

Lecture Notes in Networks and Systems 1011

Xin-She Yang
Simon Sherratt
Nilanjan Dey
Amit Joshi *Editors*

Proceedings of Ninth International Congress on Information and Communication Technology

ICICT 2024, London, Volume 1


OPEN ACCESS

 Springer

Lecture Notes in Networks and Systems

Volume 1011

Series Editor

Janusz Kacprzyk , Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Advisory Editors

Fernando Gomide, Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Türkiye

Derong Liu, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, EI Compindex, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Xin-She Yang · Simon Sherratt · Nilanjan Dey ·
Amit Joshi
Editors

Proceedings of Ninth International Congress on Information and Communication Technology

ICICT 2024, London, Volume 1

 Springer

Editors

Xin-She Yang
Middlesex University
London, UK

Simon Sherratt
The University of Reading
Reading, UK

Nilanjan Dey
Department of Computer Science
and Engineering
Techno International New Town
Kolkata, West Bengal, India

Amit Joshi
Global Knowledge Research Foundation
Ahmedabad, Gujarat, India



ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-981-97-4580-7

ISBN 978-981-97-4581-4 (eBook)

<https://doi.org/10.1007/978-981-97-4581-4>

© The Editor(s) (if applicable) and The Author(s) 2024. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

If disposing of this product, please recycle the paper.

Preface

The Ninth International Congress on Information and Communication Technology will be held during February 19–22, 2024, in a hybrid mode, physically in London, UK, and Digital Platform: Zoom. ICICT 2024 is organized by Global Knowledge Research Foundation and managed by G. R. Scholastic LLP. The associated partners were Springer and Springer Nature. The conference will provide a useful and wide platform both for display of the latest research and for exchange of research results and thoughts. The participants of the conference will be from almost every part of the world, with backgrounds of either academia or industry, allowing a real multinational multicultural exchange of experiences and ideas.

A great pool of more than 2400 papers was received for this conference from across 129 countries among which around 485 papers were accepted and will be presented physically in London and Digital Platform Zoom during the four days. Due to the overwhelming response, we had to drop many papers in the hierarchy of the quality. A total of 70 technical sessions will be organized in parallel in four days along with a few keynotes and panel discussions in hybrid mode. The conference will be involved in deep discussion and issues which will be intended to solve at global levels. New technologies will be proposed, experiences will be shared, and future solutions for design infrastructure for ICT will also be discussed. The final papers will be published in ten volumes of proceedings by Springer LNNS Series. Over the years, this congress has been organized and conceptualized with the collective efforts of a large number of individuals. I would like to thank each of the committee members and the reviewers for their excellent work in reviewing the papers. Grateful acknowledgments are extended to the team of Global Knowledge Research Foundation for their valuable efforts and support.

I look forward to welcoming you to the 10th Edition of this ICICT Congress 2025.

Amit Joshi, Ph.D.
Organising Secretary, ICICT 2024
Director—Global Knowledge Research
Foundation
Ahmedabad, India

Contents

Perceptions and Experiences of Severe Content in Content Moderation Teams: A Qualitative Study	1
Dimitra Eleftheria Strongylou, Marlyn Thomas Savio, Miriah Steiger, Timir Bharucha, Wilfredo R. Torralba Manuel III, Xieyining Huang, and Rachel Lutz Guevara	
Unveiling the Television News Puzzle: Tracking TV News Consumption Behaviors Amidst and Beyond the Covid-19 Pandemic	13
Chan Eang Teng and Tang Mui Joo	
Behaviors and Patterns of TV News Viewing in Malaysia During and After Covid-19 Pandemic	23
Tang Mui Joo and Chan Eang Teng	
Comparative Evaluation of Anomaly Detection Methods for Fraud Detection in Online Credit Card Payments	37
Hugo Thimonier, Fabrice Popineau, Arpad Rimmel, Bich-Liên Doan, and Fabrice Daniel	
Exploratory Customer Discovery Through Simulation Using ChatGPT and Prompt Engineering	51
Joseph Benjamin Ilagan, Zachary Matthew Alabastro, Claire Louise Basallo, and Jose Ramon Ilagan	
Ethical Education Data Mining Framework for Analyzing and Evaluating Large Language Model-Based Conversational Intelligent Tutoring Systems for Management and Entrepreneurship Courses	61
Joseph Benjamin R. Ilagan, Jose Ramon S. Ilagan, and Maria Mercedes T. Rodrigo	

Sensing the Emergence of New Structure in Matter and Life and Its Impact on Physics, Chemistry, and Biology, in a Light-Based Quantum Computational Model	73
Pravir Malik	
VOTUM: Secure and Transparent E-Voting System	89
Joaquin Egocheaga, William Angulo, and Cesar Salas	
Simulation and Fault Diagnostics Using I–V and P–V Curve Tracing ...	101
Kabelo Mashiloane, Peet F. Le Roux, and Coneth G. Richards	
Comparing Adopter, Tester, and Non-adopter of Collaborative Augmented Reality for Industrial Services	123
Maike Müller, Stefan Ohlig, Dirk Stegelmeyer, and Rakesh Mishra	
Cyber Victimization: Tools Used to Combat Cybercrime and Victim Characteristics	141
Marc Dupuis and Emiliya Jones	
Binary Segmentation of Malaria Parasites Using U-Net Segmentation Approach: A Case of Rwanda	163
Eugenia M. Akpo, Carine P. Mukamakuza, and Emmanuel Tuyishimire	
PLC-Based Traffic Light Control for Flexible Testing of Automated Mobility	177
Tamás Wágner, Tamás Tettamanti, Balázs Varga, and István Varga	
User Perceptions of Progressive Web App Features: An Analytical Approach and a Systematic Literature Review	187
Tulio Marchetto and Marcelo Morandini	
An Automated and Goal-Oriented Clustering Procedure	207
Oded Koren, Michal Koren, and Or Peretz	
Proposal for a New Separation Method for Reproducing Images with Properties in the Visible and Near-Infrared Spectrum	229
Jana Žiljak Gršić, Silvio Plehati, Tomislav Bogović, and Roko Vujić	
Epidemic Information Extraction for Event-Based Surveillance Using Large Language Models	241
Sergio Consoli, Peter Markov, Nikolaos I. Stilianakis, Lorenzo Bertolini, Antonio Puertas Gallardo, and Mario Ceresa	
Effects of Inductive Load on Photovoltaic Systems	253
E. Okpo, P. F. Le Roux, and O. I. Okoro	
Low Code Development Cycle Investigation	265
Małgorzata Pańkowska	

TermX—Bridging the Gap: Implementing CTS2 and FHIR Compatible Terminology Server 277
 Marina Ivanova, Igor Bossenko, and Gunnar Pihø

An Improved Technique for Generating Effective Noises of Adversarial Camera Stickers 289
 Satoshi Okada and Takuho Mitsunaga

Artificial Intelligence in Breast Cancer Diagnosis: “Synergy-Net” in Campania FESR-POR (European Fund of Regional Development—Regional Operative Program) Research Project 301
 Domenico Parmeggiani, Giancarlo Moccia, Pasquale Luongo, Francesco Miele, Alfredo Allaria, Francesco Torelli, Stefano Marrone, Michela Gravina, Carlo Sansone, Ruggiero Bollino, Roberto Ruggiero, Paola Bassi, Antonella Sciarra, Simona Parisi, Francesca Fisone, Maddalena Claudia Donnarumma, Chiara Colonnese, Paola Della Monica, Marina Di Domenico, Ludovico Docimo, and Massimo Agresti

Artificial Intelligence in Prostate Cancer Diagnosis: “Synergy-Net” in Campania FESR-POR (European Fund of Regional Development—Regional Operative Program) Research Project 313
 Domenico Parmeggiani, Marco De Sio, Giancarlo Moccia, Pasquale Luongo, Francesco Miele, Alfredo Allaria, Francesco Torelli, Stefano Marrone, Michela Gravina, Carlo Sansone, Ruggiero Bollino, Paola Bassi, Antonella Sciarra, Davide Arcaniolo, Maddalena Claudia Donnarumma, Chiara Colonnese, Lorenzo Romano, Federica Colapietra, Marina Di Domenico, Ludovico Docimo, and Massimo Agresti

Designing Reactive Route Change Rules with Human Factors in Mind: A UATM System Perspective 323
 Jeongseok Kim and Kangjin Kim

Real-Time Cyber-Physical Risk Management Leveraging Advanced Security Technologies 339
 Ramesh Chandra Poonia, Kamal Upreti, Bosco Paul Alapatt, and Samreen Jafri

Author Index 351

Editors and Contributors

About the Editors

Xin-She Yang obtained his D.Phil. in Applied Mathematics from the University of Oxford and subsequently worked at the Cambridge University and the National Physical Laboratory (UK) as a Senior Research Scientist. He is currently Reader in Modeling and Optimization at Middlesex University London and Adjunct Professor at Reykjavik University (Iceland). He is also elected Bye-Fellow at Cambridge University and the IEEE CIS Chair for the Task Force on Business Intelligence and Knowledge Management. He was included in the “2016 Thomson Reuters Highly Cited Researchers” list.

Simon Sherratt was born near Liverpool, England, in 1969. He is currently Professor of Biosensors at the Department of Biomedical Engineering, University of Reading, UK. His main research area is signal processing and personal communications in consumer devices, focusing on wearable devices and health care. He received the first place IEEE Chester Sall Memorial Award in 2006, the second place in 2016, and the third place in 2017.

Nilanjan Dey is Assistant Professor at the Department of Information Technology, Techno India College of Technology, India. He has authored/edited more than 75 books with Springer, Elsevier, Wiley, CRC Press and published more than 300 peer-reviewed research papers. He is Editor-in-Chief of the International *Journal of Ambient Computing and Intelligence*; Series Co-editor of *Springer Tracts in Nature-Inspired Computing* (STNIC); and Series Co-editor of *Advances in Ubiquitous Sensing Applications for Healthcare*, Elsevier.

Amit Joshi is Director of Global Knowledge Research Foundation, also Entrepreneur and Researcher who has completed his Masters’ and research in the areas of cloud computing and cryptography in medical imaging. He has an experience of around ten years in academic and industry in prestigious organizations. He is Active Member

of ACM, IEEE, CSI, AMIE, IACSIT-Singapore, IDES, ACEEE, NPA, and many other professional societies. He is International Chair of InterYIT at International Federation of Information Processing. He has presented and published more than 50 papers in national and international journals/conferences of IEEE and ACM. He has also edited more than 40 books which are published by Springer, ACM, and other reputed publishers. He has also organized more than 50 national and international conferences and programs in association with ACM, Springer, and IEEE to name a few across different countries including India, UK, Europe, USA, Canada, Thailand, Egypt, and many more.

Contributors

Massimo Agresti Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Eugenia M. Akpo Carnegie Mellon University Africa, Kigali, Rwanda

Zachary Matthew Alabastro Ateneo de Manila University, Quezon City, Philippines

Bosco Paul Alapatt Department of Computer Science, Christ University, Ghaziabad, India

Alfredo Allaria Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

William Angulo Peruvian University of Applied Sciences, Lima, Peru

Davide Arcaniolo Department of Woman, Child and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy

Claire Louise Basallo Ateneo de Manila University, Quezon City, Philippines

Paola Bassi Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Lorenzo Bertolini European Commission, Joint Research Centre (JRC), Ispra, Italy

Timir Bharucha TaskUs Inc, New Braunfels, TX, USA

Tomislav Bogović Zagreb University of Applied Sciences, Zagreb, Croatia

Ruggiero Bollino Bollino.It SpA, Naples, Italy

Igor Bossenko Department of Software Science, Tallinn University of Technology, Tallinn, Estonia

Mario Ceresa European Commission, Joint Research Centre (JRC), Ispra, Italy

Federica Colapietra Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, Naples, Italy

Chiara Colonnese Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Sergio Consoli European Commission, Joint Research Centre (JRC), Ispra, Italy

Fabrice Daniel LUSIS AI, Paris, France

Marco De Sio Department of Woman, Child and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy

Marina Di Domenico Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, Naples, Italy

Bich-Liên Doan CNRS, CentraleSupélec, Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, Gif-sur-Yvette, France

Ludovico Docimo Department of Woman, Child and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy

Maddalena Claudia Donnarumma Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Marc Dupuis University of Washington, Bothell, WA, USA

Joaquin Egocheaga Peruvian University of Applied Sciences, Lima, Peru

Francesca Fisone Division of General, Mini-Invasive and Obesity Surgery, University of Campania “Luigi Vanvitelli”, Oncological Naples, Italy

Antonio Puertas Gallardo European Commission, Joint Research Centre (JRC), Ispra, Italy

Michela Gravina Department of Electrical Engineering and Information Technology (DIETI), University “Federico II”, Naples, Italy

Jana Žiljak Gršić Zagreb University of Applied Sciences, Zagreb, Croatia

Rachel Lutz Guevara TaskUs Inc, New Braunfels, TX, USA

Xieying Huang TaskUs Inc, New Braunfels, TX, USA

Jose Ramon Ilagan Ateneo de Manila University, Quezon City, Philippines

Jose Ramon S. Ilagan Ateneo de Manila University, Quezon City, Philippines

Joseph Benjamin Ilagan Ateneo de Manila University, Quezon City, Philippines

Joseph Benjamin R. Ilagan Ateneo de Manila University, Quezon City, Philippines

Marina Ivanova Department of Software Science, Tallinn University of Technology, Tallinn, Estonia

Samreen Jafri Administrative Science Department, Imam Abdulrahman Bin Faisal University, Dammam, Kingdom of Saudi Arabia

Emiliya Jones University of Washington, Bothell, WA, USA

Tang Mui Joo Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia

Jeongseok Kim SK Telecom, Seoul, Republic of Korea

Kangjin Kim Department of Drone Systems, Chodang University, Jeollanam-do, Republic of Korea

Michal Koren School of Industrial Engineering and Management, Ramat-Gan, Israel

Oded Koren School of Industrial Engineering and Management, Ramat-Gan, Israel

P. F. Le Roux Tshwane University of Technology, Pretoria, South Africa

Peet F. Le Roux Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa

Pasquale Luongo Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Pravir Malik Deep Order Technologies, El Cerrito, CA, USA

Wilfredo R. Torralba Manuel III TaskUs Inc, New Braunfels, TX, USA

Tulio Marchetto Universidade de São Paulo, São Paulo, SP, Brazil

Peter Markov European Commission, Joint Research Centre (JRC), Ispra, Italy

Stefano Marrone Department of Woman, Child and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy; Department of Electrical Engineering and Information Technology (DIETI), University “Federico II”, Naples, Italy

Kabelo Mashiloane Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa

Francesco Miele Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Rakesh Mishra School of Computing and Engineering, University of Huddersfield, Huddersfield, UK

Takuho Mitsunaga INIAD, Toyo University, Tokyo, Japan

Giancarlo Moccia Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Paola Della Monica Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, Naples, Italy

Marcelo Morandini Universidade de São Paulo, São Paulo, SP, Brazil

Carine P. Mukamakuza Carnegie Mellon University Africa, Kigali, Rwanda

Maike Müller School of Computing and Engineering, University of Huddersfield, Huddersfield, UK;

Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, Frankfurt, Germany

Stefan Ohlig School of Computing and Engineering, University of Huddersfield, Huddersfield, UK;

Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, Frankfurt, Germany

Satoshi Okada INIAD, Toyo University, Tokyo, Japan

O. I. Okoro Michael Okpara University of Agriculture, Umudike, Nigeria

E. Okpo Tshwane University of Technology, Pretoria, South Africa

Simona Parisi Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy;

Division of General, Mini-Invasive and Obesity Surgery, University of Campania “Luigi Vanvitelli”, Oncological Naples, Italy

Domenico Parmeggiani Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Malgorzata Pańkowska Department of Informatics, University of Economics in Katowice, Katowice, Poland

Or Peretz School of Industrial Engineering and Management, Ramat-Gan, Israel

Gunnar Pihö Department of Software Science, Tallinn University of Technology, Tallinn, Estonia

Silvio Plehati Zagreb University of Applied Sciences, Zagreb, Croatia

Ramesh Chandra Poonia Department of Computer Science, Christ University, Ghaziabad, India

Fabrice Popineau CNRS, CentraleSupélec, Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, Gif-sur-Yvette, France

Coneth G. Richards Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa

Arpad Rimmel CNRS, CentraleSupélec, Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, Gif-sur-Yvette, France

Maria Mercedes T. Rodrigo Ateneo de Manila University, Quezon City, Philippines

Lorenzo Romano Department of Neurosciences, Reproductive Sciences and Odontostomatology “Federico II”, University of Naples, Naples, Italy

Roberto Ruggiero Bollino.It SpA, Naples, Italy;
Division of General, Mini-Invasive and Obesity Surgery, University of Campania “Luigi Vanvitelli”, Oncological Naples, Italy

Cesar Salas Peruvian University of Applied Sciences, Lima, Peru

Carlo Sansone Department of Electrical Engineering and Information Technology (DIETI), University “Federico II”, Naples, Italy;
CINI, ITEM Laboratory “C.Savy”, Naples, Italy

Marlyn Thomas Savio TaskUs Inc, New Braunfels, TX, USA

Antonella Sciarra Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Dirk Stegelmeyer Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, Frankfurt, Germany

Miriah Steiger TaskUs Inc, New Braunfels, TX, USA

Nikolaos I. Stilianakis European Commission, Joint Research Centre (JRC), Ispra, Italy

Dimitra Eleftheria Strongylou TaskUs Inc, New Braunfels, TX, USA

Chan Eang Teng Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur, Malaysia

Tamás Tettamanti Department of Control for Transportation and Vehicle Systems, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Budapest, Muegyetem rkp. 3, Hungary

Hugo Thimonier CNRS, CentraleSupélec, Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, Gif-sur-Yvette, France

Francesco Torelli Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy

Emmanuel Tuyishimire College of Science and Technology, University of Rwanda, Kigali, Rwanda

Kamal Upreti Department of Computer Science, Christ University, Ghaziabad, India

Balázs Varga Department of Control for Transportation and Vehicle Systems, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Budapest, Muegyetem rkp. 3, Hungary

István Varga Department of Control for Transportation and Vehicle Systems,
Faculty of Transportation Engineering and Vehicle Engineering, Budapest
University of Technology and Economics, Budapest, Muegyetem rkp. 3, Hungary

Roko Vujić Faculty of Graphic Arts, University of Zagreb, Zagreb, Croatia

Tamás Wágner Department of Control for Transportation and Vehicle Systems,
Faculty of Transportation Engineering and Vehicle Engineering, Budapest
University of Technology and Economics, Budapest, Muegyetem rkp. 3, Hungary

Perceptions and Experiences of Severe Content in Content Moderation Teams: A Qualitative Study



Dimitra Eleftheria Strongylou , Marlyn Thomas Savio, Miriah Steiger ,
Timir Bharucha, Wilfredo R. Torralba Manuel III, Xieyining Huang ,
and Rachel Lutz Guevara

Abstract Existing Trust & Safety policies predominantly focus on protecting content moderators' (CoMos) safety against severe content violations. Nevertheless, in reality, CoMos might encounter content across a broad severity spectrum while perceptions of content severity among CoMos are yet to be fully understood in relation to their well-being and job accuracy. The current study employs a sequential study design to qualitatively examine the views of 34 CoMos in the USA and the Philippines and of 166 professionals (i.e., mental health and learning experience staff, team leads) supporting CoMos on content severity and its perceived impact on CoMos well-being and job accuracy. Of note, consistent with ethical guidelines and wellness best practices, all CoMos received tiered wellness support services from TaskUs Inc. Collected data were thematically analyzed and two overarching themes emerged. Within the first theme, namely, 'perceived content categories', both CoMos and professionals, after considering the potential distress caused by various contents, broadly distinguished between 'moderate' and 'graphic' content, while they further attributed their own definitions to each of the two categories. Under theme two—'content impact on perceived stress- most study participants discussed how reviewing moderate content might have a more detrimental impact upon CoMos stress and job accuracy compared to reviewing graphic content in the lack of wellness support and care. This study is the first of its kind to give voice to both CoMos and support staff to discuss in depth their views on content moderation while also making participant-led recommendations for work policy and future research.

D. E. Strongylou (✉) · M. T. Savio · M. Steiger · T. Bharucha · W. R. T. Manuel III · X. Huang · R. L. Guevara

TaskUs Inc, 1650 Independence Dr, New Braunfels, TX 78132, USA
e-mail: Dimitra.Strongylou@taskus.com

M. T. Savio
e-mail: Marlyn.Savio@taskus.com

X. Huang
e-mail: irene.huang@taskus.com

R. L. Guevara
e-mail: rachel.lutzguevara@taskus.com

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_1

Keywords Content moderation · Content severity · Thematic analysis · Content categories · Trust and safety policies

1 Introduction

Content moderation emerged from the global need to regulate user generated posts on digital platforms (e.g., social media, gaming portals) that have the potential to harm or mislead others. Reports from social media giants suggest that harmful posts run in the millions within a calendar quarter (3 months)—nearly 2 billion at Facebook [1], and about 30 million at YouTube™ [2]. As a result, policies have been developed and fine-tuned over the years by social media businesses [3] and governments (e.g., UK Online Safety Bill) to detect and purge harmful content from the internet [4]. Eight broad categories of content are typically considered harmful for platform users. These include online abuse, bullying/harassment, threats, impersonation, unwanted sexual gestures, violence, self-harm/suicide, and pornography [5]. In addition to these content categories, context and target audience influence the labeling and actions of flagged posts [6]. The terms of moderation policies are, also, largely based on the severity of harmful content and how it may impact users' well-being. For instance, content containing child abuse is generally deemed highly detrimental while pornography may be viewed as comparatively innocuous. For these reasons, policies attempt to be user-centered and severity-driven.

Lesser known yet critical stakeholders in the moderation process are content moderators (CoMos). CoMos tend to inadvertently be the first human audience exposed to graphic user generated posts, oftentimes in unfiltered forms [7]. Without psychosocial support programs, such exposure (particularly to harmful content) can be detrimental for moderators' psychological health [8]. Moreover, it is especially common for CoMos to review digital posts by users from societies and groups contrary to their own beliefs. Because platform codes and guidelines are not culturally comprehensive, CoMos and platform users may have different concepts of what constitutes harmful content [9]. Likewise, regulatory laws are often localized and even markedly in contrast with those of other nations and regions [10]. In this scenario, CoMos' perceptions and experiences of content may play a key role in determining the fate of content on platforms.

In addition to content decisions, CoMos' job performance may also be impacted by their perceptions given that accuracy and handle time are key metrics in Trust & Safety evaluations [11]. For moderators to be able to deliver the speed and efficiency in their job, they arguably apply mental heuristics for a quick and reliable resolution of content queues. It then becomes critical to study and appreciate the psychological and cognitive influences on job performance.

While efforts have improved over the past years to protect CoMos' safety considering their regular exposure to potentially disturbing material, it is equally vital for policies to also be responsive to CoMos' perceptions and experiences. Much of the focus of policy and training for content moderation is targeted to extremely

severe content violations. However, CoMos encounter content across a spectrum of severity, ranging from benign to harmful [12]. This creates a need to explore how CoMos perceive severity, and how they manage harmful content without impeding job performance or personal well-being.

Theorists and researchers argue against a simplistic, singular notion of content severity. A qualitative interview-based study [13] proposed a 10-dimension framework to reflect that severity perception involves multi-factorial reasoning. The findings showed, that whether the participant was a user, CoMo, researcher or platform owner, the interpretation of severity was rarely found to be based on legality and more likely determined by the perspective (victim/perpetrator), intent (low/high harm), agency (choice/coercion), experience (personal/objective), scale (large/small), urgency (high/low), vulnerability (high/low), medium (video/text), and sphere (public/private) of the post. This highlights that being cognizant of the multitude of aspects that influence discernment of content severity can result in less ambiguous and inconsistent implementation of moderation policies.

Technological innovations such as the application of artificial intelligence (AI) in content moderation have enhanced automation and scalability options [14]. However, independent machine-led moderation of harmful content without human intervention has largely misfired including resulting in several false positives [15]. Much of cultural knowledge and awareness of new social norms is tacit and ever evolving [16]. This makes it cumbersome for AI systems to completely take over content moderation and shield human counterparts from the exposure to harmful content. Content moderation policies would, thus, benefit from insight about human perceptions of content severity. Currently, there is scarce documentation of CoMos' beliefs and practices in relation to the severity of content.

This study aimed to capture the viewpoints of different team members who conduct and/or support content moderation. By means of qualitative approaches, the present study explored how CoMos and support staff (e.g., team leaders, mental health professionals and learning experience staff) defined severe content, and how reviewing such content affected CoMos' well-being and job accuracy. The findings of this study are expected to inform the training and resources needed to improve the content moderation workflow and quality of work experience.

2 Methods

2.1 Study Design

A sequential study design was employed: a qualitative interview study followed by a qualitative online questionnaire survey were completed to holistically examine the perceptions and views of CoMos and support staff (e.g., team leaders, mental health professionals, and learning experience staff).

Semi-structured interviews were completed with CoMos in November 2020 in order to secure the in-depth and open exploration of CoMos' views on content severity. Next, an online qualitative survey consisting of open-ended questions was conducted with team leads, mental health professionals, and learning experience staff between February and March 2021, in order to capture support staffs' views on content severity in a timely manner.

2.2 *Participants*

A purposive sample of 34 CoMos from TaskUs Inc. partook in the first qualitative study. All participants were employed in lines of business involving potentially egregious content in the format of videos, images and texts, in line with the scope of the study to examine content severity perspectives. All interviewees moderated various social media platforms and dating platforms and applications. The average tenure of the sample was one-year ($M = 14.2$ months, $SD = 16.1$ months), with roughly equal proportion of male ($n = 15$) and female ($n = 19$) participants. Given that most lines of business within TaskUs Inc. are located in the Philippines, most ($n = 24$) CoMos were located in the Philippines while ten CoMos were living in the USA at the time of study completion.

It should be noted that all CoMos participating in the current qualitative study were entitled to tiered wellness support services provided by TaskUs Inc. As a result, CoMos who were more likely to be exposed to egregious content received greater wellness support and care. TaskUs Inc. wellness support services encompass both reactive and preventative wellness interventions delivered in different formats, tailored to the needs of the individuals. Popular examples of wellness interventions include self-service digital wellness services, individual- and group-level wellness sessions with licensed mental health professionals, 24/7 access to psychological services, and embedded wellness tools, such as the Centered tool, an online platform focusing on improving CoMos' emotional wellness. TaskUs Inc. wellness interventions have been proven to be effective in protecting CoMos' mental health against the potentially harmful effects of moderating egregious content [8].

The participant sample for the online questionnaire consisted of 93 team leads, 35 mental health professionals—including psychiatrists, counselors, and life coaches—and 38 learning experience staff. All participants in the sample were employed at the same organization as the CoMos. These employees were selected given their close working relationship with CoMos moderating potentially egregious content.

2.3 *Inclusion Criteria*

To be eligible for the interview, CoMos must have been working in campaigns involving potentially egregious content for at least six months. Similarly, support

staff were eligible for the questionnaire if they worked with CoMos working in campaigns that involve potentially egregious content.

2.4 Procedure and Ethical Approval

Both studies were conducted in line with the code of conduct, legal regulations, and ethical guidelines stipulated by TaskUs Inc. and were independently reviewed by an external mental health researcher. No reports of adverse outcomes were noted during the course of the two studies.

Similar recruitment strategies were utilized for both studies. A study information sheet explaining the scope of the study, participant's risks and benefits, voluntary participation, and confidentiality issues were circulated via email to CoMos and support staff, meeting the study inclusion criteria.

CoMos interested in participating in the study contacted the TaskUs Wellness + Resilience Research Team to schedule an interview. Prior to starting the interviews, CoMos were required to electronically sign a consent form. On the interview day, researchers, TB and TMT, explained in detail the scope of the study and allowed time for participants to ask any questions. Interviews lasted between 30 min and 1 h. All interviewees were given the opportunity to pause for questions or terminate the interview earlier in case they experienced discomfort. Of note, all interviews were completed as planned, without any opt-out requests.

Support staff interested in completing the online survey form were provided the survey completion link. Prior to proceeding to survey completion, participants declared consent. Data collected from participants exiting the survey before completion were disregarded prior to analysis. The interview recordings and the completed survey responses were stored securely in an encrypted drive accessible only to the researchers of the studies. Interview data were transcribed verbatim prior to analysis. Data collected from both studies were then uploaded to MaxQDA for analysis.

2.5 Materials

A semi-structured topic guide consisting of nine open questions was utilized for all interviews. Questions examined CoMos' perceptions regarding content categories, disturbing content, and the impact of moderated content on their mental health and well-being. A translated version of the topic guide was used for participants in the Philippines.

An open text questionnaire was developed based on the findings of the interviews with CoMos and administered to all support staff. Although most of the questions largely overlapped, support staff were also asked questions unique to their professions. In particular, survey questions asked: (1) professional characteristics (e.g., professional role, location, site and campaign); (2) their perceptions of CoMos' most

challenging and distressing jobs (e.g., ‘Please give me some examples of the most emotionally distressing jobs you believe CoMos encounter’); (3) their perceptions of CoMos’ least distressing and easiest jobs (e.g., Please provide examples of jobs that provide a sense of relief and are less overwhelming); and (4) their perceptions of CoMos’ content categories and workflow issues (e.g., please list the most common CoMos’ concerns when it comes to workflow).

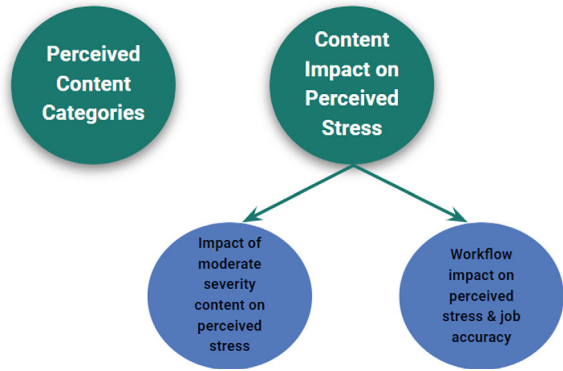
2.6 Analysis

A three-stage process was followed for the analysis of the two studies. Firstly, given the exploratory nature of the interviews, we allowed findings to emerge from the collected data (i.e., a bottom-up approach). Collected data were analyzed with inductive thematic analysis, as described by Braun and Clarke [17]. A five-step approach was taken: (1) familiarization, (2) coding (3) generating themes, (4) reviewing themes, and (5) defining themes. Familiarization refers to understanding and becoming comfortable with the responses by reading transcripts over multiple times. Next, data were closely examined, and codes were assigned to specific responses or portions of responses. Following that, coded data were grouped into categories. After carefully reviewing the coded responses and associated categories, themes were identified. This involved condensing similar categories into broader, generalizable themes. Themes were then reviewed by comparison to the responses to ensure the themes were supported by data. Finally, themes were appropriately labeled and defined. Secondly, the responses from support staff were thematically analyzed following the same 5 steps described above [17]. Finally, emerging themes from the questionnaire survey were compared against the themes already identified in the interview study. Therefore, although the findings of each study were separately generated, they were synthesized and presented together below to facilitate a comprehensive understanding of the topic of content severity under examination. To enhance rigor and reliability, two independent researchers were involved in the three stages of the analysis. One researcher led the analysis while a second researcher reviewed codes and final themes; any disagreements were discussed until reaching consensus.

3 Results

Two main themes were identified from the data (Fig. 1). The first theme, perceived content categories, represented the views of CoMos and support staff around the categorization of content reviewed into ‘graphic’ and ‘moderate’. The second theme, content impact on perceived stress, consists of two subthemes impact of moderate-severity content on perceived stress and workflow impact on perceived stress and job

Fig. 1 Themes and subthemes emerged from CoMos and support staff (e.g., learning experience staff, team leads, and mental health professionals)



accuracy which discuss the impact of graphic and moderate content, alongside the impact of workflow issues, upon CoMos’ perceived stress and job accuracy.

3.1 Theme 1: Perceived Content Categories

Within this theme, both CoMos and support staff provide an insightful account of content severity categories. Participants distinguished between graphic and moderate-severity content categories. Both groups stated that ‘graphic content’ includes child abuse, violence (e.g., torture, execution, murder, assault, etc.), animal abuse, and illicit sexual content. Both groups suggested that if CoMos do not receive timely and tailored wellness interventions, graphic content might cause increased distress.

“Like it’s, it’s just a replay video, of course, but it’s a live video. He is just, just chatting and then suddenly he gets a gun shot and then blows off her head, blows up his head. So that’s the time that I got, oh well, it’s, it’s different from movies. If you are seeing it live or you’re seeing it’s like the actual people, do it. It’s kind of different. It’s a... I get goosebumps, but then I have to watch it.” – CoMo

“For me personally, I do not want to moderate child abuse content. So that is the most disturbing for me.” – CoMo

“[The most distressing content is] imagery of suicide and self-injury jobs along with child exploitation imagery” - Team Lead

“[The most concerning content is] child abuse or pedophilia - videos where children were being raped, clients expressed their distress especially overhearing the children cry” - Support Staff

‘Moderate-severity content’ was broadly grouped under politics, bullying, harassment, and attacks on religion, race, gender, or sexual orientation by both CoMos and support staff. As most CoMos and professionals explained, compared to graphic content, moderate-severity content itself does not cause the same level of distress to CoMos.

“Those hate posts or controversial posts do not give me that kind of heavy feeling at all”- CoMo
 “Posts with nudity, bullying but with no targets...the descriptions can be funny sometimes...[these are less overwhelming]” - Support Staff

3.2 Theme 2—Content Impact on Perceived Stress

This theme illustrates how graphic content can have a greater impact at the beginning of CoMos employment. As most CoMos and mental health professionals said, CoMos usually experience the greatest impact from reviewing graphic content on their stress levels at the beginning of their employment, when they are for the first time exposed to graphic content. However, over time, CoMos acclimate and habituate to reviewing graphic content, meaning that they are not negatively impacted by graphic content to the same extent.

“I don’t feel like I’m affected as much as I was in the beginning of the campaign.” – CoMo
 “[Over time] they [CoMos] are used to what they are viewing, and they know how to normalize some views that can be disturbing for them and to others.” - Support Staff

Further, this theme consists of two sub-themes which describe the impact of reviewing moderate-severity content and the impact of workflow issues upon CoMos’ perceived stress and job accuracy, respectively.

Subtheme 1: Impact of Moderate-Severity Content on Perceived Stress

Within this first sub-theme, ‘impact of moderate-severity content on perceived stress’ CoMos and support staff discuss how moderate content can potentially have a greater negative effect upon CoMos’ stress compared to graphic content.

As most of the CoMos described, there are fewer guidelines and more ambiguity when reviewing moderate content compared to when reviewing graphic content where more succinct guidance is provided. Without clear guidelines for moderate content review, CoMos need more time to properly review and consider moderate content compared to graphic content which often has clear guidelines. As such, some CoMos feel more concerned about accomplishing performance targets when reviewing moderate content.

“Even though they are [content] graphic it is easier to be categorized or if it needs to be removed immediately. We don’t stay on it too much.”- CoMo
 “It [graphic content] was easier to review because the violation there was explicit. You already know what action to do... less disturbing [moderate content] where you need to analyze what they meant.” – CoMo

Similarly, learning experience staff and team leads suggested that the graphic content review process may be considered uncomplicated as long as guidelines and protocols are clear and CoMos do not utilize subjective assessments to moderate content. Teamleads further highlighted CoMos’ perceived ambiguity of guidelines

as central in increasing CoMos' worries about job accuracy, handle time, and meeting production goals.

"[The most common concern in workflow is] subjectivity in guidelines" - Support Staff

Subtheme 2—Workflow Impact on Perceived Stress and Job Accuracy

Within this sub-theme, CoMo interviewees and mental health professionals extensively discuss the impact that workflow issues, such as increased workload and supervisor and management issues, might have upon CoMos' perceived stress. Most of the CoMos participants reported they felt more overwhelmed by workflow and procedures rather than by moderating graphic content itself.

"It is more stressful with a higher number of jobs; you are chasing the number of jobs compared to the severity of what you see." – CoMo

Similarly, as mental health professionals discussed, during their wellness sessions with CoMos, they are more often concerned with workflow and workload related issues rather than wellness issues related to reviewing graphic content per se. In particular, as many mental health professionals pointed out, most CoMos' concerns predominantly focus on work-related issues concerning supervisor and management issues, pressure from leadership, perceived lack of support, increased stress due to workload demands, dealing with the unexpected (related to content or workflow), and abrupt changes in procedures and protocols.

"[The most common concern in workflow is] abrupt changes of schedule and protocols" - Support Staff

In addition to the impact of workflow issues on their perceived stress, CoMos also expressed their worries about how rushing through jobs, due to increased workload and work demands, could negatively impact their job accuracy. In fact, some CoMos declared they would prefer to review a lower number of queues containing graphic content rather than reviewing a higher number of queues that do not contain graphic content.

"[I would prefer reviewing more graphic content rather than taking on more jobs because] if it is graphic, it is easier to determine if there are any violations."- CoMo

"I prefer a higher number of graphic content queues because it is easier to be identified and it seems easier to take action." – CoMo

Team leads and learning experience staff further articulated that workflow issues such as unclear, fast-changing, or overlapping protocols and guidelines can negatively impact job accuracy and may potentially lead to missed jobs.

"[CoMos encounter missed jobs] when there are subjective signals that they [CoMos] tend to overthink." - Support Staff

4 Discussion

The present qualitative study explored the perceptions of CoMos and support staff regarding graphic content; and assessed the impact of reviewing graphic content on CoMos' perceived stress and job accuracy. Regarding our first research aim, the findings of our study are consistent with and further add to the existing limited research [5], showing that the main categories of graphic content predominantly concern child abuse, violence, animal abuse, and illicit sexual content.

Our results cast a new light on how moderate content and workflow issues can have a large, negative impact upon CoMos' perceived stress and job accuracy. These findings suggest that scarce guidelines and lack of clarity for reviewing moderate content can negatively affect CoMos' perceived stress and job accuracy. Likewise, workflow issues impact upon some CoMos' stress levels.

Previous research has thoroughly discussed the detrimental impact of repeated exposure to egregious content upon CoMos' mental health—especially in the lack of appropriate wellness support [18, 19]. For instance, as Rees et al. [20] showed, the scarcity of appropriate aid among CoMos might result in elevated risk for burnout, compassion fatigue, secondary traumatic stress, depression, and anxiety. However, CoMos participating in our study did not discuss and highlight any negative impact of reviewing graphic content upon their mental health. This might be potentially explained by the fact that all interviewees in our study were offered tiered wellness services, designed to address trauma-related experiences. These wellness services have been proven effective in boosting CoMos' mental health, by maintaining baseline resiliency levels and protecting against burnout risk [8]. As such, CoMos in our study might have been more concerned with workflow-related issues rather than the graphic content per se.

Interestingly, the CoMos interviewees of this study pointed out that they often experienced the greatest impact from reviewing graphic content at the beginning of their employment. However, over time, they acclimate and habituate to graphic content. This finding is in agreement with Bharucha et al. [21] qualitative study that discussed how CoMos experience the greatest impact at the start of employment. These authors further identified various factors predicting startle response and habituation among content moderators [21].

This is the first study to examine and synthesize the views of CoMos and support staff about the impact that moderate or severe content may have on CoMos' well-being and job accuracy. The qualitative approach provided an in-depth investigation of the impact on CoMos' perceived stress while considering the insights of support staff. However, convenience samples can pose limitations about generalizability of our findings to different populations and settings. An additional limitation of this study is that support staff were not individually interviewed. This may have prevented support staff from extensively expressing any additional thoughts on the topic. Future qualitative studies may be designed to gather more detailed accounts from different settings (e.g., companies, regions, clients) to understand whether

CoMos and professionals share similar experiences and views to the participants of this study.

The current findings also suggest the need for further research to explore the impact of moderate content and workflow issues on CoMos' wellness. Prospective studies may focus on the commonly used coping mechanisms by CoMos. Additional research might also focus on examining the link between moderate content and workflow issues, and extensively document the specific wellness facets that may be negatively impacted as a result of reviewing moderate content.

Our findings further highlight the importance of developing additional interventions, policies and strategies that focus on moderate content. Therefore, it is recommended that workplace wellness interventions that address work-related stress and boost active coping skills (e.g., time management and cognitive skills), in addition to content impact, are developed and delivered to CoMos. This recommendation is further substantiated by past research showing that enhanced cognitive skills such as self-talk or imagery can lead to decreased stress levels in the workplace and, subsequently, to improved productivity [22]. As our study indicated, a lack of clear guidelines on content moderation requirements can have a detrimental impact upon CoMos' perceived stress and job accuracy. This finding should be taken into consideration by policy makers and clients. When developing and updating content moderation protocols and guidelines, clients should assess protocol clarity by reflecting on the feedback provided by CoMos. Finally, content moderation training, particularly at the beginning of CoMos' employment, should be further updated and improved to include training focused on enhancing CoMos' problem-solving skills and confidence when dealing with complex and gray area guidelines and protocols.

Acknowledgements We thank, first and foremost, the people who participated in this study. We would also like to thank Priyanka Manchanda who provided assistance with data entry.

References

1. Richter F (2020) Toxic content runs rampant on Facebook. Statista com
2. McCarthy N (2018) The content getting flagged on youtube [Internet]. [1 Sept 2023]. <https://tinyurl.com/3t8235u7>
3. Trust & Safety Professional Association (2023) Policy models [Internet]. 2023 [1 Sept 2023]. <https://tinyurl.com/27ecv2y4>
4. Feingold S (2023) The UK's online safety bill could transform the internet. Here's how [Internet]. [1 Sept 2023]. <https://tinyurl.com/bdfbs7jm>
5. UK Safer Internet Centre (2023) Reporting harmful content [Internet]. [1 Sept 2023]. <https://tinyurl.com/yezywyk7>
6. Morrow G, Swire-Thompson B, Polny JM, Kopec M, Wihbey JP (2022) The emerging science of content labeling: contextualizing social media content moderation. *J Am Soc Inf Sci* 73(10):1365–1386
7. Karunakaran S, Ramakrishan R (2019) Testing stylistic interventions to reduce emotional impact of content moderation workers. *Proceed AAAI Conferen Hum Comput Crowdsour* 7(1):50–58

8. Steiger M, Bharucha TJ, Torralba W, Savio M, Manchanda P, Lutz-Guevara R (2022) Effects of a novel resiliency training program for social media content moderators. In: Proceedings of seventh international congress on information and communication technology: ICICT 2022, London, vol 4, Springer
9. Jiang JA, Scheuerman MK, Fiesler C, Brubaker JR (2021) Understanding international perceptions of the severity of harmful content online. *PLoS ONE* 16(8):e0256762
10. Wardle C (2019) Challenges of content moderation: define “harmful content” [Internet]. [1 Sept 2023]. <https://tinyurl.com/ys5wc3t6>
11. Bhatlapenumarthy H, Gresham J (2023) Metrics for content moderation [Internet]. [1 Sept 2023]. <https://tinyurl.com/2p9vhdyb>
12. Steiger M, Bharucha TJ, Venkatagiri S, Riedl MJ, Lease M (2021) The psychological well-being of content moderators: the emotional labor of commercial moderation and avenues for improving support. In: Proceedings of the 2021 CHI conference on human factors in computing systems
13. Scheuerman MK, Jiang JA, Fiesler C, Brubaker JR (2021) A framework of severity for harmful content online. *Proceed ACM Hum Comput Interact* 5(CSCW2):1–33
14. Darbinyan R (2022) The growing role of AI in content moderation [Internet]. [1 Sept 2023]. <https://tinyurl.com/yc86v6mk>
15. Kersley A (2023) The one problem with AI content moderation? it doesn't work [Internet]. [1 Sept 2023]. Available from: <https://tinyurl.com/4tmyedx5>
16. Duarte N, Llanso E, Loup A (2017) Mixed messages? the limits of automated social media content analysis
17. Braun V, Clarke V (2006) Using thematic analysis in psychology. *Qual Res Psychol* 3(2):77–101
18. Roberts ST (2019) Behind the screen. Yale University Press
19. Wohn DY (2019) Volunteer moderators in twitch micro communities: how they get involved, the roles they play, and the emotional labor they experience. In: Proceedings of the 2019 CHI conference on human factors in computing systems
20. Rees CS, Breen LJ, Cusack L, Hegney D (2015) Understanding individual resilience in the workplace: the international collaboration of workforce resilience model. *Front Psychol* 6
21. Bharucha T, Steiger ME, Mere R, Manchanda P (2022) Content moderator startle response: a qualitative study
22. Michie S (2002) Causes and management of stress at work. *Occup Environ Med* 59(1):67–72

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Unveiling the Television News Puzzle: Tracking TV News Consumption Behaviors Amidst and Beyond the Covid-19 Pandemic



Chan Eang Teng  and Tang Mui Joo

Abstract Concerns have been raised about the relevance of TV news with the fast growing of media technology. Increasing opportunities of choices have been held responsible for diverting viewers away from TV news. However, the Movement Control Order due to Covid-19 pandemic has paved a new life to TV news which was once claimed to be outdated and irrelevant as people started to select online platforms such as online news websites and social media as the main source of information. In view of the limited and insufficient research about TV news viewing where concerns are always placed on newspaper news, a survey was undertaken to determine the TV news viewing behavior in Sarawak during and after Covid-19 pandemic. A total of 306 respondents was collected through the use of snowball sampling from Central, Southern and Northern Sarawak. Despite bearing a paradox, the findings offer a paradigm shift to the previous research, as it indicates that watching TV news is still important to get the latest information. TV news via other devices only intensifies the shift of forms but not the content of the news. The content of the TV news is still the major consideration to stay engaged with the viewers. Result shows not much changes of the patterns of use of TV news during and after the Covid-19 pandemic. This implies the persistent importance and relevance of TV news in the daily life of people living in Sarawak. Producing quality content which suits the needs of the viewers is the only pathway to sustain the TV news.

Keywords TV news · Uses and gratification theory · Sarawak

1 Introduction

Television plays an important role to obtain various information, particularly during the Covid-19 Pandemic where the movement of people has been restricted and information about public health care from the government is crucial. In view of this, media

C. E. Teng (✉) · T. M. Joo

TAR University of Management and Technology, 53300 Kuala Lumpur, Malaysia
e-mail: eamteng@hotmail.com

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_2

are expected to help government and health-care facilities to provide transparent and clear communication with front liners and the public to create the greatest awareness to the community [1]. TV news might remain as a widely used and important news source, and will remain for older people for years to come. Nonetheless, the way people consume news has undergone significant transformation due to the rise of new media. In the past, individual used to adhere to scheduled time slots and watch TV for their daily news updates. In the environment of the increasing number of innovations in the media, particularly the digitization of media, the news viewing behavior has changed due to all the flexibility and convenience of technology. Nowadays, TV news is accessible via alternative platforms such as FB, Instagram, Youtube, etc. The attributes of social media are advantageous for individual to promptly assess and easily navigate news sources. In Malaysia, the TV viewership has increased when the government implemented the MCO [2]. Media Prima's news program viewership has also increased by 56% during the MCO as compared to pre-pandemic period [3]. According to Nielsen Television Audience Measurement report [4], the TV viewership has risen with an average of 30% viewers after the implementation of MCO in Malaysia. Castriota, Delmastro and Tonin [5] found that TV news is gaining higher recognition as the main source of news and information during the pandemic. These changes have surprisingly surpassed the recent digital revolution which helped to create another breakthrough in the development of TV industries. Focus is still placed in television as it is still considered as the most prominent and trusted source of information in many countries [6]. Davood Mehrabi [7] shows that TV is one the most credible sources to get the timeliness and credible news. A substantial portion of previous research has concentrated on news within newspaper and social media in Malaysia. However, a significant gap exist in the existing literature regarding TV news in a broader context, and more specifically, the consumption of TV news in Sarawak. Consequently, the extent to which these findings can be generalized is constrained. This research is intended to determine the behavior and patterns of television news consumption during Covid-19 pandemic in Sarawak. It is also to investigate the significance of television news on Sarawakian during the pandemic and to ascertain the changes of television news behavior during and after the pandemic. This study holds importance because the insight gained from this research can contribute to ensuring the TV industry's continued relevance in the digital era. This research also serves as a base to enhance the TV news viewing experience and increase the exposure and the revenue of the TV industry which used to be claimed as outdated. Furthermore, it will undoubtedly aid the government in disseminating information crucial for the nation's development and offer valuable research guidance for future studies focused on rural communities.

2 Literature Review

2.1 *TV Viewing and Consumption Pattern*

Local news consumption among youth is a bit low (reason unclear)-might be due to the lack of interest, low news information need, poor perception of media credibility [8]. However, a research by Centre of Independent Journalism [9] found out that the main sources of media and information are telephone (26%), social media (19%), word of mouth (18%), television (13%), and newspaper (7%). Nevertheless, TV remained as the most trusted source of information which stands at 38% followed by radio (24%). TV is perceived as the most credible media [10]. TV news still holds higher credibility compared to social media news as it provides better means of engagement with its audience [11]. The same research also claimed there is a small yet positive relationship between perceived TV credibility and TV news consumption. According to YouGov-Min Millennial from India's premium business news publications survey [12] predicted that TV news is dead as youth today are using online news instead of TV news, and only a few are using TV news to get their news. Genner and Suss [13] indicated that more and more people today are choosing to use online media to receive news because of the speed of distribution. This is also supported by a survey conducted by Statista [14] that more than 70% respondents from Malaysia cited social media as their source of news. Another study also showed that most of the users chose to use social media to get online news due to its easy availability and accessibility [15]. It is notable that watching TV for news also happens simultaneously with other media such as social media [11].

2.2 *TV News During and After Pandemic*

The media play a crucial role in facilitating transparent and clear communication between the government, health-care facilities, frontline workers, and the public to raise awareness within the community [1]. During previous pandemics, such as the H1N1 outbreak, there was a lack of sufficient information. From 2016–2017, TV news consumption has dropped 7% while people getting news from social media has shown an increment of 5% [16]. However, the Covid-19 pandemic has revitalized the TV industry with the shift in news viewing pattern back to television [17]. Due to the implementation of Movement Control Order, the average time that Malaysian turn on TV has increased more than one hour [14]. Mohamad et al. [18] claimed that the MCO has made the Malaysian to be highly dependent on the official information and local updates regarding the Covid-19. According to Nielsen Television Audience Measurement [4], there is a 30% increase in TV viewing during the MCO period. This is equivalent to about one million Malaysian viewers watching TV per average minute during the MCO period [19]. The report of Statista showed that, 2014–2019 Malaysian average time spent watching TV news was about 3 h 52 min [2]. Yet,

during the Covid-19, it showed an average time spent watching was 7 h and 8 min. Obviously this shows a clear increment of TV viewing time since the implementation of MCO in Malaysia on March 18, 2020. TV channels act as a bridge between the government and general public in dissemination of Covid-19-related information to avoid confusion [19]. TV channels have been trying to educate the wide audiences about the ongoing pandemic situation where daily updates and special reports, breaking news coverage, and announcements by officials or health experts can be seen on everyday TV news' content. Public had a high demand for credible news in the Covid-19 period to avoid trusting in inaccurate information [20]. The presence of gatekeepers in TV news has increased the news' credibility and accuracy of information as professional journalists would filter the news content and only broadcast news which are credible and educational to the audiences [21]. According to Samani et al. [22], mass media is a necessity that is a medium to obtain various information. Research in the United States also shows that Network TV and Cable TV have gained the highest consumption and trust by the citizens during the pandemic [23].

3 Theoretical Framework

3.1 *Uses and Gratification Theory*

Uses and gratification theory is a principle of media usage from mass communication studies that directs the evaluation of user motivation for media use and access [24]. This theory suggests that individuals use media to fulfill specific desires and needs. These needs include time passing, force of habit, escapism and to obtain information [25]. In this study, the needs of the public are examined to determine the relationship between the usage of the media and the level of satisfaction of them. As stated by Fitzgerald [26], studies show that traditional TV viewers in the United States have increased by a whopping 8.3 million people due to the pandemic. At the same time, Rochyadi-Reetz, Maryani, and Agustina [27] also show that India's media use and gratification during the pandemic is dominated by private TV, with public TV being the 5th most used platform to gain information during the pandemic. Pahayahay and Khalili-Mahani [28] found out that more than 60% of people used TV as a distraction from COVID-19 as they are in denial. Increased TV viewing is associated with the level of stress individuals found from the pandemic. As a result, the users choose the medium and its usage to satisfy their diverse needs.

4 Methodology

Online descriptive survey was conducted with a sample of 308 respondents from Kuching, Sibul, Miri, Limbang, and Marudi using snowball sampling where a few starting points were identified to keep rolling in order to get the targeted sample. The data were collected from December 2022–March 2023 via online google survey form. The survey questions were divided into 4 sections, such as personal information, TV news watching behavior, and patterns during Covid-19 pandemic, TV news watching behavior and Patterns in the post Covid-19 pandemic and TV news credibility. The respondents were asked to report their gender, age, and their race with the response options of Malays, Chinese, Bumiputera or Indians. The design of the survey questionnaire was generally to gauge the changes of TV news viewing behavior and patterns during and after Covid-19 pandemic in Sarawak. A pilot study was conducted prior to the actual data collection stage in which 29 pilot samples were selected. Overall, result shows the Cronbach's Alpha of 0.926, and therefore, it is deemed reliable to proceed with actual study. The collected survey data were utilized via univariate analysis and bivariate analysis.

5 Result

Table 1 presents the respondents' profile and a total of 306 respondents were received. The biggest majority of the respondents were aged between 25–64 years (59.5%), followed by 15–24 years (37.3%) and 65 years old and above (3.3%). Female respondent makeup 68% of the sample whereas males make up the remaining 32% of the sample. Majority of the respondents were Chinese with 74.5%, followed by Bumiputera with 17% and Malays with 8.5%. Apparently most of the respondents started their TV news watching habit when they have nothing to do (52.3%), followed by since they were young (41.8%), when they attended school (19.6%), and when they were forced to (7.8%). Generally, respondents prefer to watch national news (81.7%) over international news. The finding challenged the conclusion drawn in Ezhar Tamam [8] which suggested that local news consumption among youth is somewhat low, without a clear explanation. The results indicate that majority of the respondents exhibit a greater inclination to watch national news rather than international news. This preference become particularly evident during the Covid-19 pandemic, as it allows them to stay informed about developments impacting their lives. Previous research shows that respondents watch TV news via RTM and Media Prima channels mainly to focus on knowledge related to government information, especially current developments in the country and abroad [29].

During Covid-19 and post Covid-19, a significant portion of the population occasionally watches TV news. The findings do not reveal a substantial variation in TV news viewing habits between these two periods. The time duration on TV news watching is kept below 30 min per day during and post Covid-19. The majority

Table 1 The profiles of the respondents

Gender	Frequency	Percent
Female	208	68
Male	98	32
Total	306	100
<i>Race</i>		
Bumiputera	52	17
Chinese	228	74.5
Malays	26	8.5
Total	306	100
<i>Age</i>		
15–24	114	37.3
25–64	182	59.4
65 and above	10	3.3
Total	306	100

Source Online survey conducted from December 22, 2022 to March 21, 2023

of individuals allocate less than 15 min to watch TV news. Smartphones are the predominant device utilized for watching TV news both during and after Covid-19, with TV set follow closely behind. During the Covid-19 period, the favored channels for watching TV news include Astro, 8TV and TV3, while in the post Covid-19 era, Astro, TV3 and 8TV maintain their popularity. During and in the post Covid-19, people prefer watching TV news during night time, followed by afternoon then morning [29]. During Covid-19, people lean toward watching TV news on weekdays, while in the post Covid-19 era, their preference shift to weekends. They normally watch TV news at home be it during Covid-19 or in the post Covid-19 pandemic. The purpose of TV news watching during and in the post Covid-19 are getting the latest information, and keep themselves update with what is happening. The result shows social connection (28.8%) as one of the reason for TV news consumption [29]. As Thompson [30] argued, it is pertinent for broadcast stations to know their target audiences so that they can develop services that satisfy the audience’s need, tastes, and lifestyles. Overall, TV news remains significant, with the result indicating its importance at a level of 8 out of 10.

The result of the Pearson correlation coefficient shows TV news watching attention is positively weak correlated with the importance of TV as an information provider, education, keeping up with events around the world, and creating awareness. The finding indicates that those who pay more attention in the TV news watching acknowledge the importance of TV news as an information provider which includes Covid-19 related information. It also shows that those who pay more attention in the TV news watching paramount the importance of TV to educate them on how to interpret and understand the issues and matters surrounding them. Besides that, those who pay more attention in the TV news watching also rely on TV to keep them

updated with the events around the world, and create awareness on what is happening in the society. This implies the more attention they put on TV news, the more they acknowledge the importance of TV as the source of information. The result of the Pearson's correlation coefficient also indicates that the attraction to watch TV news via devices such as smartphones, laptop, Ipad, and computer is positively weakly correlated with the elements of unbiased, fair, can be trusted, accurate and tells the whole story. This implies that the unbiased, fair, can be trusted, accurate and tells the whole story might attract more people to watch TV news via other devices besides TV set such as how interesting the news are presented. The result of the Pearson correlation coefficient also indicates that the content of TV news via TV set is interesting and is positively moderately correlated with the credibility of TV news via TV set as it tells the whole story. This implies that people find the TV news to be more interesting as it has higher credibility (clearly presented), and at the same time tells the whole story to the viewer. This is compatible with the previous research that has demonstrated that TV is the most-trusted source of information [9]. This finding is important because it shows how to have more engagement of the viewers to the TV news.

6 Conclusion and Discussion

The findings appear to indicate that there were minimal alterations observed in TV news viewing habit during and following the Covid-19 pandemic in Sarawak. It seems that the most noticeable shift during and after the Covid-19 pandemic is the preferred time of viewing. During Covid-19 pandemic, 51.6% watched TV news during weekdays, while after Covid-19 pandemic, 52.9% watched TV news during weekends. In addition, the preferred TV news viewing time is at night, followed by afternoon and morning. These changes are evidently a result of the flexibility and increased free time available during the weekdays (especially at night) during the Covid-19 pandemic due to Movement Control Order (MCO) measure, as opposed to the typical post-MCO period when people tend to watch TV news on weekends (particularly at night). Another issue worth pondering is people in Sarawak generally still watch TV news as indicated in the result as "I watch TV news sometimes". This has shown not much changes during and after the Covid-19 pandemic. This indicates that TV news still plays an important role in the life of people in Sarawak. This contradicts with the previous research indicating that respondents have never and rarely watched TV for a week [29]. As TV news is always kept short and precise due to its digitization, the TV news watching time is kept below 30 min where 43.1% spend less than 15 min on TV news viewing and 33.3% watch 16–30 min per day. This outcome suggests that TV news should be concise and to the point, considering that viewers typically allocate less than 30 min to their viewing. People prefer to watch TV news via smartphone, followed by TV set. Again, this suggests that TV news might benefit from having distinct versions tailored to suit both smartphones and TV sets in order to cater to its diverse viewership. The study finds out the popular

TV channels to view news are Astro (Astro AEC & Astro Awani), 8TV followed by TV3. Both public and private TV stations play distinct roles in delivering news and information to the public, offering various languages and news presentation. This finding is in line with Rochyadi-Reetz, Maryani, and Agustina [27] which shows that India's media use and gratification during the pandemic is dominated by private TV, with public TV being the 5th most used platform to gain information during the pandemic. One potential rationale might be the public's perception of the credibility of news provided by both public and private TV stations. It is worth noting that content plays a significant role in the choice of news source. News presented on TV set is often seen as clear and straightforward, whereas news presented through other devices is perceived as engaging and interesting. Meanwhile, people prefer to watch TV news mostly at home, followed by coffee shops and the workplace. In general, individuals watch TV news for various reasons, including staying informed with the latest information, acquiring knowledge, staying updated on current events, fostering social connection, and spending quality time with their families. All of these purposes are aimed at fulfilling their personal needs. This is consistent with Vinney that individuals use media to fulfill specific desires and needs which include time passing, force of habit, escapism, and to obtain information [25]. It is worth emphasizing that primary objectives of watching TV news are to obtain the most current information and to stay abreast of ongoing developments. This aligns with result of prior research, as exemplified by Korhan and Ersoy [24]. Nonetheless, one limitation needs to be acknowledged. The snowball sampling method could be insufficient in a multi-racial context, potentially introducing biases into the collected data. In view of this, ratio sampling could enhance the representativeness of this study and contribute to improving the generalizability of the findings. In addition to enhancing the result's quality, future research should also broaden the investigation into the reliability and dependence on TV news within society.

Acknowledgements This research project is sponsored by the Internal Grant of **TAR University of Management and Technology**, Kuala Lumpur, Malaysia. Thank you to all respondents who are involved in this research.

References

1. Karasneh RA, Sayer I, Al-Azzam KH, Alzoubi LK, Rababah SM, Muflih M (2020) Health literacy and health related behaviour: a community based cross sectional study from a developing country. *J Pharmaceut Health Serv Res*. <https://doi.org/10.1111/jphs.12370>
2. Hanadian Nurhayati-Wolff (2020) Average time watching TV in Malaysia from 2014 to 2019. viewed 15 February 2021. <http://www.statista.com/statistics/1051333/malaysia-average-time-watching-tv/>
3. Media Prima (2020) Media Prima AGM 2020. viewed 3 February 2021. http://www.insage.com.my/Upload/Docs/MEDIA/MPB%20AGM2020_FINAL.pdf

4. The Nielsen Company (2020) *Movement restrictions are driving tv viewing, but Malaysia's markers must adapt to take full advantage*. Viewed 4 Feb 2021. <http://www.nielsen.com/my/en/insights/article/2020/movement-restrictions-are-driving-tv-viewing-but-malysias-market-ers-must-adapt-to-take-full-advantage>
5. Castriota S, Delmastro M, Tonin M (2020) National or local? the demand for news in Italy during Covid-19. CESifo Working Papers. No 8699. 1–27
6. Gottfried J, Barthel M, Shearer E, Mitchell A (2016) The 2016 presidential campaign-a news event that's hard to miss. <http://www.journalism.org/2016/02/04/the-2016-presidential-campaign-a-news-event-that-s-hard-to-miss/>
7. Davood M, Abu Hassan M, Muhamad Sham SA (2009) News media credibility of the internet and television. *Euro J Soc Sci* 11(1)
8. Tamam E (2011) Consumption of local news in television and newspaper and national pride among Malaysian youth. *Pertanika J Tropica Agricult Sci* 19(1):71–80
9. CIJ (2021) Information ecosystem assessment (IEA) report. Sarawak, Malaysia. <https://cijmalaysia.net/wp-content/uploads/2021/06/FINAL-IEA-ENG-SWK.pdf>
10. Saodah W, Ezhar T, Suci EM (2011) Patterns of news media consumption and news discussion among youth: a test of agenda setting theory. *Global Media J Malaysia Ed* 1(1):1–31
11. Affendi AK, Shifa F, Sara C (2021) Audiences' perception and engagement with Malaysian public broadcaster, radio television Malaysia's (RTM) prime news. *Int J Res Eng Sci (IJRES)* 9(6):17–26
12. NL Team (2018) Young people don't care much for TV news, YouGov- Mint Millennial Survey
13. Genner S, Suss D (2017) Socialisation as media effect. *Int Encyclopedia Effects*. <https://doi.org/10.1002/9781118783764.wbieme0138>
14. Statista (2020) Malaysia: time watching TV average per person during COVID-19 2020
15. Zanuddin H, Shaid N (2021) Social Perceived Value on social media and online news portal: benefits to the aborigines women in Malaysia. *Int J Interact Mob Technol (IJIM)* 15(04). <https://doi.org/10.3991/IJIM.V15I04.20189>
16. Poynter (2018) New Pew study says local TV news viewing dropping fast. 15 Feb 2021. <http://www.poynter.org/ethics-trust/2018/new-pew-study-says-local-tv-news-viewing-dropping-fast-2>
17. Ding D, del Pozo Cruz B, Green MA, Bauman AE (2020) Is the COVID-19 lockdown nudging people to be more active: a big data analysis. *Br J Sports Med* 54(20):1183–1184
18. Mohamad E, Tham JS, Ayub SH, Hamzah MR, Hashim H, Azlan AA (2020) Relationship between COVID-19 information sources and attitudes in battling the pandemic among the Malaysian public: cross-sectional survey study. *J Med Int Res* 22(11)
19. Shalvee S, Sambhav S (2020) Role of Mass media and communication during pandemic Covid 19. *Int J Creat Res Thoughts (IJCRT)* 8:3786–3790
20. Mathur A (2020) Pandemic shows importance of credible news, analysts say. *Press Freedom*. https://www.voanews.com/a/press-freedom_pandemic-shows-importance-credible-news-analysts-say/6199918.html
21. Tutheridge G (2017) What is the role of gatekeeping journalist's in today's media environment? <https://medium.com/@gabrielletutheridge/what-is-the-role-of-gatekeeping-journalists-in-today-s-media-environment-2034a30ba850>
22. Samani MC, Maliki J, Rashid NA (2011) Literasi media:ke arah melahirkan pengguna media berpegetahuan. *Jurnal Penjajian Media Malaysia* 13(20):41–64
23. Casero-Ripolles A (2020) Impact of Covid-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak. *El Profesional de la informacion* 29(2):1–11
24. Korhan O, Ersoy M (2015) Usability and functionality factors of the social networks sites application users from the perspectives of users and gratification theory. *Portal Komunikacji Naukowej*. <https://doi.org/10.1007/s11135-015-0236-7>
25. Vinney C (2019) What is uses and gratification theory? definition and examples. 16 Feb 2021. <http://www.thoughtco.com/uses-and-gratifications-theory-4628333>

26. Fitzgerald T (2020) For the first time in almost 10 years, time watching tv is up. <https://www.forbes.com/sites/tonifitzgerald/2020/04/28/for-the-first-time-in-almost-10-years-time-watching-tv-is-up/?sh=230415f120a8>
27. Rochyadi-Reetz M, Maryani E, Agustina A (2020) Public's media use and gratification sought during corona virus outbreak in Indonesia: a national survey. *Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia* 5(4):1–3
28. Pahayahay A, Khalili-Mahani N (2020) What media helps, what media hurts: a mixed methods survey study of coping with covid-19 using the media repertoire framework and the appraisal theory of stress. *J Med Int Res* 22(8):20
29. Alan R, Hassan MS, Vikibg J, Osman MN, Lepun P, Kamarudin S (2021) Descriptive analysis: television uses among community in rural area, Sarawak. *Int J Acad Res Bus Soc Sci* 11(11):1258–1272
30. Thompson R (2010) *Writing for broadcast journalist*, 2nd edn. Routledge, New York

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Behaviors and Patterns of TV News Viewing in Malaysia During and After Covid-19 Pandemic



Tang Mui Joo and Chan Eang Teng

Abstract The demise of Television (TV) news before Covid-19 has been an attention when people have shifted to online platforms for online news. Covid-19 Pandemic has changed the demand for TV news. The media landscape in Malaysia has then been reengineered by the impact of it. Driven by the theory of Fear Appeal in the information searching about Covid-19 during Malaysian Movement Control Order (MCO), TV news viewing has increased. This research aims to study the behaviors and patterns of TV news consumption during Covid-19 Pandemic in Malaysia. It is also to study the significance of TV news to Malaysians during the Pandemic. This research targets to ascertain the change of trend in TV news viewing behaviors and patterns among Malaysians during and after the Pandemic. In-depth interviews are conducted with 10 interviewees from Northern, Central and Southern Sarawak, the largest state in Malaysia. Purposive sampling is applied where the interviewees are from the background of media practitioners, educators, and undergraduates. The interviews are conducted from end March to early April 2023. Written consent has been obtained and ethical clearance is gotten. It is found that the TV news viewing behaviors and patterns have changed starting from MCO. It is corresponding to the various statistics reviewed and to the theory of Fear Appeal in the updates of Covid-19. After MCO, Malaysians expect various contents, presentations, and promotional activities that can engage the audiences. It is also to cope and compete with the digitalization in the new phase of challenge.

Keywords TV news · Viewing behaviors and patterns · MCO · TV contents · The role of TV · Fear appeal theory · Sarawak · Malaysia

T. M. Joo (✉) · C. E. Teng

Tunku Abdul Rahman University of Management and Technology, 53300 Kuala Lumpur, Malaysia

e-mail: muijoo@hotmail.com

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_3

23

1 Introduction

TV has been gradually neglected due to the advancement of technology when new media have been booming up [1]. The use of TV news demised when people have started to opt for online platforms in the search for online news either from social media platforms or online news websites. Such usage allows preferred genre with audience sustainability. Eventually, the audience are less likely to switch to other platforms for information search [2].

The media landscape in Malaysia has been reengineered by the impact of the Covid-19 Pandemic. Covid-19 accelerates the speed of technology adoption [3]. Malaysians watch and listen to the news to be kept updated of Covid-19 news, for entertainment, learning, and shopping in online platforms. Somehow at this point of time, Malaysian's TV viewing pattern has changed due to Covid-19 Pandemic. From the report shown by Statista [4], the TV viewership has increased when the Malaysian government implemented the Movement Control Order (MCO) on March 18 2020. When Malaysians are to stay and work from home, the TV viewing behaviors has changed relatively. The TV viewership has risen with an average of 30% viewers after the implementation of MCO in Malaysia [5].

Observing from the increase of statistics in the overall TV viewership in Malaysia during MCO, this research focuses on the TV news and geographical area in the East Malaysia, namely Sarawak, the largest state in the country. The research purposes have been as below:

- (a) To study the behaviors and patterns of TV news consumption during Covid-19 pandemic in Malaysia.
- (b) To study the significance of TV news on Malaysians during the Pandemic.
- (c) To ascertain the change of trend in TV news viewing behaviors among Malaysians during and after the Pandemic.

To achieve the purposes of this research, this research paper further discusses and reviews the impact of the Pandemic toward Malaysian focusing on Malaysians in TV viewing behaviors and patterns, the roles of TV to Malaysians during the Pandemic, changes in Malaysian TV news viewership during and after the Pandemic and factors to the changes in Malaysian TV news viewership at the transition of the Pandemic. Fear appeal theory is used to support this research. For data collection, structured in-depth Interviews have been conducted with the interviewees from the background of education and media industry. Thematic analysis with inductive approach is used to determine the themes to be correlated with the behaviors and patterns of TV news viewership in Malaysia. Conclusion is drawn from the data collected based on the themes produced.

2 Literature Review

2.1 *Impact of the Covid-19 Pandemic Toward Malaysians TV Viewing Behaviors and Patterns*

Covid-19 has indirectly altered the patterns and behaviors of Malaysian in TV news viewing. When most of the countries implement lock down, most people are to stay at home and to work from home, the consumption of TV has relatively increased [6]. According to the statistics of Malaysian TV viewership [4], the average time of Malaysians spent on TV in 2020 has increased for more than an hour, from 5 h 36 min to 7 h 8 min. During the Pandemic, Badminton's Thomas and Uber Cup Finals, Tokyo 2020 Olympics, women's football league in England and most of the popular sports are to be stopped or postponed [7]. The TV viewers of sports channels has decreased and the viewable live sports advertisement dropped from 6% to less than 1% [8].

In Malaysia, the trend of media content during MCO is that, 51% has followed news closely every other hour on how to protect themselves against Covid-19. Other media content apart from news are 67% of movies; 62% of TV series or dramas; 35% of cooking shows; 28% of online streaming sites; and, 27% of documentaries. The information has marked a change of trend in the search of media content. People watched and listened to the news, to be kept informed of updates on the Covid-19 situation, entertainment, learning, and shopping [3].

Though digital platforms have also shown an increase in its usage, the change of trend in media content is remarkable when the TV news consumption rebounds with an increase. Looking at such a scenario in TV news consumption, this research is to study the behaviors and patterns of TV news consumption during Covid-19 pandemic in Malaysia.

2.2 *The Roles of TV to Malaysians During the Covid-19 Pandemic*

TV is perceived as the most credible mass media [9]. Due to the advancement of technology and global trends, media consumption over the Internet among Malaysians has shown an increase especially in 2017 and 2018, while consumption of radio has declined. Although the number of hours spent on viewing content through TV has remained relatively stable in the last five years, it is expected that Internet consumption will surpass TV in the next few years [10]. On the other hand, the demise of TV viewership is indicated in other countries like USA, where from 2016 to 2017, even when technology is not fully developed, TV news consumption has dropped 7% from 57 to 50% while people getting news from social media increase around 5% from 38 to 43% [11]. Whereas in UK, TV viewership has been dropping since 2017

to 2020, from 73 to 61%, it is expected to continue dropping to 49% in 2023 [12]. In following the global trends in media consumption, TV viewership in Malaysia should be demising following the footsteps of other countries after 2019.

The Pandemic has overthrown the expectation on TV in Malaysia. As of March 31, 2020, the share of medium viewers, those who watched TV between 8 to 16 h a day, increased to 61%, compared to 49% of the total viewers at the onset of the COVID-19 outbreak in Malaysia. There was a marked increase in TV viewership in Malaysia since the implementation of the Movement Control Order (MCO) on March 18, 2020 [4]. The roles of TV have since then diversified into other areas of media content. While the rise in viewing is expected as Malaysians tune in to the TV to stay informed as well as to stave off boredom, the extent of the increase was extraordinary as it surpassed TV viewing levels seen before the recent digital revolution. Malaysians are not just tuning in to the news to get their information but are also watching a lot of reality, cooking, and female-orientated shows during the MCO [13].

The contradiction to expectation on TV viewership reflected in statistics and different genre of media content required during Pandemic have shown a change of behaviors and patterns in TV not only on news but media content as well. This research is here to study the significance of TV news on Malaysians during the Pandemic.

2.3 Factors to the Changes in Malaysian TV News Viewership During the Pandemic

During the Covid-19 Pandemic, understanding the effects of location information has significant implications for public crisis management and health communication [14]. The public is having high demand for credible news in this Covid-19 Pandemic period to avoid trusting in inaccurate information [15]. Media are to help the government and health care facilities to provide transparent and clear communication with front liners and the public to create the greatest awareness to the community [16].

The public has begun to rely on credible media outlets to understand and stay aware of what is going on around them in this Pandemic period which is full of uncertainties [17]. TV channels act as a bridge between the government and general public in dissemination of Covid-19-related information to avoid confusion [18]. TV news channels have also been trying to educate wide audiences about the on-going Pandemic situation where daily updates and special reports, breaking news coverage, and announcements by officials or health experts can be seen on everyday TV news' content [19].

In Malaysia, Ministry of Health was playing the role of Covid-19 information provider to the nation through mass media. Malaysia is one of the first countries to come out with various quick responses to protect its citizens from COVID-19. The main aim is to minimize economic and social impacts, limit its spread, and

provide care for its citizens. There are actions taken by the Malaysian media from mainstream to social media. COVID-19 has caused fear, anxiety, and confusion. The media, celebrities, and other influencers have appealed to the public to stay at home and avoid mass gatherings. The media has started to use the hashtag #stayhome. This hashtag has been used widely in the media, and it is hoped that important messages to stop the spread of Covid-19 can reach all levels of society [20].

In view to the changing behaviors and patterns of TV news viewership in Malaysia due to the various factors caused by Covid-19, this research intends to ascertain the change of trend in TV news viewing behaviors among Malaysians during and after the Pandemic.

2.4 Fear Appeal Theory in TV News Viewership

Fear appeals are commonly used in many types of marketing communication. The fundamental idea is where, if you do not do this, either to buy, to vote, to believe, to support or to learn, some particular consequences will occur [21]. Fear appeals can be direct or indirect. A direct fear appeal focuses on the welfare of the message recipient. An indirect fear appeal focuses on motivating people to help others in danger [22]. Whether direct or indirect, there are three additional factors contributing to success. The additional factors are that to design ads which motivate changes in individual behavior, to distribute the ads to the appropriate target audience, and to use a sustained communication effort to bring about change [23]. The theory indicates that threat information can influence individuals' threat appraisal and emotion like fear and anxiety [24]. Further to this view, it is also indicated that emotions can have an impact on coping appraisals, for instance, motivation to obtain information and to pay attention to information [25].

In Italy, TV has become the main source of information during Covid-19 Pandemic. Due to the people's depression, stress and anxiety over the Pandemic, the number of hours spent in watching TV to obtain information has also increased [26].

This theory is used to support this research in the view of Covid-19 Pandemic as a factor in changing and further increasing Malaysians' TV news viewing behaviors and patterns. This theory will be the base to ascertain the change in behaviors and patterns of TV news viewing, referring to the condition of the urge for information through TV news after the Pandemic.

3 Methods

In the approach to data collection, the researchers use Qualitative method, open-ended questions and in-depth interviews with 10 purposive samples from the background of media practitioners who are staff of TV stations and reporters, educators and undergraduates. TV stations covered are TVS, a TV station from the region of Sarawak, and RTM, namely Radio Television Malaysia. Whereas reporters invited for the interviews are from newspapers like Sin Chew Daily and See Hua Daily. Educators attended the interviews are public university lecturers and school teachers. Undergraduates are all from a public university in Sarawak. It is in the views that these groups of people are highly sensitive to information searching where news is included as part of the information. The respondents are selected from the biggest state in Malaysia, namely Sarawak, to participate in the interviews. Sarawak of Malaysia is selected as it is the largest state in the country, in terms of its land area, with its population of 2.74 million over the total population of 31.60 million [27]. The very much concern of this research's choice of sample is that, there are 46 percent of them are living in the rural areas. This research focuses at the respondents who are living in the urban areas where online and offline media accessibility is not an issue during the period of MCO.

There are 10 interviewees selected for the interviews in the three regions where bigger region like Southern Sarawak has 4 interviewees and two regions of Sarawak, Northern, and Central Sarawak have 3 interviewees each. The letters of invitation and emails are sent out prior to the interviews, together with the interview questions and consent forms. The respondents are adults who are living in the regions during and after MCO. Other than that, interviewees are also approached through WhatsApp in fixing the interviews. The interview appointments are all fixed and conducted during the period of researchers visiting the regions between end of March to early April of 2023. The whole interview processes are audio recorded.

The research has received full ethical approval from ethics committee of Tunku Abdul Rahman University of Management and Technology (TARUMT). All interviewees provide written consent to the interviews. The interviews occur in their work places or institutions on a prearranged and mutually agreed day and time. The interviews have been carried out in a various day between end of March to early April of 2023. The interview questions are divided into two parts related to the behaviors and patterns of TV news viewing during and after MCO. The interview questions can be found in Table 1 under the column of IQ (Interview Question). The interviews will then be analyzed and themed.

The data collected is aligned to thematic analysis. Inductive method of coding is applied to allow the data to determine the themes [28] in this research. Therefore, the whole process may require the transcriptions of interview recordings, coding, theming, reporting, and eventually drawing conclusion to the research.

Table 1 Illustrative examples of transcript analysis based on interview questions (IQs) with theming applied (*Source* In-depth Interviews conducted from end March to early April 2023)

IQ	Illustrative Analysis / Direct quotes	Themes (T)	Themes related to whom?
IQ1. Where do you normally get news?	<ul style="list-style-type: none"> • Newspaper, TV • Online news, social media, newspapers, radio stations • TV, verified Twitter source, online news, social media 	T1: Social media, online, digitalization, phones T2: Media gets news from authorities, government, wire service,	<ul style="list-style-type: none"> • Staff from TV stations • Educators and Reporters • Undergraduates
IQ2. How do you find TV news in Malaysia? (Content, length, presentation, time, slots, etc....)	<ul style="list-style-type: none"> • “15 min is an ample time, a quick update, quite enough” • More to human interest • Effective with audio and visual • Short and informative. Comes on at just the right time 	correspondents and stringers T3: Keep it short T4: Variety of topics coverage like human interest, news for the people, community, localized and niche T5: Lively, creative, better presentation, short and informative	<ul style="list-style-type: none"> • Staff from TV stations • Reporters • Educators • Undergraduates
IQ3. What do you expect from TV news?	<ul style="list-style-type: none"> • “we can show their culture, their food, daily life in their village area, ... and opportunity for us to get content” • More engagement • The presentation can be more lively • More news locally and internationally, covering social issues 	T6: Off line and online engagement with the audience or public T7: TV news on current affairs is still relevant T8: Consume less TV news T9: On TV news more during MCO T10: Content is the key T11: Fact check and to educate the public T12: Better talent pool	<ul style="list-style-type: none"> • Staff from TV stations • Reporters • Educators • Undergraduates

(continued)

Table 1 (continued)

IQ	Illustrative Analysis / Direct quotes	Themes (T)	Themes related to whom?
<p>IQ4. During MCO, has your TV news viewing habit changed? You started to watch more or watch less of TV news?</p>	<ul style="list-style-type: none"> • “I watched TV the most, to follow current issues and announcement from government at that time” • “During MCO, we consume through digital, our devices” • “During MCO I have to spend time at home, so I have more time to watch TV news as compared to before MCO” • “I am more interested in watching TV news during MCO period to keep myself up to date with the MCO policies” 		<ul style="list-style-type: none"> • Staff from TV stations • Reporters • Educators • Undergraduates
<p>IQ5. Are you still watching TV news? If yes, what kind of information do you focus on?</p>	<ul style="list-style-type: none"> • “I am looking at what the current government can do for us” • “Yes, politics, international news, shopping and livelihood issues” • Yes, current issues • “No, I watch less TV news. The TV news resume back to its old patterns in which it concentrates on politics, less social issues and international issues” 		<ul style="list-style-type: none"> • Staff from TV stations • Reporters • Educators • Undergraduates

(continued)

Table 1 (continued)

IQ	Illustrative Analysis / Direct quotes	Themes (T)	Themes related to whom?
IQ6. Apart from Covid-19-related news, will you continue watching TV news to get update info in future?	<ul style="list-style-type: none"> • “I will still follow the current issues” • Yes • “Depends on the information covers and provided by the tv news. Depending on its focus and agenda-setting” 		<ul style="list-style-type: none"> • Staff from TV stations • Reporters and Educators • Undergraduates
IQ7. Other than TV news, do you get news from other media?	<ul style="list-style-type: none"> • “Yes, we actually have a few synergies, the local newspapers, we have radio and social media platforms that we take note, other than our wires” • Yes, from social media and print media • “I do, I usually get information and updates from the online media platforms” 		<ul style="list-style-type: none"> • Staff from TV stations • Reporters and Educators • Undergraduates
IQ8. In your opinion, how to engage more viewers to watch TV news?	<ul style="list-style-type: none"> • “We do boost about our news segment frequently through our social media “ • “News is not necessarily to be in visual and audio, it can be infographics, easy” • “The presentation of news has to be more creative and with more innovation to attract audience’s attention” 		<ul style="list-style-type: none"> • Staff from TV stations and Reporters • Staff from TV stations • Educators and Undergraduates

(continued)

Table 1 (continued)

IQ	Illustrative Analysis / Direct quotes	Themes (T)	Themes related to whom?
IQ9. What are the challenges of TV news now?	<ul style="list-style-type: none"> • “I think we need local talents who meet the industry’s standard” • “Fact check the news to ensure the accuracy of it while being the first to break the news, and the competition in all aspects is getting intense” • Pattern and way of news presentation need changes. News readers shall be more energetic to make news more impactful 		<ul style="list-style-type: none"> • Staff from TV stations • Reporters • Educators and Undergraduates
IQ10. What are your suggestions to make TV news to continue surviving in the digital age?	<ul style="list-style-type: none"> • It is not about audience watching news at home, it is about the content • Keep up with the times to make necessary changes, be creative and incorporate some unique elements that others don’t have • Broaden their coverage on issues, include social issues such as crimes and more, less advertisements, and more international news 		<ul style="list-style-type: none"> • Staff from TV stations • Reporters and Educators • Educators and Undergraduates

4 Results

The data collected from the in-depth interviews has been transcribed, analyzed, and summarized into Table 1. Table 1 contains transcript analysis with some direct quotes based on the interview questions. Theming process performed after the illustrative analysis, and they are to be related to the interviewees from the background of media practitioners, educators and undergraduates.

This thematic analysis uses the method of inductive in the coding which allows the data to determine the themes. There are no predetermined themes expected. There are data of 10 interviewees collected. Therefore, the frequency of themes that strengthen the themes has not been into the consideration for any emphasis. The analysis has come out with twelve themes from the interviews. The themes have been indicated with their sources and summarized into Table 1.

T1 and T2 refers to the source of news and perceptions on news during MCO. The interviewees are getting news from social media, mainly from an online environment. Whereas those media practitioners are getting news from wire services, correspondents, and stringers, which are mainly through online too. In terms of views and expectation on TV news contents during MCO, there comes to T3 to T7. It is good to keep the content short with more variety of topics other than the updates of Covid-19. It is expected to have lively, creative, better presentation, short, and informative TV news during MCO. It is also expected to have more engagement with the audiences. T8 and T9 shows TV news viewing behaviors and patterns during MCO. There are more interviewees watching TV news though some have remained online for the source of news. Some are into TV news and online news at the same time.

For T10 to T12, they are themed from the interview questions based on the TV news viewing after MCO as of the condition during the interviews. For T10, interviewees are still watching TV news on various contents but the viewing is reducing. There is one interviewee highlighted that she is watching less now as the TV news resumes back to its old patterns that it concentrates on political news, less social and international news. This is therefore themed as that content is the key to engage the audience. For all the interviewees, they will continue to follow TV news not only on Covid-19. Again, content is the key as one has indicated that the engagement is depending on what the information covers and is providing in the agenda setting.

In T11, other than TV news, the interviewees are also engaged with news from other platforms like radio, printed materials and online media which includes social media. In order to engage more audience to TV news, audience engagement can be boosted through social media, news presentations which are creative and innovative and provide various contents. The challenges of TV news now are themed to the better talent pool. For those media practitioners, the challenges are to make sure of the fact check and to produce news that can educate the public.

In order for TV news to continue surviving in the digital age, after MCO, interviewees suggest to have relevant contents to the audience. It is not about watching news at home now but it is more about keeping up with the changes. It is time to incorporate contents with some unique elements. Broaden the coverage on issues, less advertisements and more of international news is suggested by the media practitioners interviewed.

5 Conclusion

TV news viewing behaviors and patterns have changed starting from MCO. It is corresponding to the various statistics reviewed earlier that the number of viewing has increased during MCO. It is mainly for the updates of Covid-19 though some other contents are also the purposes and urge of viewing as people are staying at home as compared to those days where people are mobile to go around. With this change of TV news viewing behaviors and patterns, it is as similar to the drive of Fear Appeal Theory that emotions can have an impact on coping appraisals that it leads to information obtaining and to pay attention to information. Same like many other countries, this theory has reflected the scenario in Malaysia where TV has become the main source of information during Covid-19 Pandemic. TV news plays the role to minimize the emotions of fear and security during Covid-19 Pandemic.

As the whole world is now in the period of post pandemic, the scenario of increase in TV news viewing is again the concern of the media industry. In Malaysia, the updates of Covid-19 may not be much of the concern among audiences. People are expecting various contents than before. Broaden the coverage on issues, less advertisements and more of international news may be able to engage the audiences. In order to retain the audiences, innovation and creativity in news presentation shall be looked into. Energetic presentation may be more impactful too. It is also suggested that “news is not necessarily to be in visual and audio, it can be infographics, easy”.

Of all, promotional activities and publicity of TV news shall be on going. It is an on-going effort of media practitioners to boost their news segment through their social media. It is also suggested to have a better pool of talent to cope and compete with the digitalization. After MCO when lives have gone back normal, TV news viewing may not be in front of the TVs anymore. It is going online, to-go scenario allows people to watch TV news from a different platform. The design of news presentation, technical support and uptrend contents should come into the focus to cope with the digitalization. This is another phase of challenge we are facing, to be addressed and to be researched.

Acknowledgements This research project is sponsored by the Internal Grant of **TAR University of Management and Technology**, Kuala Lumpur, Malaysia. Thank you to all respondents who are involved in this research.

References

1. Page C (2015) How technology is changing the way we watch television. <https://www.itbriecase.net/how-technology-is-changing-the-way-we-watch-television>
2. Hardwick J, Stox P, Oh S (2022) How search engines work. <https://ahrefs.com/blog/how-do-search-engines-work/>
3. Malaysian Communications and Multimedia Commissions (2020) Industry performance report 2020. https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/HighRes-MCMC_12102021_spread.pdf

4. Statista (2023) Average time watching TV in Malaysia from 2014 to 2021. <https://www.statista.com/statistics/1051333/malaysia-average-time-watching-tv/>
5. Chalil M (2020) TV viewing in Malaysia increases 30pc during MCO. https://malaysia.news.yahoo.com/tv-viewing-malaysia-increases-30pc-090640096.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAJuXROXFmfNOpR8MLxSDYi5P4QjSPRf8efQvJ6l0-E-A2mUbhrGanK6WqS92MrQ28BwoDoUuEhrCQCCMR2LMMXKJ4V3GrSaZfJFNFXwQ0KMpduSyXLRUEbVjhbyHPTiteskrFLZJ-e2Z4YNomMhjnXVI
6. Nielsen (2020) COVID-19: Tracking the impact on media consumption. <https://www.nielsen.com/insights/2020/covid-19-tracking-the-impact-on-media-consumption/>
7. Evans AB, Blackwell J, Dolan P, Fahlén J, Hoekman R, Lenneis V, McNarry G, Smith M, Wilcock L (2020) Sport in the face of the COVID-19 pandemic: towards an agenda for research in the sociology of sport. *European J Sport Soc* 17(2):85–95. <https://doi.org/10.1080/16138171.2020.1765100>
8. Webster T (2020) How COVID-19 has changed TV viewing habits. <https://www.tvisioninsights.com/resources/how-covid-19-has-changed-tv-viewing-habits>
9. Saodah W, Ezhar T, Suci Elsa M (2010) Pattern of the news media consumption and news discussion among youth: a test of agenda setting theory. In: International communication and media conference (ICOME' 10), pp 18–20
10. Asia Video Industry Association (2019) Malaysia in view. Executive summary, Avia
11. Poynter (2018) New Pew study says local TV news viewing dropping fast. <https://www.poynter.org/ethics-trust/2018/new-pew-study-says-local-tv-news-viewing-dropping-fast-2/>
12. Lee P, Westcott K, Bohm K (2021) Traditional TV wanes: television is about to dip below half of all UK video viewing. Magazine article, Deloitte Insights
13. Syok (2020) Malaysians watch more TV during the MCO. <https://en.syok.my/lifestyle/malaysians-watch-more-tv-during-the-mco>
14. Wu G, Deng W, Liu B (2021) Using fear appeal theories to understand the effects of location information of patients on citizens during the COVID-19 pandemic. In: National library of medicine. <https://doi.org/10.1007/s12144-021-01953-8>
15. Mathur-Ashton A (2020) Pandemic shows importance of credible news, analysts say. <https://www.voanews.com/press-freedom/pandemic-shows-importance-credible-news-analysts-say>
16. Karasneh R, Al-Azzam S, Muflih S, Soudah O, Hawamdeh S, Khader Y (2020) Media's effect on shaping knowledge, awareness risk perceptions and communication practices of pandemic COVID-19 among pharmacists. In: National library of medicine. <https://doi.org/10.1016/j.sapharm.2020.04.027>
17. Muno M (2020) Opinion: coronavirus, the media and credibility. <https://www.dw.com/en/opinion-coronavirus-the-media-and-credibility/a-53172087>
18. Shalvee K, Sambhav S (2020) Role of mass media and communication during pandemic Covid' 19. *Int J Creat Res Thoughts (IJCRT)* 8:3786–3790
19. Karam J (2020) The role of radio and television during Covid-19 pandemic. https://www.itu.int/en/ITU-D/Study-Groups/2018-2021/Documents/events/2020/Webinar%20Q2-1July03/Webinar_3%20July%20_Role%20of%20broadcasting%20COVID_19_JK.pdf
20. Ain Umaira MS, Syafiqah Nur AS, Rathedevi T, Nor Kamariah N, Azmawani AR, Zamberi S, Aini I, Mohamed Thariq HS (2020) COVID-19 outbreak in Malaysia: actions taken by the Malaysian government. *Int J Infect Dis* 97:108–116. <https://doi.org/10.1016/j.ijid.2020.05.093>
21. Witte K, Allen M (2000) A meta-analysis of fear appeals: implications for effective public health campaigns. *Health Educ Behav* 27(5):591–615. <https://doi.org/10.1177/109019810002700506>
22. Williams KC (2012) Fear appeal theory. *Int J Econ Bus Res* 5(February):63–82
23. Abernethy AM, Wicks JL (1998) Television station acceptance of aids prevention psas and condom advertisements. *J Advert Res* 38(5):53–62
24. Floyd DL, Prentice-Dunn S, Rogers RW (2000) A meta-analysis of research on protection motivation theory. *J Appl Soc Psychol* 30(2):407–429. <https://doi.org/10.1111/j.1559-1816.2000.tb02323.x>

25. So J, Kuang K, Cho H (2016) Reexamining fear appeal models from cognitive appraisal theory and functional emotion theory perspectives. *Commun Monogr* 83(1):120–144. <https://doi.org/10.1080/03637751.2015.1044257>
26. Scopelliti M, Pacilli MG, Aquino A (2021) TV news and COVID-19: Media influence on healthy behavior in public spaces. *Int J Environ Res Public Health* 18(4):1879. <https://doi.org/10.3390/ijerph18041879>
27. Malaysian Aviation Commission (2023) Profile of East Malaysia. <https://www.mavcom.my/en/industry/public-service-obligations/profile-of-east-malaysia/>
28. Braum V, Clarke V (2006) Using thematic analysis in psychology. *Qual Res Psychol* 3(2):77–101. <https://doi.org/10.1191/1478088706qp063oa>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Comparative Evaluation of Anomaly Detection Methods for Fraud Detection in Online Credit Card Payments



Hugo Thimonier, Fabrice Popineau, Arpad Rimmel, Bich-Liên Doan, and Fabrice Daniel

Abstract This study explores the application of anomaly detection (AD) methods in imbalanced learning tasks, focusing on fraud detection using real online credit card payment data. We assess the performance of several recent AD methods and compare their effectiveness against standard supervised learning methods. Offering evidence of distribution shift within our dataset, we analyze its impact on the tested models' performances. Our findings reveal that LightGBM exhibits significantly superior performance across all evaluated metrics but suffers more from distribution shifts than AD methods. Furthermore, our investigation reveals that LightGBM also captures the majority of frauds detected by AD methods. This observation challenges the potential benefits of ensemble methods to combine supervised, and AD approaches to enhance performance. In summary, this research provides practical insights into the utility of these techniques in real-world scenarios, showing LightGBM's superiority in fraud detection while highlighting challenges related to distribution shifts.

Keywords Imbalanced learning · Anomaly detection · Fraud detection

1 Introduction

Detecting fraudulent behaviors has emerged as a critical problem that has garnered significant attention from practitioners and scholars alike. In sectors such as banking, frauds incurred an estimated annual cost of \$28.58 billion in 2021, as highlighted

H. Thimonier (✉) · F. Popineau · A. Rimmel · B.-L. Doan
CNRS, CentraleSupélec, Laboratoire Interdisciplinaire des Sciences du Numérique,
Université Paris-Saclay, 91190 Gif-sur-Yvette, France
e-mail: name.surname@liscn.fr

F. Daniel
LUSIS AI, Paris, France

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_4

by the Nilson Report 2021.¹ To address the challenge of identifying frauds within regular credit card payments, banks have increasingly turned to machine learning techniques, known for their effectiveness in many classification tasks, particularly with unstructured data.

Two critical features of fraud detection pose challenges to constructing effective and accurate classifiers: highly imbalanced classes and distribution shifts.

Imbalanced datasets arise due to the significant disparity in the number of genuine transactions compared to fraudulent ones, making it difficult for traditional classification algorithms to generalize accurately. Highly imbalanced datasets are a specific subset of imbalanced datasets in which the positive class represents less than 1% of the samples; this situation is also referred to as rarity [3]. Moreover, distribution shift occurs as fraudsters constantly adapt their strategies, causing a discrepancy between the training and testing data distributions, thereby hampering the performance of machine learning models.

Learning from imbalanced datasets is a critical topic with implications across various real-life applications. Extensive research has highlighted the consequences of imbalanced learning on canonical classifiers, revealing that most standard classifiers are ill-suited for imbalanced settings. For instance, the limitations of standard machine learning techniques when confronted with imbalanced datasets are examined comprehensively by Yanmin et al. [42]. This study also sheds light on the struggles faced by backpropagation algorithms in converging within imbalanced setups, as the dominant majority class can overwhelm the gradient vector used for weight updates in neural networks. In contrast, gradient boosted decision trees (GBDTs) are often considered more resilient to imbalanced settings due to their focus on particularly challenging examples [11], thus enabling them to prioritize the minority class more effectively. Highly imbalanced datasets are among the most challenging as research [21] has shown how learners display decreasing performance as imbalance becomes more severe.

In addition to imbalanced datasets, distribution shift is another crucial challenge in detecting fraud. Standard machine learning techniques perform well when the training and testing dataset distributions are similar, if not identical. However, domain shift occurs when the distribution of the test dataset deviates from the original training distribution and thus hinders standard classifiers' performance. Fraud detection involves an iterative game between fraudsters and banks. Fraudsters continually strive to produce increasingly inconspicuous fraudulent behaviors, while banks aim to detect frauds as accurately as possible while avoiding false negatives. This dynamic nature of fraud detection presents significant challenges for machine learning algorithms.

These characteristics of fraud detection underscore the need for methodologies capable of effectively handling distribution shift and highly imbalanced datasets. In response to these challenges, researchers have proposed using anomaly detection (AD) methods, which promise to exhibit robustness in case of both distribution shift and extreme class imbalances. Specifically, AD involves the identification of anomalies within a dataset by delineating deviations from a predefined notion of

¹ <https://nilsonreport.com/>.

normality. Anomaly detection methods typically characterize the normal distribution solely based on normal samples during training. Consequently, AD has been regarded as particularly well-suited for imbalanced and extremely imbalanced settings. By design, AD methods do not experience performance deterioration when faced with highly skewed class distributions, as the training process solely requires normal samples. Moreover, assuming only fraudulent behaviors change over time, AD models should be more robust to distribution shift than standard supervised approaches. Indeed, if the normal distribution is well characterized, they should always be able to exclude new types of anomalies.

In this work, we empirically investigated AD for fraud detection tasks, exploring their capabilities and limitations. By empirically evaluating various AD methods on a real-world dataset characterized by distribution shifts and extreme class imbalances, we aim to provide insights into the suitability and effectiveness of these techniques for addressing the challenges inherent to fraud detection. In addition to evaluating the performance of AD methods, we conduct a comparative analysis with gradient boosted decision trees (GBDTs), the prevalent choice for machine learning tasks on tabular data [16], to gauge the added value of AD approaches. We rely on the LightGBM implementation [23] and show that GBDT suffers significantly more from distribution shift than AD methods while displaying substantially better fraud detection performance than all tested AD methods. Our paper is structured as follows: in the next section, we discuss works related to our application; in Sect. 3, we discuss in detail the experiments conducted on our dataset; in Sect. 4, we present the obtained results; in Sect. 5, we discuss the results, and finally in Sect. 6, we conclude.

2 Related Works

Anomaly detection encompasses two types of algorithms: supervised and unsupervised. In the case of supervised AD, one disposes of the label and indirectly uses it in the training process by building a training set solely composed of samples belonging to the *normal* class.² Unsupervised AD methods involve situations where the label is unavailable, and anomalies must be directly identified within a dataset containing both *normal* samples and anomalies. While supervised approaches can be used in situations where the imbalance is too severe for standard supervised approaches to work, unsupervised approaches are usually confined to applications that consist in removing samples that may hinder another models' performance on a particular task, e.g., mislabeled samples or outliers. Since we dispose of labels in the context of fraud detection, we will focus on supervised anomaly detection.

Supervised anomaly detection methods differ from standard supervised approaches because labels are only used indirectly. Indeed, standard supervised approaches consist in training a classifier using a dataset

² Throughout this paper, the term *normal* relates to the notion of normality.

$$\mathcal{D}_{\text{train}} = \{(x_i, y_i) : x_i \in \mathcal{X}, y_i \in \mathcal{Y}\}_{i=1}^n, \quad (1)$$

using both sample features x_i and labels y_i . Moreover, $\mathcal{D}_{\text{train}}$ contains samples from each class in \mathcal{Y} . On the contrary, supervised anomaly detection methods use the label to build a training set solely composed of a single class, referred to as the *normal* class. In the case of binary classification, the *normal* class is the majority class, e.g., the legit payments in the case of fraud detection, and the training set can then be constructed as

$$\mathcal{D}_{\text{train}}^{AD} = \{x_i : y_i = 0\}_{i=1}^n. \quad (2)$$

In this anomaly detection framework, the overall goal is to characterize the normal distribution, $p(x | y = 0)$. In inference, this characterization is used to determine whether a sample belongs to the normal distribution or should be seen as an anomaly.

As a field of research, anomaly detection can be divided into several non-exhaustive categories: one-class classification (OCC), reconstruction-based methods, and self-supervised methods.

One-class classification In contrast to traditional machine learning classification problems, one-class classification (OCC) approaches aim to identify samples that do not belong to a specific class by characterizing the distribution of that class. These discriminative models learn a decision boundary using only samples from the designated *normal* class, thereby circumventing the direct estimation of the class distribution. During the inference phase, samples are classified as either belonging to the *normal* class or not, without making any assumptions about the *anomaly* class. One-class support vector machines (OCSVM) [35] and support vector data description (SVDD) [39] are popular OCC methods that rely on kernels to map the data space to a Hilbert space, where a decision boundary is learned. Recently, [33, 34] have introduced extensions to OCC methods that incorporate deep neural networks to alleviate the computational complexity associated with kernels. Other OCC tree-based approaches can be found in the OCC such as isolation forest (IForest) [27], extended isolation forest [19], Robust Random Cut Forest (RRCF) [18], and PID-Forest [13]. Other methods have relied on sample-sample dependencies to identify anomalies such as TracInAD [41] relying on influence measures or approaches based on k-nearest neighbors (KNN). In the latter, anomalies are identified by measuring the distance of each sample to its k-nearest neighbors [2, 32]: higher distance indicating abnormality.

Reconstruction-based methods Reconstruction-based anomaly detection methods rely on the assumption that different distributions generate normal samples and anomalies. Consequently, training a model to reconstruct samples from the *normal* distribution aims to achieve low reconstruction error for any sample belonging to this distribution. Conversely, anomalies that are believed to stem from a distinct distribution should exhibit significantly higher reconstruction errors.

One of the most prevalent shallow reconstruction-based anomaly detection methods employs principal component analysis (PCA) or Bayesian PCA [9, 20]. Autoen-

coders [10], regularized autoencoders like Variational Autoencoders (VAEs) [30] and memory-augmented deep autoencoders [12], have also been leveraged for anomaly detection. Recently, [24] proposed a novel methodology for anomaly detection using autoencoders that incorporate the hidden representations of the original and reconstructed samples. Instead of solely comparing the reconstructed sample and the original sample, the authors suggest comparing the hidden representations of both samples by passing the reconstructed sample through the autoencoder. In addition, recent approaches have explored attention-based architectures for reconstructing masked features of samples, as exemplified by NPT-AD [40]. Other related methods do not directly compute a reconstruction error and only focus on estimating either the entire *normal* distribution such as ECOD [26] or local *normal* distributions as proposed in local outlier factor (LOF) [6].

Self-supervised methods The literature also features self-supervised approaches employing pretext tasks for anomaly detection [4, 31, 38]. In GOAD [4], several affine transformations are applied to each sample in the training set, while a classifier is trained to predict the specific transformation applied to a transformed sample. During testing, since the classifier was exclusively trained on *normal* samples, it is expected to struggle in correctly predicting the transformation for anomaly samples. Similarly, [31] propose NeuTraL-AD, a contrastive framework in which they transform samples using neural mappings instead of affine transformations. The objective is to learn transformations that maintain similarities in a semantic space between transformed samples and their untransformed counterparts while different transformations are easily distinguishable. In inference, the contrastive loss utilized to optimize the parameters serves as the anomaly score. More recently, [36] introduced a self-supervised methodology for anomaly detection that maximizes the mutual information among the elements of a sample's features using contrastive learning. By maximizing the mutual information, the method effectively captures the underlying structure of normal samples and identifies deviations indicative of anomalies.

Supervised classification on tabular data Although deep-learning models have become ubiquitous for a broad range of tasks involving natural language processing (NLP) and computer vision (CV), applying these models to tabular data remains very challenging. Some recent methods [15, 22, 37] have shown promising results when applying deep learning models tailored for tabular data. However, in recent work [14, 16], authors discuss how neural networks tend to struggle with this data type in comparison with other methods based on gradient boosted decision trees (GBDTs). In most scenarios, approaches such as XGBoost [7] or LightGBM [23] have been shown to surpass deep learning algorithms. This type of approach remains the go-to method for practitioners due to its strong classification performance and its simplicity to train in comparison with deep methods. Moreover, GBDT models such as LightGBM and XGBoost are often considered particularly suited for imbalanced and extremely imbalanced setups since these models focus on particularly hard-to-classify samples [11], generally the minority class, and thus offer strong performance in comparison to other standard machine learning models.

3 Experiments and Datasets

3.1 Dataset

We dispose of a labeled dataset of online credit card payments made available by a major French bank. Our dataset contains 145 features describing raw characteristics of payments (e.g., amount, currency) as well as features that we computed, such as rolling sums or rolling means. Our dataset contains 480 million online transactions from the first day of 2018 until the last day of 2021. The dataset is comprised of two independent datasets merged, one from 2018 to 2019 and the other from 2020 to 2021. This dataset displays the characteristic of highly imbalanced classes³ discussed in Sect. 1 since the proportion of legit payments vastly outnumbers the proportion of frauds. We remove cards with less than 50 payments, cards for which the proportion of frauds exceeds 50%, and cards with too few payments since they would not have derived features (e.g., rolling means) with meaningful values and would risk hindering the models’ performance. Similarly, we also omit cards with too many frauds in their payment history since they would also be problematic as they might pollute the fraud distribution. Overall, this preprocessing reduces the dataset to 192 million payments. Most methods we wish to test would be intractable with such dataset size and require further dimension reduction. Thus, we restrict our analysis to two countries with 3 million payments (1.5% of total dataset size) and 20 million payments (10.3% of total dataset size), respectively. Moreover, restricting our analysis to one country at a time should help models learn since payment distributions likely differ between countries.

Distribution shift We argue that our datasets undergo a distribution shift and that the distributions of legit and fraudulent payments differ between 2018 and the end of 2021. For instance, as supported by Gu et al. [17] COVID-19 has caused online payment behaviors to change over time drastically. To further support our statement, we display in Fig. 1 the t-sne [28] and UMAP [29] representations for years 2018–2019 and 2020–2021 for both countries. We observe significant dissimilarities between datasets. For country A, we observe an entire subsample of payments made in 2020–2021 at the bottom left of the UMAP graph, which does not exist for the 2018–2019 period. Similarly, for the t-sne representation of country A, we observe a similar pattern with few data sample overlays between periods. Moreover, for country B, we also observe a very scarce overlay on the graph between periods, especially for the UMAP representation. To further investigate whether distribution shift is present in our datasets, we rely on the Optimal Transport Dataset Distance method (OTDD) [1] to measure the distance between datasets from each period. This method relies on optimal transport, which measures the distance between distributions. For each country, we created two subsamples of 5000 observations for each period and compared the distance between the subsample of the same period and between periods.

³ Due to confidentiality, we cannot discuss the exact proportion of frauds within our dataset.

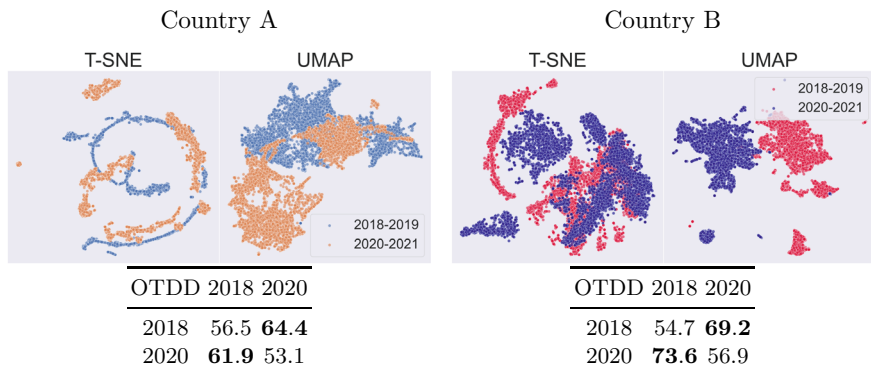


Fig. 1 T-sne [28] and UMAP [29] bi-dimensional representation of payments in countries A and B for each period. These graphs give evidence of a distribution shift for both countries' payment behaviors between 2018 and 2020 since we observe very few sample overlays. Tables give the Optimal Transport Dataset Distance [1] between subsamples for each country. We observe a much higher distance between subsamples from the same country between different periods than for the same period

Results of this analysis are shown in the tables in Fig. 1 and indicate that the distance between the dataset increases across periods. This is especially true for country B.

Data splits and preprocessing For a fair comparison between supervised approaches and anomaly detection methods, we split the 2018 datasets of each country into two separate datasets constituted of legit payments and frauds. We take a training set representing 75% of the 2018 dataset for each country and use a test set for the 25% remaining. We include the frauds in the training set for LightGBM, while for anomaly detection approaches, the frauds are excluded from the training set. The 2020 dataset of each country serves entirely as test sets. Overall, the considered dataset used for every tested model contains features describing characteristics of the payment (e.g., amount of the transaction, the currency used, duration since the last transaction), among which eight are categorical. All eight categorical features are encoded using CatBoost encoding following [5]. Continuous features are scaled to be in (0, 1) through standard normalizing by removing the mean and reducing to unit variance.

3.2 Experimental Settings

In order to evaluate the suitability of anomaly detection methods for fraud detection, we conduct a comprehensive analysis using state-of-the-art approaches on our dataset. We investigate both deep learning-based techniques designed for tabular data, such as the self-supervised approaches of GOAD [4], NeuTraL-AD [31] the contrastive approach proposed in [36], and the reconstruction-based approach of

NPT-AD [40]. Additionally, we include non-deep learning methods, namely Isolation Forest [27], ECOD [26], COPOD [25], and the KNN AD approach [2, 32], as they have demonstrated remarkable performance on various tabular datasets. As a baseline, we employ LightGBM [23], a well-established supervised classification method, to assess the added value of anomaly detection compared to standard supervised techniques in this context.

To effectively compare the performance of various models in detecting fraud, we employ three commonly used metrics from the anomaly detection literature and the banking industry for real-life model evaluation. The anomaly detection literature has widely adopted the F1-score and the AUROC as evaluation metrics. While the AUROC is suitable for balanced class distributions, it may not fully account for class proportions, which is crucial in assessing performance in imbalanced setups. We include the Area Under the Precision-Recall Curve (AUPRC) to address this limitation, which is better suited for imbalanced datasets. In support of this choice, [8] has demonstrated that a model dominates in the ROC space if and only if it also dominates in the PR space.

For the F1-score, whose value depends on a threshold, we adhere to the practices of the anomaly detection literature by selecting the threshold for the anomaly score that predicts an equal number of fraud cases as those present in the dataset. This approach ensures a fair and consistent evaluation across all models.

4 Results

4.1 Models Hyperparameters

We implemented the non-deep models using the PyOD library [43] with default hyperparameter values and LightGBM [23] also with default parameters. For the deep learning approaches, we adopted the hyperparameters suggested in the original papers for the dataset that most resembled our datasets. Specifically, we set the hyperparameters to those used for the KDD dataset for NeuTraL-AD [31], GOAD [4], and NPT-AD [40] using their official implementations available on GitHub. Regarding the approach proposed by Shenkar and Wolf [36], we kept the parameters at their default values as specified in their implementation. Deep models were trained on 4 Nvidia GPUs V100 16Go/32Go, while non-deep models were trained on 2 Intel Cascade Lake 6248 processors (20 cores at 2.5 GHz), thus 40 cores.

4.2 Fraud Detection Performances

Metrics reported in Table 1 are averaged over 10 runs and we performed t-tests on the highest metrics to assess whether models obtained significantly different results.

Among the AD approaches, we observe that the non-deep methods demonstrate the best performance, specifically ECOD, COPOD, and KNN. However, it is worth noting that the deep learning approach NPT-AD also yields comparable results. While the AD methods achieve satisfactory results regarding the AUROC, their performances are consistently poor for both the F1-score and AUPRC metrics across both countries and periods. In contrast, LightGBM exhibits significantly better performance across all metrics, consistently achieving the highest values, except for the F1-score on the 2020 dataset of Country B. The overall poor performance of all models, including LightGBM, for the F1-score and AUPRC, highlights the inherent challenges associated with fraud detection. However, a noteworthy observation is the substantial performance gap between LightGBM and the anomaly detection methods.

5 Discussion

5.1 *Distribution Shift*

The obtained results for both countries support the hypothesis that a distribution shift occurred between 2018 and 2020 in our dataset. Across most tested approaches, we observe a significant decrease in all metrics between the 2018 and 2020 datasets for both countries. Notably, the distribution shift appears more pronounced in country B, with metrics experiencing a more significant decline. The findings presented in Fig. 1 further corroborate this statement as the dataset distance is higher between periods than for country A, and the bi-dimensional representations also show less overlap between the periods than for country A. Although the AD methods display poor overall performance, they exhibit more resilience in the face of distribution shift than LightGBM. This trend is particularly evident for ECOD and COPOD, which maintained relatively similar metrics between the 2018 and 2020 datasets for both countries. Conversely, KNN and NPT-AD experienced a significant decline in performance across both periods and datasets compared to ECOD and COPOD. While LightGBM still achieves the highest metrics for most of the 2020 dataset, it suffers a substantial drop in performance between the two periods. This drop is especially pronounced in the AUPRC and F1-score metrics. Notably, most AD methods outperform LightGBM by a significant margin in terms of the F1-score for the 2020 dataset of country B.

Based on these findings on our dataset, it appears that LightGBM is a favorable choice in the fraud detection framework when labels are available and no distribution shift occurs. However, in the presence of a distribution shift, retraining LightGBM on an updated dataset becomes crucial to prevent a significant performance decline.

Table 1 Performance metrics of AD models in comparison with LightGBM [23]

Model	Country A						Country B					
	F1 (\uparrow)		AUROC (\uparrow)		AUPRC (\uparrow)		F1 (\uparrow)		AUROC (\uparrow)		AUPRC (\uparrow)	
	2018	2020	2018	2020	2018	2020	2018	2020	2018	2020	2018	2020
LightGBM	21.52	0.7	89.98	66.49	18.74	0.31	17.15	0.48	93.5	75.51	34.84	2.68
ECOD	1.6	0.33	0.41	2.38	1.78	0.05	1.93	0.39	0.31	1.24	1.48	0.29
	0.48	0.2	62.2	62.49	0.4	0.25	0.57	1.04	54.02	51.59	1.09	0.76
	0.2	0.22	0.64	1.02	0.02	0.01	0.26	0.38	0.54	0.87	0.06	0.04
COPOD	0.34	0.16	64.77	62.64	0.43	0.25	0.5	1.12	51.7	50.15	1.0	0.73
	0.21	0.15	0.51	0.98	0.02	0.01	0.2	0.44	0.59	0.91	0.05	0.04
Isolation forest	0.16	0.19	64.14	60.55	0.43	0.23	0.71	1.3	60.52	46.86	1.35	0.67
	0.12	0.19	0.75	0.76	0.02	0.01	0.23	0.34	0.52	0.84	0.07	0.03
KNN	0.34	0.01	68.87	55.6	0.51	0.18	0.78	0.38	65.92	49.28	1.58	0.63
	0.12	0.04	0.76	0.85	0.02	0.01	0.21	0.27	0.64	0.67	0.08	0.02
GOAD	0.14	0.19	53.72	52.73	0.17	0.17	0.7	0.69	50.36	64.45	0.67	<i>1.03</i>
	0.09	0.13	1.41	1.69	0.01	0.01	0.36	0.34	2.35	1.25	0.05	0.06
NeuTraL-AD	0.6	0.02	59.12	51.52	0.35	0.15	1.45	0.38	53.23	45.19	1.08	0.58
	0.22	0.08	3.56	1.15	0.05	0.01	0.44	0.21	1.9	1.75	0.07	0.03
Internal Cont.	0.64	0.0	39.43	46.7	0.18	0.13	1.21	0.23	45.63	50.66	0.87	0.68
	0.08	0.0	1.05	0.17	0.0	0.0	0.16	0.07	2.46	0.9	0.1	0.03
NPT-AD	<i>0.97</i>	0.66	67.21	53.2	<i>0.81</i>	0.18	<i>1.67</i>	0.58	<i>66.21</i>	53.45	1.28	0.61
	0.07	0.06	1.25	0.65	0.01	0.03	0.11	0.03	1.14	0.43	0.11	0.06

Results are averaged over 10 runs for 10 different splits of the data, and the standard deviation is displayed below the metrics. We report the F1-score in terms of percentage. The highest metric over all models is highlighted in bold, while the highest metrics among the AD method are italics. We perform 5% t-test between the highest metrics to measure whether they are statistically different

5.2 Anomaly Detection Methods for Ensembling

One potential advantage of anomaly detection (AD) methods is their ability to identify fraud cases that differ from those flagged by supervised approaches. If AD models can successfully detect fraud instances supervised models cannot identify, resorting to ensembling techniques could enhance overall fraud detection performance. To investigate this further, we focused on ECOD, one of the top-performing AD methods on the dataset consisting of payments in *country A*. We examined whether the fraud cases detected by ECOD differed from those identified by LightGBM. Across the 10 iterations, we observed that, on average, 3.28% of the fraud cases in the test set were detected by ECOD but not by LightGBM. Conversely, LightGBM detected 20.41% of the fraud cases in the test set that ECOD did not flag. Notably, the 3.28% represents 10.98% of the fraud cases detected by ECOD. In other words, 89.02% of the fraud cases detected by ECOD were also detected by LightGBM. As a result, ensembling models by combining AD methods with LightGBM to enhance fraud detection performance may prove ineffective.

6 Conclusion

In conclusion, our study highlights several key findings concerning applying machine learning techniques for fraud detection. Our results demonstrate that LightGBM consistently outperforms the tested AD methods across various evaluation metrics, emphasizing its efficacy in fraud detection tasks compared to other methods. However, we also observed that LightGBM's performances are susceptible to degradation due to distribution shift. This finding underscores the importance of retraining LightGBM on updated datasets when there is suspicion or evidence of a distribution shift. By adapting the model to the changing data distribution, it is possible to mitigate the drop in performance and maintain its effectiveness in fraud detection. Furthermore, our investigation revealed that ensembling techniques with AD methods would not significantly improve overall fraud detection performance. Despite the potential for AD methods to detect frauds that may elude supervised approaches, our analysis showed that LightGBM also detected most of the frauds identified by AD methods. This finding suggests limited benefits in combining AD methods with LightGBM in our specific fraud detection framework. We believe these insights may contribute to advancing the field of fraud detection and inform practitioners in selecting appropriate models and strategies for robust and accurate fraud detection systems.

Future work may involve replicating our analysis on other credit card payment datasets to determine whether our obtained results can be generalized. Furthermore, enhancing the robustness of GBDT models against distribution shifts emerges as a critical direction for further exploration. Addressing this challenge is paramount for financial institutions, as it enables them to embrace machine learning techniques in fraud detection systems confidently.

Acknowledgements This work was granted access to the HPC resources of IDRIS under the allocation 2023-101424 made by GENCI.

This research publication is supported by the Chair “Artificial intelligence applied to credit card fraud detection and automated trading” led by CentraleSupélec and sponsored by the LUSIS company.

References

1. Alvarez-Melis D, Fusi N (2020) Geometric dataset distances via optimal transport. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) Advances in Neural information processing systems 33: annual conference on neural information processing systems 2020, NeurIPS 2020, December 6–12, 2020, virtual. <https://proceedings.neurips.cc/paper/2020/hash/f52a7b2610fb4d3f74b4106fb80b233d-Abstract.html>
2. Angiulli F, Pizzuti C (2002) Fast outlier detection in high dimensional spaces. In: European conference on principles of data mining and knowledge discovery. Springer, pp 15–27 (2002)
3. Bauder RA, Khoshgoftaar TM, Hasanin T (2018) An empirical study on class rarity in big data. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA), pp 785–790. <https://doi.org/10.1109/ICMLA.2018.00125>
4. Bergman L, Hoshen Y (2020) Classification-based anomaly detection for general data. In: International conference on learning representations. https://openreview.net/forum?id=H1IK_IBtvS
5. Bourdonnaye FDL, Daniel F (2021) Evaluating categorical encoding methods on a real credit card fraud detection database. CoRR abs/2112.12024. <https://arxiv.org/abs/2112.12024>
6. Breunig MM, Kriegel H, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. In: Chen W, Naughton JF, Bernstein PA (eds) Proceedings of the 2000 ACM SIGMOD international conference on management of data, May 16–18, 2000. ACM, Dallas, Texas, USA, pp 93–104. <https://doi.org/10.1145/342009.335388>. <https://doi.org/10.1145/342009.335388>
7. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
8. Davis J, Goadrich M (2006) The relationship between precision-recall and roc curves. In: Proceedings of the 23rd international conference on machine learning, ICML '06. Association for Computing Machinery, New York, NY, USA, pp 233–240. <https://doi.org/10.1145/1143844.1143874>
9. Dutta H, Giannella C, Borne KD, Kargupta H (2007) Distributed top-k outlier detection from astronomy catalogs using the DEMAC system. In: Proceedings of the seventh SIAM International conference on data mining, April 26–28, 2007. SIAM, Minneapolis, Minnesota, USA, pp 473–478. <https://doi.org/10.1137/1.9781611972771.47>
10. Finke T, Krämer M, Morandini A, Mück A, Oleksiyuk I (2021) Autoencoders for unsupervised anomaly detection in high energy physics. J High Energy Phys 2021(6). [https://doi.org/10.1007/jhep06\(2021\)161](https://doi.org/10.1007/jhep06(2021)161). [http://dx.doi.org/10.1007/JHEP06\(2021\)161](http://dx.doi.org/10.1007/JHEP06(2021)161)
11. Frery J (2019) Ensemble learning for extremely imbalanced data flows. Theses, Université de Lyo. <https://tel.archives-ouvertes.fr/tel-02899943>
12. Gong D, Liu L, Le V, Saha B, Mansour MR, Venkatesh S, van den Hengel A (2019) Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: 2019 IEEE/CVF international conference on computer vision (ICCV), pp 1705–1714
13. Gopalan P, Sharan V, Wieder U (2019) Pidforest: anomaly detection and certification via partial identification. In: Neural information processing systems. <https://api.semanticscholar.org/CorpusID:202766416>

14. Gorishniy Y, Rubachev I, Kartashev N, Shlenskii D, Kotelnikov A, Babenko A (2023) Tabr: unlocking the power of retrieval-augmented tabular deep learning
15. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A (2021) Revisiting deep learning models for tabular data. In: Beygelzimer A, Dauphin Y, Liang P, Vaughan JW (eds) *Advances in neural information processing systems*. https://openreview.net/forum?id=i_Q1yrOegLY
16. Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on typical tabular data? In: *Thirty-sixth conference on neural information processing systems datasets and benchmarks track*. https://openreview.net/forum?id=Fp7__phQszn
17. Gu S, Ślusarczyk B, Hajizada S, Kovalyova I, Sakhbiava A (2021) Impact of the COVID-19 pandemic on online consumer purchasing behavior. *J Theoret Appl Electron Commerce Res* 16(6):2263–2281
18. Guha S, Mishra N, Roy G, Schrijvers O (2016) Robust random cut forest based anomaly detection on streams. In: *ICML*. <https://www.amazon.science/publications/robust-random-cut-forest-based-anomaly-detection-on-streams>
19. Hariiri S, Kind MC, Brunner RJ (2021) Extended isolation forest. *IEEE Trans Knowl Data Eng* 33(4):1479–1489. <https://doi.org/10.1109/TKDE.2019.2947676>
20. Huang L, Nguyen X, Garofalakis MN, Jordan MI, Joseph AD, Taft N (2006) In-network PCA and anomaly detection. In: Schölkopf B, Platt JC, Hofmann T (eds) *Advances in neural information processing systems 19, Proceedings of the twentieth annual conference on neural information processing systems, December 4–7*. MIT Press, Vancouver, British Columbia, Canada, pp 617–624. <https://proceedings.neurips.cc/paper/2006/hash/2227d753dc18505031869d44673728e2-Abstract.html>
21. Japkowicz N, Stephen S (2002) The class imbalance problem: a systematic study. *Intell Data Anal* 6(5):429–449
22. Kadra A, Lindauer M, Hutter F, Grabocka J (2021) Well-tuned simple nets excel on tabular datasets. In: *Thirty-fifth conference on neural information processing systems*
23. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY (2017) Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 30:3146–3154
24. Kim KH, Shim S, Lim Y, Jeon J, Choi J, Kim B, Yoon AS (2020) Rapp: novelty detection with reconstruction along projection pathway. In: *International conference on learning representations*. <https://openreview.net/forum?id=HkgeGeBYDB>
25. Li Z, Zhao Y, Botta N, Ionescu C, Hu X (2020) Copod: Copula-based outlier detection. In: *2020 IEEE international conference on data mining (ICDM)*. <https://doi.org/10.1109/icdm50108.2020.00135>. <http://dx.doi.org/10.1109/ICDM50108.2020.00135>
26. Li Z, Zhao Y, Hu X, Botta N, Ionescu C, Chen G (2022) Ecod: unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Trans Knowl Data Eng*:1–1. <https://doi.org/10.1109/TKDE.2022.3159580>
27. Liu FT, Ting KM, Zhou ZH (2008) Isolation forest. In: *2008 Eighth IEEE international conference on data mining*, pp 413–422. <https://doi.org/10.1109/ICDM.2008.17>
28. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605. <http://www.jmlr.org/papers/v9/vandermaaten08a.html>
29. McInnes L, Healy J, Melville J (2020) Umap: uniform manifold approximation and projection for dimension reduction
30. Pol AA, Berger V, Cerminara G, Germain C, Pierini M (2019) Anomaly detection with conditional variational autoencoders. In: *ICMLA 2019—18th IEEE international conference on machine learning and applications, 18th international conference on machine learning applications*. Boca Raton, United States. <https://hal.inria.fr/hal-02396279>
31. Qiu C, Pfrommer T, Kloft M, Mandt S, Rudolph M (2021) Neural transformation learning for deep anomaly detection beyond images. In: *International conference on machine learning*. PMLR, pp 8703–8714
32. Ramaswamy S, Rastogi R, Shim K (2000) Efficient algorithms for mining outliers from large data sets. In: *ACM sigmod record*, vol 29. ACM, pp 427–438

33. Ruff L, Vandermeulen R, Goernitz N, Deecke L, Siddiqui SA, Binder A, Müller E, Kloft M (2018) Deep one-class classification. In: Dy J, Krause A (eds) Proceedings of the 35th international conference on machine learning, Proceedings of machine learning research, vol 80. PMLR, pp 4393–4402. <https://proceedings.mlr.press/v80/ruff18a.html>
34. Ruff L, Vandermeulen RA, Görnitz N, Binder A, Müller E, Müller KR, Kloft M (2020) Deep semi-supervised anomaly detection. In: International Conference on learning representations. <https://openreview.net/forum?id=HkgH0TEYwH>
35. Schölkopf B, Williamson R, Smola A, Shawe-Taylor J, Platt J (1999) Support vector method for novelty detection. In: Proceedings of the 12th international conference on neural information processing systems, NIPS'99. MIT Press, Cambridge, MA, USA, pp 582–588
36. Shenkar T, Wolf L (2022) Anomaly detection for tabular data with internal contrastive learning. In: International conference on learning representations. https://openreview.net/forum?id=_hszZbt46bT
37. Somepalli G, Goldblum M, Schwarzschild A, Brass CB, Goldstein T (2021) SAINT: improved neural networks for tabular data via row attention and contrastive pre-training. CoRR abs/2106.01342. <https://arxiv.org/abs/2106.01342>
38. Tack J, Mo S, Jeong J, Shin J (2020) Csi: Novelty detection via contrastive learning on distributionally shifted instances. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) NeurIPS. <https://proceedings.neurips.cc/paper/2020>
39. Tax D, Duin R (2004) Support vector data description. Machine Learn 54:45–66. <https://doi.org/10.1023/B:MACH.0000008084.60811.49>
40. Thimonier H, Popineau F, Rimmel A, Doan BL (2023) Beyond individual input for deep anomaly detection on tabular data. <https://arxiv.org/abs/2305.15121>
41. Thimonier H, Popineau F, Rimmel A, Doan BL, Daniel F (2022) TraInAD: measuring influence for anomaly detection. In: 2022 International joint conference on neural networks (IJCNN), pp 1–6. <https://doi.org/10.1109/IJCNN55064.2022.9892058>
42. Yanmin S, Wong A, Kamel MS (2011) Classification of imbalanced data: a review. Int J Pattern Recogn Artif Intell 23:687–719. <https://doi.org/10.1142/S0218001409007326>
43. Zhao Y, Nasrullah Z, Li Z (2019) Pyod: a python toolbox for scalable outlier detection. J Mach Learn Res 20(96):1–7. <http://jmlr.org/papers/v20/19-011.html>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Exploratory Customer Discovery Through Simulation Using ChatGPT and Prompt Engineering



Joseph Benjamin Ilagan , Zachary Matthew Alabastro ,
Claire Louise Basallo , and Jose Ramon Ilagan 

Abstract Entrepreneurship and the tech startup journey are complex, dynamic, risky, and uncertain. Risk-taking needs to consider the complexity and interconnection of different aspects of the entrepreneurial and startup context. A simulation is a model of an existing complex system and experimenting with the model to understand the whole system's behavior. Computer simulations have been widely used to study complex environments such as entrepreneurship. Large language models (LLMs) such as ChatGPT, by nature of their training and design, are models of humans and likely possess latent social information. As such, LLMs could extend their usefulness from mostly being assistants to simulators of human behavior. Technology startups iteratively manage risks involving the uncertainty of their business models through Lean Startup Approaches (LSAs) combined with *customer development*. The first step in customer development involves *customer discovery*, where startup co-founders start with a vision and a set of assumptions (or “hypotheses”) about their business model and seek feedback from their prospective customers. This study explores using ChatGPT as a simulation tool for customer development for technology startups. The validation of the simulator involves coming up with baseline behavior and feedback without prompt preparations, followed by preparing synthetic prospective customers as agents in the virtual environment. This involves *endowment* of demographic characteristics and getting behavior and feedback again afterward.

WWW home page: <https://www.ateneo.edu/>.

J. B. Ilagan (✉) · Z. M. Alabastro · C. L. Basallo · J. R. Ilagan
Ateneo de Manila University, Quezon City, Philippines
e-mail: jrilagan@ateneo.edu

Z. M. Alabastro
e-mail: zachary.alabastro@obf.ateneo.edu

C. L. Basallo
e-mail: claire.basallo@obf.ateneo.edu

J. R. Ilagan
e-mail: jbilagan@ateneo.edu

Keywords Computer simulation · Simulation · Large language models · LLM · ChatGPT · GPT · Startup · Customer development · Lean startup · Agent-based modeling · Agent-based simulation

1 Introduction

1.1 Context

Entrepreneurship pursues “novel or better products or business models amidst constraints” [1, 2]. In pursuing novelty with limited resources, the entrepreneur has to manage four risks: demand, technology, execution, and financing [1]. The entrepreneurial context or system involves a decision environment that is complex, dynamic, and inherently uncertain [3]. A simulation can help design a model of an existing complex system found in entrepreneurial scenarios and experiment with the model to understand the system’s behavior as a whole or to evaluate strategies to operate it [4]. Simulations are low-cost alternatives to setting up real ventures for learning purposes [5], and they provide dynamic testbeds for conducting experiments and opportunities for analyzing scenarios [6] or situations otherwise involving legal, moral, or privacy concerns [7]. Despite the models’ simple and limited theoretical grounding, simulations allow elaboration and exploration through experimentation [8]. Agent-based Modeling and Simulation (ABMS) covers interactions among independent agents [9]. ABMs are helpful when it is essential to model complex and dynamic interaction with other agents [9], which is the case in entrepreneurial and startup contexts. However, one limitation with ABMSs is that you must program the agents as both “judge and jury” and then see what they do [10].

Generative artificial intelligence (GAI) is a machine learning framework that generates content using probability and statistics derived through training from existing digital content (e.g., text, video, images, and audio) [11]. A large language model (LLM) is a GAI based on a mathematical model of the statistical distribution of tokens in the large body of human-generated text publicly available [12]. GPT uses the *transformer* [13] architecture for deep neural networks. Transformer-based models train large amounts of publicly available data in parallel. From the training, LLMs can produce human-like language [14].

Prompts are instructions to an LLM to ensure specific qualities (and quantities) of generated output through enforcing rules and automating processes [15]. LLMs can perform different types of natural language processing (NLP) tasks without task-specific training (which is also known as *zero-shot*), and rather only by conditioning the model on appropriate prompts [16, 17]. Related terms to zero-shot are in-context learning (ICL) and few-shot learning [18], which is the ability to learn from limited examples [16, 19]. With chain-of-thought (CoT) prompting [20], LLMs produce intermediate steps like smaller building blocks before answering [17] and this approach to prompting has been found to yield significant performance improve-

ments over other prompting methods. A conversational agent or AI assistant based on LLM is ChatGPT [21], a chat interface to GPT. Due to low barriers to entry, prompt engineering is usually done by beginners with no AI expertise [22].

There have been attempts to shift the use of LLMs as assistants to those related to modeling and simulation (M&S) tasks for their natural language generation (NLG) use [23] and for possibilities of implicit human computational modeling abilities [10]. The main argument of [10] (and also cited in [24]) is that LLMs, as they are trained, are (1) computational models of humans, or “homo silicus” [10, 25], and (2) likely already have “latent social information” [10]. Whether or not these models are wrong, they may still be helpful. In contrast to ABMS, human simulation under an LLM setup is out of the researchers’ purview, though there is a way to “influence their behavior with endowments of beliefs, political commitments, experiences, etc.” [10].

Entrepreneurial ventures such as digital and technology startups launching their products and services have used Lean Startup Approaches (LSAs) to test and validate their business model [26]. These LSAs include customer development [27, 28], and Lean Startup [29]. Customer development goes through four stages or steps (not necessarily in linear progression): (1) *customer discovery*, (2) *customer validation*, (3) *customer creation*, and (4) *company-building*. With customer discovery, founders leave the guesswork behind and “get out of the building” to test customer reaction to each “hypothesis” (or guess), gain insights from their feedback, and adjust the business model [26, 28]. However, getting out of the building may still be costly and time-consuming, and complementing the effort with a simulator may help speed up the process. Customer discovery shall be the focus of the simulator in this study.

1.2 Objectives and Research Questions

This study explores using ChatGPT as a simulation tool for customer development, followed by technology startups [26]. Experimentation involves coming up with baseline behavior and feedback without prompt preparations, followed by preparing prospective customers as agents in the simulation environment, which involves the endowment of demographic characteristics and getting behavior and feedback again afterward.

RQ1 What general approach may be followed to simulate customer discovery in preparation for “getting out of the building?”

RQ2 How must the simulation environment be constantly validated and refined to be useful as a valuable co-pilot for startup co-founders?

This study does not assert that the simulations done through ChatGPT should replace performing customer development with real customers. Instead, it is a proposal to consider any resulting products of this research as valuable co-pilots to startup co-founders as they go through their business model discovery efforts.

To address RQ2, *Validation*, from the perspective of computerized marketing models, should possess a satisfactory range of accuracy matching the simulated model to the real-world model [30]. As this study is still in its initial stages, reliance on what is real-world will be based on what is available on the Internet, with room for future studies involving human participants.

2 Review of Related Work

Experiments show that CoT prompting improves performance on tasks involving arithmetic, commonsense, and symbolic reasoning [20]. The approach is based on the idea of *few-shot learning* [16], that is, “one can simply prompt the model with a few input-output exemplars demonstrating the task” instead of fine-tuning the whole model [20]. An attempt to maximize the performance of the simulator in this study shall use CoT techniques applied to the general approach taken by [24].

Role-playing with LLMs can “mimic personas, from fictional characters to historical or contemporary figures” [31]. The role-playing capabilities of LLMs have been extended through CoT prompting. It also allows the extraction of psychological concepts used to understand human behavior through beliefs, desires, goals, ambitions, emotions, and so on [32]. In [10], studies were presented where LLMs were prompted to play different agents with certain demographic properties. Even with similar demographic characteristics, agents may differ stochastically depending on the “temperature” given to the model. Experiments were cited, with the first run having no agents endowed with demographic characteristics, followed by another run with endowment. The structure of the experiments mentioned here will serve as a loose framework for the experiments of this study.

GPT’s training data “may include product reviews, messaging boards, and other online forums with contributions” from various consumers discussing the products they buy [24]. As LLMs respond to prompts with the most likely next text sequence, responses to surveys will mirror the types of responses that would have come from the customers as influenced by the training data [24]. If the responses from the synthetic customers in LLMs are aligned with human behavior as demonstrated in existing studies, then they serve as a cost-effective and immediate method of providing the information otherwise acquired through more expensive and time-consuming means [7]. The study by Brand et al. [24] in turn involves studies with encouraging results, and each study conducted may be broken down into these general steps: (1) involve prompts varying consumer choices and attributes, (2) evaluate the results on how realistic they are and comparing with economic theory or expected realistic response attributes. A recent study involving economic rationality of GPT output [25] strengthens the argument to consider this route. This study will borrow some techniques used for *willingness to pay* (WTP).

However, this study emphasizes using this approach in the context of a co-pilot and a simulator, not as a replacement for actual studies. It is also worth noting, as other studies cited here, that there is no certainty beforehand that GPT will be able to generate helpful responses.

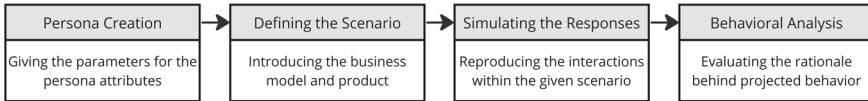


Fig. 1 Proposed prompting approach to conduct WTP simulations on synthetic personas

A study by Aher et al. [7] suggests that it is plausible to use an LLM to simulate human samples from a population and represent different subpopulations and their accompanying multiple biases [33] by injecting prompts with demographic hints. This study shall use the *algorithmic fidelity* of LLMs to accurately represent the distribution of subpopulation characteristics (including biases) to generate multiple personas representing the target market through *silicon sampling* [33].

3 Method

This study centers around eliciting WTP for a product a startup plans. Businesses often conduct price sensitivity surveys (and analyses on the responses) to determine how prospective customers respond to different price points.

This study aims to explore using a large language model (LLM), particularly OpenAI’s GPT-4.0, to simulate a price sensitivity analysis. In lieu of a survey, the optimal price point is determined by asking GPT-4.0 to create a number of personas, describe their purchasing behavior, and simulate their responses to a given product.

This work uses chain-of-thought prompting as its overarching framework. With this, each process step is carried out individually to elicit in-depth responses.

Figure 1 outlines the process: GPT-4.0 creates 30 personas based on attributes, adopts a business analyst persona, clarifies inputs, simulates persona responses with 5 price points, and seeks insights for price sensitivity analysis. For Ateneo de Manila University students, attributes include school, age, year level, gender, strand graduation, scholarship, senior high school, city of residence, and cumulative QPI.

GPT-4.0’s predictive and simulation capabilities evaluate spending habits, interests, and attitudes. Qualitative analysis is performed to understand how it connects persona characterizations with the provided business model.

4 Results

The hypothetical business model consists of a tutoring service startup in the Ateneo de Manila campus. As a business primarily catered toward university students as the target market, the service is as follows:

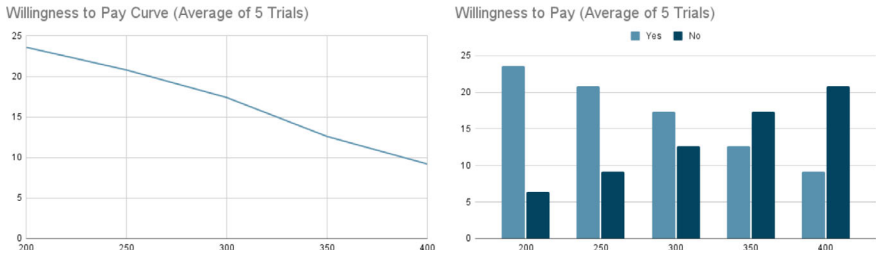


Fig. 2 Number of personas willing to pay versus price, illustrated as a curve and graph

1. The service outsources professional educators for different subjects and areas of study. However, university students may also apply to become tutors.
2. The service focuses on core and major subjects such as Math, Literature, and Science.
3. The service charges a tutoring fee by the hour (however, a half-hour is permitted). Fees remain constant among the tutors or the subject. If in the case that a student extends past 10 min, the price for a half-hour is added.
4. The service may be held on campus or outside, depending on the agreement of the student and tutor. Both online and onsite sessions may be performed.

Upon completion of the entire simulation process, GPT-4.0 managed to provide the willingness to pay (WTP) of the 30 generated customer personas per price point. The curve on the left of Fig. 2 represents the WTP of each persona through a simple Yes-or-No response. The bar graph to the right of Fig. 2 shows the WTP curve of the personas. These charts account for the average of all five trials tested.

Based on the responses garnered from these trials, the optimal price ranges from Php 250–300. Furthermore, upon consultation with the LLM on what it recommends as the suggested tutoring fee, the same range was given among the five. Moreover, GPT-4.0 managed to suggest real-world business strategies and marketing in connection with the recommended price that the co-founder/s can use when designing a business model.

To do initial model validation, it will help to search for any empirical data showing a satisfactory range of accuracy matching the simulated model to the real-world model [34]. An informal Google search with the phrase “metro manila hourly tutorial rates” yields links containing rate info:

- Glassdoor: PhP 100–150
- FB page of an established tutorial center: PhP 750
- FB group of a tutorial services interest community: PhP 120–PhP 300

Note that the prompts used in our simulation had no implicit instructions to let the synthetic students know about these price ranges. Rather, price discovery seems to have occurred indirectly through data used in pre-training. As this is an initial and exploratory attempt, a more rigorous search for empirical data to validate the model may be pursued in future iterations.

To create more precise decisions toward the simulation, the prompt was modified to ensure that the output of all 30 personas was complete, with no brevity or shortcuts. This ensures the GAI has enough information to feed into its conversation memory and provide a complete customer discovery of all personas. Through this, the capabilities of GPT-4.0 were experimented on further, observing how it understood human behavioral patterns and attributes. Indications of economic theory and explicit references to willingness to pay (WTP) were omitted from the prompts to ensure that GPT-4.0 is capable of producing demand without pre-trained concepts into its memory file. By using the LLM's generated personas, adjustments were implemented to evaluate how realistic the simulations were compared to economic principles, determining its response to human attributes. By properly assessing if the prompting process above can emulate customer demand from generated human subjects, the viability of GPT-4.0 and other LLMs as a low-cost alternative for business model refinement and customer discovery can be determined. To ensure that the LLM is given enough time to think through precise prompts, several iterations of the process were done. This led to the final prompting process, which was performed five times to collect the average WTP of all personas. Throughout the experiment, reevaluation was performed to determine if the LLM understood the goal of the prompt accurately.

5 Discussion

Revisiting the research questions with the developed context yields these answers:

RQ1 What general approach may be followed to simulate customer discovery in preparation for “getting out of the building?”

Utilizing ChatGPT to gauge customer reactions provides entrepreneurs the flexibility for industry-contextualized simulations, allowing for the refinement of the business model. Integrating the prompting frameworks discussed, one can better understand customer behavior for each hypothetical scenario and prepare for “getting out of the building.” This expedites customer analysis, identifying optimal price points for revenue generation, thus fortifying the business model's foundations before entering the real market.

RQ2 How must the simulation environment be constantly validated and refined to be useful as a valuable co-pilot for startup co-founders?

To create realistic simulations with ChatGPT, it is vital to ensure the realistic identities of the generated personas. Assuming co-founders input specific customer demographics, the GAI must produce personas following certain attribute dependencies (e.g., 1st-year students are generally 17–18 years old). The business model must also be feasible enough to be simulated, as GPT-4.0 focuses on the prompt itself rather than real-world data. Clarity and specificity are essential for ensuring meaningful

responses from the GAI for customer discovery and analysis. Both qualitative and quantitative responses may be requested to ensure that the rationale behind the simulation is applicable to the actual environment of such business, which can then prompt the co-founder/s to reevaluate their models and price points.

Ethics must be considered for this use case of GAIs. Potential biases and misrepresentation due to societal stereotypes and boundaries found on the Internet, including the exclusion of minorities, may yield inaccurate information on customer dynamics. Privacy and legal issues, particularly regarding data privacy laws, may also arise when providing real-world customer data to prompt the LLM for expanded customer analysis and discovery, potentially risking customer safety. Additionally, the use of GAIs may result in the loss of jobs and autonomy, as the work of lead generators and customer hunters may be accomplished by artificial intelligence instead.

6 Conclusion

This paper explored how LLMs may be used as simulators of prospective consumers to perform customer discovery tasks through prompting techniques only (i.e., without document retrieval and fine-tuning). The results suggest that there is reason to pursue further work in this area and that if further work is successful, this approach may lower costs for market validation activities.

No custom pre-training means using the stock and static knowledge of the LLM. The simulator may also take on pre-trained influenced behavior, which may not necessarily be consistent with what the simulator creators value more [35]. Future work in this area may include additional few-shot training and fine-tuning. Before this approach can be deployed in practical settings, work may need to be done on validating whether LLMs demonstrate knowledge of non-intuitive consumer WTP and whether LLMs can accurately identify whether a price range is, in its entirety, too high or too low to be meaningful.

Acknowledgements We want to thank the John Gokongwei School of Management, the Rizal Library Open Access Journal Publication Grant, and the University Research Council of the Ateneo de Manila University for making our participation and future presentation of this study possible.

References

1. Eisenmann TR (2013) Entrepreneurship: a working definition. *Harvard Bus Rev* 10:2013
2. Ilagan JB (2022) The design and use of agent-based modeling computer simulation for teaching technology entrepreneurship
3. Haynie M, Shepherd DA (2009) A measure of adaptive cognition for entrepreneurship research. *Entrepreneurship Theory Practice* 33(3):695–714. SAGE, Los Angeles, CA
4. Gibbons B, Fernando M, Spedding T (2022) Innovation through developing a total enterprise computer simulation: teaching responsible decision making. *J Manage Educ* 46(1):16–

42. <https://doi.org/10.1177/1052562920987591>, 1 citations (Crossref) [2023-09-23] Publisher: SAGE Publications Inc
5. Kapp KM, Blair L, Mesch R (2014) The Gamification of learning and instruction. Ideas into practice
6. Jackson I, Saenz MJ (2022) From natural language to simulations: applying GPT-3 codex to automate simulation modeling of logistics systems. arXiv preprint [arXiv:2202.12107](https://arxiv.org/abs/2202.12107)
7. Aher G, Arriaga RI, Kalai AT (2022) Using large language models to simulate multiple humans. arXiv preprint [arXiv:2208.10264](https://arxiv.org/abs/2208.10264)
8. Carayannis EG, Provanca M, Grigoroudis E (2016) Entrepreneurship ecosystems: an agent-based simulation approach. *J Technol Transf* 41:631–653. Springer
9. Macal CM, North MJ (2008) Agent-based modeling and simulation: ABMS examples. In: *Proceedings—winter simulation conference*, pp 101–112. <https://www.scopus.com/inwardrecord.uriid=2s2.0-60849084592&doi=10.1109%2fWSC.2008.4736060&partnerID=40&md5=b5d6f62b004bdba79921f0260a5c67ae>, 54 citations (Crossref) [2023-09-23]
10. Horton JJ (2023) Large language models as simulated economic agents: what can we learn from homo silicus? Tech. rep, National Bureau of Economic Research
11. Baidoo-Anu D, Owusu Ansah L (2023) Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484
12. Shanahan M (2023) Talking about large language models. <http://arxiv.org/abs/2212.03551>
13. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: *Advances in neural information processing systems*, vol 30
14. Cooper G (2023) Examining science education in chatgpt: an exploratory study of generative artificial intelligence. *J Sci Educ Technol* 32(3):444–452. Springer
15. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC (2023) A prompt pattern catalog to enhance prompt engineering with ChatGPT. <http://arxiv.org/abs/2302.11382>, [arXiv:2302.11382](https://arxiv.org/abs/2302.11382) [cs]
16. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D (2020) Language models are few-shot learners. In: *Advances in neural information processing systems*, vol 33. Curran Associates, Inc., pp 1877–1901. <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>
17. Qin C, Zhang A, Zhang Z, Chen J, Yasunaga M, Yang D (2023) Is ChatGPT a general-purpose natural language processing task solver? <http://arxiv.org/abs/2302.06476>
18. Bragg J, Cohan A, Lo K, Beltagy I (2021) FLEX: unifying evaluation for few-shot NLP. In: *Advances in neural information processing systems*, vol 34, pp 15787–15800. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2021/hash/8493eeaccb772c0878f99d60a0bd2bb3-Abstract.html>
19. Zhao Z, Wallace E, Feng S, Klein D, Singh S (2021) Calibrate before use: improving few-shot performance of language models. In: *International conference on machine learning*. PMLR, pp 12697–12706
20. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le Q, Zhou D (2023) Chain-of-thought prompting elicits reasoning in large language models. <http://arxiv.org/abs/2201.11903>, [arXiv:2201.11903](https://arxiv.org/abs/2201.11903) [cs]
21. Introducing ChatGPT (2023). <https://openai.com/blog/chatgpt>. OpenAI
22. Zamfirescu-Pereira J, Wong RY, Hartmann B, Yang Q (2023) Why Johnny Can't prompt: how non-AI experts try (and fail) to design LLM prompts, pp 1–21. ACM, Hamburg Germany. <https://dl.acm.org/doi/10.1145/3544548.3581388>, 1 citations (Crossref) [2023-09-10]
23. Giabbanelli PJ (2023) GPT-based models meet simulation: how to efficiently use large-scale pre-trained language models across simulation tasks. <http://arxiv.org/abs/2306.13679>, [arXiv:2306.13679](https://arxiv.org/abs/2306.13679) [cs]

24. Brand J, Israeli A, Ngwe D (2023) Using GPT for market research. Available at SSRN 4395751
25. Chen Y, Liu TX, Shan Y, Zhong S (2023) The emergence of economic rationality of GPT. arXiv preprint [arXiv:2305.12763](https://arxiv.org/abs/2305.12763)
26. Ghezzi A (2019) Digital startups and the adoption and implementation of lean startup approaches: effectuation, bricolage and opportunity creation in practice. *Technol Forecasting Soc Change* 146:945–960, 91 citations (Crossref) [2023-09-23]
27. Blank S (2013) Why the lean start-up changes everything. *Harvard Buss Rev*
28. Blank S, Dorf B (2005) The path to epiphany: the customer development model. The four steps to the epiphany, pp 17–28
29. Ries E (2011) *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Currency
30. Garcia R, Rummel P, Hauser J (2007) Validating agent-based marketing models through conjoint analysis. *J Bus Res* 60(8):848–857. <https://www.sciencedirect.com/science/article/pii/S0148296307000410>, 43 citations (Crossref) [2023-09-23]
31. Kong A, Zhao S, Chen H, Li Q, Qin Y, Sun R, Zhou X (2023) Better zero-shot reasoning with role-play prompting. arXiv preprint [arXiv:2308.07702](https://arxiv.org/abs/2308.07702)
32. Shanahan M, McDonell K, Reynolds L (2023) Role-play with large language models. arXiv preprint [arXiv:2305.16367](https://arxiv.org/abs/2305.16367)
33. Argyle LP, Busby EC, Fulda N, Gubler JR, Rytting C, Wingate D (2023) Out of one, many: Using language models to simulate human samples. *Political Anal* 31(3):337–351. Cambridge University Press
34. Cadotte E (2014) The use of simulations in entrepreneurship education: Opportunities, challenges and outcomes. In: *Annals of entrepreneurship education and pedagogy*, pp 280–302. Edward Elgar
35. Phelps S, Ranson R (2023) Of Models and Tin Men—a behavioural economics study of principal-agent problems in AI alignment using large-language models. arXiv preprint [arXiv:2307.11137](https://arxiv.org/abs/2307.11137). <https://arxiv.org/abs/2307.11137>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Ethical Education Data Mining Framework for Analyzing and Evaluating Large Language Model-Based Conversational Intelligent Tutoring Systems for Management and Entrepreneurship Courses



Joseph Benjamin R. Ilagan , Jose Ramon S. Ilagan ,
and Maria Mercedes T. Rodrigo 

Abstract Educational data mining (EDM) can be used to design better and smarter learning technology by finding and predicting aspects of learners. Amend if necessary. Insights from EDM are based on data collected from educational environments. Among these educational environments are computer-based educational systems (CBES) such as learning management systems (LMS) and conversational intelligent tutoring systems (CITSs). The use of large language models (LLMs) to power a CITS holds promise due to their advanced natural language understanding capabilities. These systems offer opportunities for enriching management and entrepreneurship education. Collecting data from classes experimenting with these new technologies raises some ethical challenges. This paper presents an EDM framework for analyzing and evaluating the impact of these LLM-based CITS on learning experiences in management and entrepreneurship courses and also places strong emphasis on ethical considerations. The different learning experience aspects to be tracked are (1) learning outcomes and (2) emotions or affect and sentiments. Data sources comprise Learning Management System (LMS) logs, pre-post-tests, and reflection papers gathered at multiple time points. This framework aims to deliver actionable insights for course and curriculum design and development through design science research (DSR), shedding light on the LLM-based system's influence on student learning, engagement, and overall course efficacy. Classes targeted to apply this framework have 30–40 students on average, grouped between 2 and 6 members. They will involve sophomore to senior students aged 18–22 years. One entire semester takes

WWW home page: <https://www.ateneo.edu/>.

J. B. R. Ilagan (✉) · J. R. S. Ilagan · M. M. T. Rodrigo
Ateneo de Manila University, Quezon City, Philippines
e-mail: jbilagan@ateneo.edu

J. R. S. Ilagan
e-mail: jrilagan@ateneo.edu

M. M. T. Rodrigo
e-mail: mrodrigo@ateneo.edu

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_6

about 14 weeks. Designed for broad application across diverse courses in management and entrepreneurship, the framework aims to ensure that the utilization of LLMs in education is not only effective but also ethically sound.

Keywords LLM · Large language models · Conversational intelligent tutoring systems · CITS · Generative AI · GenAI · GPT · ChatGPT · Design science research

1 Introduction

Educational data mining (EDM) can be used to design better and smarter learning technology and to inform learners and educators better [1]. Data for use in EDM may come from various computer-based education system (CBES) sources. LMS software, a type of CBES, covers course delivery capabilities such as administration, documentation, tracking, and reporting of programs, classroom and online events, course content, and e-learning. LMS logs “provide granular, near-real-time records of student behavior as learners interact with online course materials in digital learning environments” [2]. An LMS records student activities, such as reading, writing, taking tests, performing specific tasks, and commenting on events with peers [3]. A conversational intelligent tutoring systems (CITSs), also a type of CBES, facilitates tutoring content through natural language to cover concepts, break down the learning material into conversations, ask and answer questions, determine knowledge gaps, and provide contextual feedback and corrective interventions [4].

Generative artificial intelligence (GAI) is a machine learning framework that generates content using probability and statistics learned from existing digital content such as text, video, images, and audio through training data [5]. A large language model (LLM) is a GAI trained from text created by humans in order to produce human-like language [5, 6]. Generative pre-trained transformer (GPT), an LLM-based system, is designed to generate sequences of words, code, or other data, starting through source input called the *prompt* [7]. GPT is based on the *transformer* architecture [8], which can train from large public digital data quickly in parallel. GAI is already being applied [9]. ChatGPT is a conversational interface to GPT [10]. The use of ChatGPT as chatbots in education has also been reported in blog posts and social media [11].

Innovation in education through DSR involving GAI will require classrooms to be ready for EDM to determine whether technology interventions help improve learning experiences. However, the data collection and integration face ethical challenges [12], which, if not thoroughly addressed, can compromise the integrity of the research. This study proposes a framework that is efficient and deeply rooted in ethical principles. Among the ethical concerns that need addressing are the location and interpretation of data, informed consent, privacy, de-identification of data, and the classification

and management of data [13]. The aim is to make the process more understandable, streamlined, and manageable for researchers, encouraging more ethical and effective research in education.

1.1 Objective and Research Questions

This study proposes a framework to fit university management and entrepreneurship classes with the instrumentation of LMS, LLM-based CITS technologies, and non-CITS interfaces for continuous course and curriculum improvement. The framework follows a multi-dimensional evaluation strategy, combining quantitative and qualitative methods to assess the effects on learning experiences.

RQ1. How should classes be set up with instrumentation rooted in ethical principles using LMS and CITS technologies for EDM? **RQ2.** What useful data can be captured by a CBES, whether a CITS or a regular interface, for rich student behavior and insights on emotions through EDM? **RQ3.** How can EDM help evaluate LLM-based CITS for effectiveness in achieving student learning outcomes with the right balance of thoroughness and ethical considerations? **RQ4.** How can EDM from CITS help in the course and curriculum development?

With this scaffolding, research may progress more smoothly with strengthened compliance with ethical clearance guidelines. The structures will have ethics considerations baked in and not simply an afterthought.

2 Review of Related Literature

Beyond trying to improve learning outcomes [3], researchers have also used EDM to model affect (emotion in context), engagement, meta-cognition, and collaboration [1]. Learning analytics (LA), related to the concept of EDM and traditional education research, involves the collection, measurement, analysis, and reporting of data about learners and their contexts for purposes of understanding and improving learning and the environments in which it occurs [14]. Types of data used in LA were proposed in [14] based on a study involving various cases. This study will make use of the following: (1) data resulting from measurement of artifacts (e.g., communication artifacts): blogs, group discussions, discussion boards, and social network data; (2) repurposed data initially collected for other purposes and reanalyzed: survey data; (3) transformed data, or primary data after transformation processes: LMS logs.

Activity indicators in LMS logs can serve as proxies of student engagement under certain conditions [2]. A learning analytics information system (LAIS) architecture proposed by Nguyen et al. [15] involves a data pipeline covering the following stages: (1) logging services and event trackers from learning and teaching service sources, (2) batch extraction-transformation-loading (ETL) and events transformation through event adapters and services, and (3) learning analytics services. Three

broad overlapping categories of issues involving ethics in LA were proposed by Slade et al. [13]: (1) the location and interpretation of data, (2) informed consent, privacy, and de-identification of data, and (3) management, classification, and storage of data. Personally identifiable information (PII) is “any information that can identify an individual, and de-identification is used to prevent revealing individual identity and keeping the PII confidential” [12]. This study shall use these categories of [13] as a checklist for addressing EDM ethical concerns. De-identification techniques shall be borrowed from the work of [12]: anonymization (de-identifying data while preserving its original format), masking (replacing sensitive data with fictional data while still making records usable), and blurring (adding noise to records). Design science research (DSR) calls for “creating innovative artifacts to solve real-world problems” iteratively [16]. DSR is relevant to this study in two ways: First, attempts to use LA to support DSR for learning and teaching in higher education have emerged [15]. Second, the ethical EDM framework that is the subject and artifact of this study shall also evolve through DSR principles. Ethics in research involving human subjects involve the following: integrity and honesty, justice and fairness, (3) safety and beneficence, and respect for human rights and dignity [17].

3 Methodology

This study intends to use the EDM framework for management and entrepreneurship classes from a management school in the Philippines. The classes usually span one full semester of 14 weeks. Each class may range from 6 to 50 students, 30–40 on average, and has about an equal number of males and females from 18 to 22 years old. The students will interact with the CBES iteratively throughout. The students are also expected to answer course-specific reflection assignments at the start, middle, and end of the semester. The CBES sessions will be conducted either onsite in the classrooms or remotely online. The underlying research protocol, which may be updated regularly, will be applied and customized to specific classroom requirements as needed.

The sources of data for quantitative and qualitative analytics are the following: (1) LMS logs, (2) CBES logs (whether from simulators, CITS, and non-CITS front-ends), (3) reflection papers, and (4) pre- and post-session survey questionnaires. From these data sources, qualitative and quantitative analytic methods range from thematic and sentiment analyses of reflection papers to statistical evaluations of performance indicators.

LMS logs capture user interactions within the LMS, such as session durations, page views, assignment submissions, and others [2]. They offer valuable insights into student behavior, engagement, and learning activity. From the CBES, the simulator logs record actions and decisions participants make during simulation exercises. These are useful for assessing performance and decision-making skills. The CITS conversation logs capture the content and metadata of conversations within the chat-based CBES, helping to analyze communication effectiveness. The web logs document the network-level interactions, providing useful data for LA. Other logs for

events not covered by the previous categories track specific events triggered within the CBES, such as completion of tasks or participation in collaborative activities.

For reflection paper assignments, instead of separate informed consent forms, the informed consent clauses are placed at the start of each reflection assignment prompt. The clauses were derived from a template provided by the university's research ethics office. At the start of each semester, a questionnaire asks about the student's background, interests, and previous experiences in technology entrepreneurship. These questions may be embedded in a start-of-semester reflection if one exists. Reflection papers may ask about emotions based on [18]. During the semester, pre- and post-assessments and evaluations will be done coincidentally to interact with the CBES to determine if learning outcomes are met. Results will then be contextualized through insights from EDM. Student participation may follow this general path: For the first session, students complete a pre-session survey, including questions about the student's background, interests, and previous experience in technology entrepreneurship (if not yet collected and shared previously). Then, the student engages with the CBES to learn various concepts related to management and entrepreneurship. The student will be asked to make decisions, solve problems, and answer questions in a simulated environment. The CBES could either be a CITS or a regular web interface. After the session, the students will complete a post-session survey to provide feedback on the experience and understanding of the concepts taught through the CBES. For the second session, students will use the computer simulation with updated content based on the feedback from the first session. After, the students will complete a post-session survey similar to the first session. Additional sessions may be introduced as necessary. The CBES in simulators shall capture logs throughout the session. The logs will be fed into an EDM data platform and matched with their corresponding student information. Any of the following setups for sessions may be employed throughout the semester:

- **Within-subject design:** The CITS may be used in some topics for some students, while others won't use the CITS.
- **Experimental:** For pre-identified topics, half of the students in the class chosen at random will be assigned the CITS, and the rest will not be assigned the CITS. For this setup, no grades will be given.
- **Cross-over design:** There will be switches between the CITS and non-CITS interface at different points.

At the end of each semester, a questionnaire prompts the student's reflections on learning related to using the various forms of CBES. These questions may be embedded in an end-of-semester reflection if one exists.

The reflection assignments serve as sample templates and are intended for educational use. They may be modified as needed, provided that they adhere to established ethical guidelines. A new ethics review will be submitted for approval if changes are significant. The question sets will be updated periodically. Version control will be maintained to reflect such changes. Table 1 illustrates sample reflection papers answered by students at the start of the semester. The reflections in the middle and at the end of the semester would have a similar flow.

Table 1 Sample reflection paper (start of semester)

<p>Start-of-semester reflection points: 100 submitting a text entry box, a website URL, a media recording, or a file upload</p>	<p>Consent and data privacy clause: Before proceeding with this mandatory assignment, please read the following information carefully</p> <p>Educational and research use: We seek to evaluate the effectiveness of various class interventions in achieving student learning experiences. The data we will collect and analyze will help in the course and curriculum development. While completing this reflection assignment is a course requirement, the responses will be used for educational purposes</p> <p>Opt-out option: Participation in the research component means that you agree to have your answers included in the research. While reflection prompts are graded, research participation is voluntary. If you decide not to participate, there will be no penalty or negative consequence. If you wish to opt-out, please include "I Opt-Out of Research Use" at the end of your submission. If you no longer wish to be part of the research study and withdraw your data, please contact the researcher as soon as possible</p> <p>Confidentiality, anonymization, and security: All data used in the research will be stripped off of personal identifiable information and securely stored, complying with data protection standards. Your full name will not appear on any of the questionnaires, and information identifying you will not appear in any report or publication of this research. Only the principal investigator and designated researchers will know the identity associated with the information collected for this study, and they will not reveal it to anyone else</p> <p>Payments: You will not have to pay anything, nor will you be paid for participating in the research component</p> <p>Contact: For concerns about this process or data privacy, please contact [Name of proponent] (email address). Do not agree to participate if you are unsure or have remaining questions. Please ask the researcher to explain anything unclear to you, including any words in this assignment's text</p> <p>Part 0: Background and interests: Describe your academic background briefly. What led you to take up this program? What are your interests related to technology entrepreneurship? Do you have any previous experiences in technology entrepreneurship? If so, describe them</p> <p>Part 1: Skill areas and emotional connections coding and development</p> <p>What primary emotion best describes your feeling toward coding and development? (Joy/Trust/Fear/Surprise/Sadness/Disgust/Anger/Anticipation)</p> <p>What factors contribute to this emotion? ...</p> <p>By submitting this assignment, you confirm your agreement to the consent and data privacy clause stated above. Thank you</p>
---	--

Table 2 Sample post-session survey questions

<p>Post-session assessment whether required and graded will be determined on a case-to-case basis</p>	<p>Consent and data privacy clause ... <i>[Same as in Table 1]</i> Please indicate the system ID shown on the screen Sentiments after session: Choose an emotion to describe your feelings after this session. (Options: interest, surprise, joy, sadness, anger, fear) Learning outcomes: In free-form text, what have you learned during this session? CBES feedback: Did the computer-based educational system (CBES) help you achieve your learning goals? Explain briefly Suggestions for improvement: Any suggestions to improve the learning experience?</p>
---	---

Similar to the reflection prompts, informed consent clauses are placed in the beginning of the pre- and post-session survey questionnaires. Table 2 shows a sample post-session survey questions. The pre-session survey is almost similar but gets baseline information.

Data collected in this study may be shared with third-party researchers for future research but will undergo a rigorous anonymization process to remove PII. Any research agreement shall be bound by a memorandum of agreement (MOA) and accompanied by a data protection covenant (DPC). All shared data will be anonymized and retained for up to 10 years to facilitate future research endeavors. After the retention period, the data will be securely deleted following institutional guidelines and applicable laws. Interested parties must adhere to ethical guidelines outlined by an institution’s ethics board before accessing the data. Due to these procedures, the researchers may discover a condition previously unknown to the participant (e.g., disease). The researchers will have to contact the respondents again to ask if an opt-out would be considered when this happens. The categories in Table 3 (columns 2–4) are from [13]. The additional ethical considerations and disclaimers in this section were derived from the Informed consent form guidance and template document from the University Research Ethics Office of the Ateneo de Manila University, formulated and counterchecked with the assistance of OpenAI’s ChatGPT to ensure comprehensive coverage of ethical issues. There is not one prompt that led to the following operational clauses.

Each data source will be tagged with a unique, random identifier to combine multiple data sources under one identifier. Data with the same unique identifier will be consolidated into a student profile, including metrics and insights from LMS logs, CBES logs, reflection, and survey assignments. After aggregation, specific attributes

Table 3 Categorized ethical considerations

Data source	The location and interpretation of data	Informed consent, privacy, and de-identification of data	Management, classification, and storage of data
LMS logs	LMS server, or on protected cloud storage	No PII. De-identification of data records. Restricted academic third-party researchers-party data access.	Quantitative data for machine learning
CBES logs	Simulator and conversation logs in cloud storage	Pop-up screen asking for consent	CITS conversations for text mining
Reflection assignments	LMS server, or on protected cloud storage	Consent asked explicitly; PII removed; right to withdraw from the study anytime, with data deleted	Quantitative data for machine learning; Qualitative data for text mining
Pre/Post-session surveys	LMS server, or on protected cloud storage	<i>Same as reflection assignments</i>	<i>Same as reflection assignments</i>

Italic shows the information about reflection assignments

from different sources will be mapped to standardized variables for combined analysis. All PII will be stripped from the data or anonymized, masked, and blurred. Data will be processed in secure, authorized cloud-based storage and computing accounts. Each step of the process will be logged and audited for compliance with privacy guidelines. Consolidated data may be encrypted in the end.

Figure 1 shows the ethical EDM framework's data infrastructure architecture inspired by Nguyen et al. [15].

4 Discussion

This section addresses the study's research questions. For **RQ1** involving class setup with ethical LMS and CITS instrumentation, a data management architecture for the LAIS will be introduced, emphasizing ethical considerations and data privacy. Informed consent will be part of reflection papers and session questionnaires. Addressing **RQ2** touching on data captured by a CBES, the LAIS will manage sources like logs, reflection papers, and session questions, covering both learning outcomes and emotional experiences. **RQ3**, related to evaluating LLM-based CITS effectiveness, will involve both qualitative and quantitative analyses based on trans-

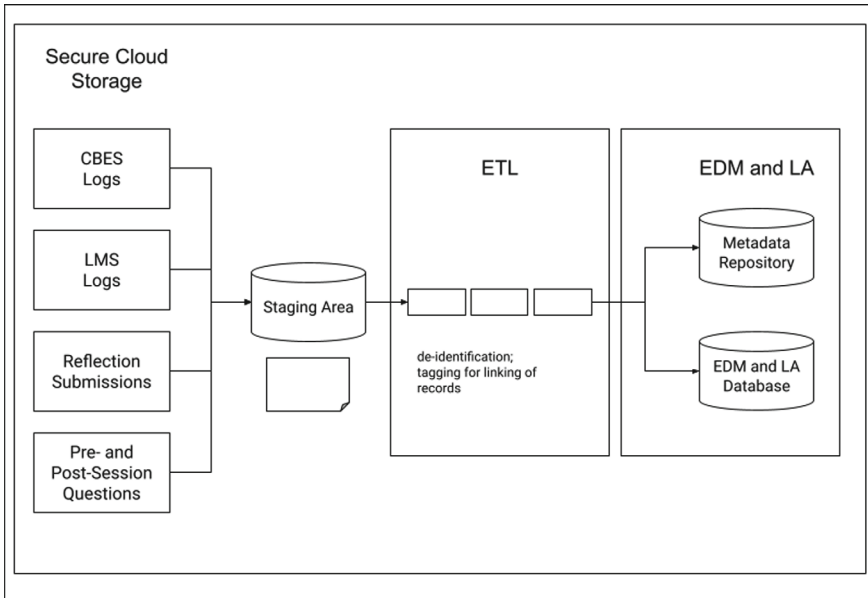


Fig. 1 EDM data infrastructure architecture

formed data from various sources. Lastly, **RQ4**, which is about course and curriculum development, insights from EDM will guide the process, possibly within the design science research scope.

5 Conclusion

This study aims to provide a method for understanding student engagement and performance through a multi-source data collection approach. Integrating various data sources will enable a complete view of students' learning experiences. With a heavy focus on ethical considerations, the study ensures the ethical integrity of the research process. The study aims to serve as a model and streamline the process of instrumenting classes to balance the need for in-depth analytics with the ethical imperatives of privacy and confidentiality. The plan is to formalize this study into a comprehensive research protocol to serve as a guide for similar research efforts.

Acknowledgements We want to thank the Department of Quantitative Methods and Information Technology of the John Gokongwei School of Management, the Rizal Library Open Access Journal Publication Grant, and the University Research Council of the Ateneo de Manila University for making our participation and future presentation of this study possible.

References

1. Baker RS (2014) Educational data mining: an advance for intelligent systems in education. *IEEE Intell Syst* 29(3):78–82. IEEE
2. Motz B, Quick J, Schroeder N, Zook J, Gunkel M (2019) The validity and utility of activity logs as a measure of student engagement. In: *Proceedings of the 9th international conference on learning analytics and knowledge*, pp 300–309. ACM, Tempe AZ USA. <https://dl.acm.org/doi/10.1145/3303772.3303789>, 11 citations (Crossref) [2023-09-30]
3. Romero C, Ventura S (2013) Data mining in education. *Wiley Interdisc Rev Data Mining Knowl Discov* 3(1):12–27. Wiley
4. Holmes M, Latham A, Crockett K, O’Shea JD (2017) Tomorrow’s learning: involving everyone. learning with and about technologies and computing. In: *11th IFIP TC 3 world conference on computers in education, WCCE 2017, Dublin, Ireland, July 3–6, 2017, Revised Selected Papers. IFIP Advances in Information and Communication Technology*, vol 515, pp 251–260. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041551804&doi=10.1007%2f978-3-319-74310-3_27&partnerID=40&md5=d0f6929e9b416a1e6036891b725905ab, 2 citations (Crossref) [2023-07-11]
5. Baidoo-Anu D, Owusu Ansah L (2023) Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484
6. Cooper G (2023) Examining science education in chatgpt: an exploratory study of generative artificial intelligence. *J Sci Educ Technol* 32(3):444–452. Springer
7. Floridi L, Chiriatti M (2020) GPT-3: its nature, scope, limits, and consequences. *Minds Mach* 30:681–694
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Process Syst* 30
9. Adiguzel T, Kaya MH, Cansu FK (2023) Revolutionizing education with AI: exploring the transformative potential of ChatGPT. *Contemporary Educ Technol* 15(3):ep429. Bastas
10. Introducing ChatGPT (2023). <https://openai.com/blog/chatgpt>. OpenAI
11. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, Agyemang B (2023) What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn Environ* 10(1):15. Springer
12. Khalil M, Ebner M (2016) De-identification in learning analytics. *J Learn Anal* 3(1):129–138. <https://learning-analytics.info/index.php/JLA/article/view/4519>
13. Slade S, Prinsloo P (2013) Learning analytics: ethical issues and dilemmas. *Am Behav Sci* 57(10):1510–1529. <http://journals.sagepub.com/doi/10.1177/0002764213479366>, 400 citations (Crossref) [2023-09-26]
14. Nistor N, Hernández-García (2018) What types of data are used in learning analytics? An overview of six cases. *Comput Hum Behav* 89:335–338. Elsevier. https://www.sciencedirect.com/science/article/pii/S0747563218303601?casa_token=VrQyzC4OTjoAAAAA:BFd2G1u2j33n6TlrH_J4IEKOn-_zIKBgZQltzN0T0EJZxbcdRlk6TrzDEA1pWFvdTP6sWiNa14_2
15. Nguyen A, Tuunanen T, Gardner L, Sheridan D (2021) Design principles for learning analytics information systems in higher education. *Eur J Inf Syst* 30(5):541–568. <https://www.tandfonline.com/doi/full/10.1080/0960085X.2020.1816144>, 17 citations (Crossref) [2023-10-08]
16. Hevner A, Chatterjee S (2010) Design science research in information systems. In: *Design research in information systems*, vol 22, pp 9–22. Springer US, Boston, MA. http://link.springer.com/10.1007/978-1-4419-5653-8_2, series title: *Integrated Series in Information Systems*
17. Rodrigo MMT (2023) Is the AIED conundrum a first-world problem? *Int J Artif Intell Educ*. <https://link.springer.com/10.1007/s40593-023-00345-2>
18. Plutchik R (1980) A general psychoevolutionary theory of emotion. In: *Theories of emotion*, pp 3–33. Elsevier. <https://www.sciencedirect.com/science/article/pii/B9780125587013500077>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Sensing the Emergence of New Structure in Matter and Life and Its Impact on Physics, Chemistry, and Biology, in a Light-Based Quantum Computational Model



Pravir Malik 

Abstract Recognition of new patterns in layers of matter and life is heavily dependent on the foundational schema or model that animates a sensing computational system. A symmetrical, fourfold light-based quantum computational model that links together layers such as quantum particles and molecular plans in cells, among other layers is positioned as the schema required to model a range of existing and even emerging patterns in and across levels. Fourfoldness in the model—related to four parts of a single pattern—stipulates the default way in which pattern recognition occurs. When conditions in the model are fulfilled such that the fourfoldness has reached a threshold level, then a fifth property implicit in the model surfaces, which allows specific emerging categories of patterns related to fivefoldness to be recognized. The prerequisite condition for such recognition is related to fourfold meta-function in source layers being more easily reflected in the external layers of the model, which morphs the mathematics of the model itself, to allow patterns related to the implicit or fifth property to also be recognized. Therefore, dynamics that we associate with space, time, energy, gravity, or the layer of quantum particles, or the layer of atoms, or the layer of molecular plans, can be parsed in terms of fourfold patterns, subsequently allowing the possibility of recognition of emerging patterns related to fivefoldness. Recognition of such patterns will provide additional insight into the changing physics, chemistry, and biology of target systems.

Keywords Complex adaptive systems · Symmetry · Light · Pattern recognition · Computational intelligence

P. Malik (✉)
Deep Order Technologies, El Cerrito, CA 94530, USA
e-mail: pravir.malik@deepordertechnologies.com

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_7

1 Introduction

There is fourfold structure that informs and animates the layers of matter and life. This is evident in considering the fourfold space, time, energy, and gravity parameters of cosmos, the fourfold quark, lepton, boson, Higgs boson structure of quantum particles, the fourfold s-Shell, p-Shell, d-Shell, f-Shell structure of the periodic table, and the fourfold nucleic acid, polysaccharide, lipid, and protein structure of molecular plans instrumental in living cells.

Physics—primarily concerned with the space–time–energy–gravity layer and the quantum particle layer; chemistry—primarily concerned with the periodic table layer; and biology—primarily concerned with the living cell layer are generally considered to be stable and fixed.

This paper, however, will make the case that due to the persistent quantum computation that animates cosmos to ensure its operation as a complex adaptive system (CAS) as made evident in the symmetrical light-based model explored in prior IEEE papers [1] and books on cosmology of light [2], it is inevitable that a fifth construct, implicit in the fourfold model of light, will emerge to add further complexity to layers of matter and life, thereby enhancing the possibilities currently encapsulated by physics, chemistry, and biology. The right base model of the type proposed in this paper will allow sensing mechanisms to become sensitive to such changes.

Section 2, the Fourfold Quantum-Computational Model of Light, summarizes the previously introduced fourfold, symmetrical model of light, and the notion of quantization that connects one layer to each subsequent cascading layer in the model. This model becomes the basis for recognizing a genre of fourfold patterns in layers of matter and life.

Section 3, the Fifth Construct in a Quantum-Computational Model of Light, suggests the emergence of a fifth construct implicit in the fourfold model of light. This emergence is tied to viewing cosmos as a complex adaptive system [3] and will allow sensitivity to emerging fivefold patterns in layers of matter and life.

Section 4, Sensing the Fivefold Electromagnetic Spectrum, suggests how the electromagnetic layer instrumental to physics and operation in the cosmos may itself change due to the CAS reality of the cosmos. Sensitivity to the emerging fivefold pattern would allow subtle changes in the spectrum to be sensed.

Section 5, Sensing the Fivefold Structure of Quantum Particles, suggests how the structure of quantum particles may change due to the underlying CAS reality. Sensitivity to the emerging fivefold pattern would allow subtle changes in quantum particles to be sensed.

Section 6, Sensing the Fivefold Structure of the Periodic Table, suggests how the structure of atoms may change due to the underlying CAS reality. Sensitivity to the emerging fivefold pattern would allow subtle changes in atoms to be sensed.

Section 7, Sensing the Fivefold Structure of Molecular Plans in Cells, suggests how the structure of molecular plans may change due to the underlying CAS reality. Sensitivity to the emerging fivefold pattern would allow subtle changes in the structure of molecular plans to be sensed.

Section 8, Summary and Conclusion, integrates the core theses and implications suggested by this paper.

2 The Fourfold Quantum-Computational Model of Light

At the heart of the model which will facilitate sensing of the emergence of new structure in matter and life, lies a symmetrical, multi-dimensional model of conceptual space [4, 5]. This model will be seen to have an intimate bearing on how matter and life emerge. The symmetry and multi-dimensionality derive from imagining light traveling at several constant speeds, beyond the known speed of ‘ c ’, 186,000 miles per second in vacuum in the physical universe. Light is used to construct these spaces since it is known that the way it travels has an effect on space, time, and the movement and even perception of objects [6].

To begin to construct the first layer of conceptual space, imagine light traveling infinitely fast. Note that conceptual spaces created by such faster-than-light speeds perhaps are envisioned to be similar to property spaces [7].

Consider a volume of any size and imagine a light source at the center. What can be surmised is that since light is traveling infinitely fast, it will be present in that volume instantaneously. That is, light will be present everywhere all at once. This suggests a property of light true in that conceptual space of ‘presence’. This property will be referred to as ‘Pr’. Now, since this light is instantaneously present everywhere if anything were to arise or disappear in that volume it will immediately be recorded in the fabric of light. That is, the light will have a ‘knowledge’ of everything that is happening in its space. This property of knowledge will be referred to as ‘K’. Further, since everything that arises or disappears, or exists in that volume will be connected by the pervasive fabric of light, that connection will create a property of harmony. This harmony will be referred to as ‘H’. Finally, since light is all pervasive, if anything not of the nature of light arises, it will sooner or later be overpowered by the nature of light. This property of ‘power’ can be referred to as ‘Po’.

Using mathematical notation the conceptual space created by light imagined traveling infinitely fast is summarized by Eq. (1) as a set containing four properties, and where R_{C_∞} depicts reality (R) when light is traveling infinitely fast (C_∞):

$$R_{C_\infty}: [\text{Pr}, \text{Po}, K, H]. \quad (1)$$

A second conceptual space is constructed where the imagined speed of light is less than C_∞ but much, much greater than the known speed of light c . While Einstein’s Theory of Relativity states that it is not possible to accelerate to c from a slower speed [8], it does not rule out light speeds that are faster than c . This speed will be referred to as c_K . At R_{C_∞} , the four properties that define the conceptual space can be imagined as merging together. As the imagined speed of light is slowed down though, a phenomenon of quantization, depicted by (\downarrow), occurs so that the four properties at R_{C_∞} are further differentiated. Note that this first differentiation, expressed by

($\downarrow R_{C_K} = f(R_{C_\infty})$), implies that reality (R) at c_K , referred to as R_{C_K} , is a function (f) of reality (R) at c_∞ , referred to as R_{C_∞} . This first-differentiation segment is an integral part of Eq. (5), to appear soon. Mathematically, all that may be contained in Pr, Po, K , and H can be thought of as expanding into the corresponding sets S_{Pr} , S_{Po} , S_K , and S_H . This can be summarized in mathematical notation by Eq. (2), which depicts the reality (R) in the conceptual space created where light travels at c_K , (R_{C_K}):

$$R_{C_K}: [S_{Pr}, S_{Po}, S_K, S_H]. \quad (2)$$

A third conceptual space that further quantizes or differentiates these sets can be imagined by light slowing down to an imagined speed much less than c_K but much greater than the known speed of c . This constant speed will be referred to as c_N , and the quantization function depicted soon in (5) can be summarized by ($\downarrow R_{C_N} = f(R_{C_K})$), which suggests that reality (R) at c_N , referred to as R_{C_N} , is a function (f) of reality (R) at c_K , referred to as R_{C_K} . In this conceptual space, the differentiation is imagined to be such that elements from each of the sets S_{Pr} , S_{Po} , S_K , and S_H , combine in unique permutations to create a very large number of unique seeds or functions. Mathematically, these unique functions are summarized as being a function (f) of possible combinations (X) of the four sets. The reality (R) where light travels at the imagined constant speed of c_N is depicted by R_{C_N} in Eq. (3):

$$R_{C_N}: f(S_{Pr} \times S_{Po} \times S_K \times S_H). \quad (3)$$

A final quantization connects the layer R_{C_N} to the layer R_{C_U} . R_{C_U} is the familiar physical layer where light travels at the known speed of c , depicted here by C_U . This quantization, as will appear in (5), is specified by ($\downarrow R_{C_U} = f(R_{C_N})$). Here, the reality (R) at c_U referred to as R_{C_U} , and is a function (f) of reality (R) at c_N , R_{C_N} .

Note that this final quantization results in space, time, energy, and gravity. The practically infinite seeds theorized in (3) suggest an infinite granularity that defines space (S) [9]. Time (T) becomes the means by which what is contained or meant by the function embedded by the granularity of space comes to fruition or can express itself. Energy (E) is related to the process by which the transformation from the subtle seed or meta-function contained in the seed changes to become more material. Gravity (G) suggests an order that is made evident by the way sets of seeds interrelate with each other. Equation (4) summarizes the reality (R) when light travels at the speed c_U :

$$R_{C_U}: [S, T, E, G]. \quad (4)$$

When one considers the nature of space (S), it is also evident that it is a vehicle of the property of knowledge ‘ K ’ envisioned in (1). The nature of time (T) can be seen to be related to the property of power (Po) in (1). Energy can be seen to be related to the property of presence (Pr) in (1). Gravity can be seen to relate to the property of harmony (H) in (1). In other words, there is a symmetry in these conceptual or ‘property’ spaces created by layers of light imagined traveling at different speeds in

which the conceptual spaces are depicting the same property differently. Equation (5) ‘Multi-Layered Fourfold Light-Based Model’ ties the symmetrical, multi-layered model of light together:

$$\left[\begin{array}{l}
 R_{C_\infty} : [\text{Pr}, \text{Po}, K, H] \\
 (\downarrow R_{C_K} = f(R_{C_\infty})) \\
 R_{C_K} : [S_{\text{Pr}}, S_{\text{Po}}, S_K, S_H] \\
 (\downarrow R_{C_N} = f(R_{C_K})) \\
 R_{C_N} : f(S_{\text{Pr}} \times S_{\text{Po}} \times S_K \times S_H) \\
 (\downarrow R_{C_U} = f(R_{C_N})) \\
 R_{C_U} : [S, T, E, G]
 \end{array} \right]_{\text{Light}} \tag{5}$$

From (5), it follows that space, time, energy, and gravity (STEG) emerge when light slows down to *c*. This implies that STEG exists in subtle form before light slows down to *c* and in fact is nothing other than a cosmic-level symmetrical face of the four properties of light proposed in the model. Further, as suggested in related works on genetics [5, 10] STEG is nothing other than a language or code that writes ‘laws’ for subsequent symmetrical emergences [11] such as the electromagnetic spectrum, quantum particles, atoms, molecular plans, and so on.

3 The Fifth Construct in a Quantum-Computational Model of Light

The IEEE article on the oscillating universe [3] provides evidence that a single, universal complex adaptive system (CAS) can be developed from a multi-layered light-based model, due to the underlying dynamism and adaptability facilitated by a system of feedback loops. Quantum computation in a light-based model is found to be the basis of a universal CAS, with adaptation and genetics causing STEG-quantization (note: STEG refers to space–time–energy–gravity) and a change to the STEG-fabric, as well as the dynamics of expansion and contraction of the cosmos, since it is space itself that is affected by STEG dynamics. Ultimately, this leads to a different kind of equilibrium at R_{C_U} which provides the implicit fifth construct, to be discussed in this section, with the ability to establish itself. This implicit construct is evident regardless of the speed with which light is imagined traveling in (5).

Considering (1), light remains one, even as we conceptualize it to have properties of presence, power, knowledge, and harmony. Previous IEEE papers [1] elaborate on the forms in matter and life that these four properties take and will be further discussed in the forthcoming sections. When we see the emergence of quantum particles, atoms, and molecular plans though, one thing that is evident is that all categories of quantum particles are required to act together to create even a single

atom. Further, it is the play of a vast combination of all types of atoms that create the richness of molecules and allow material diversity. Finally, it is the combination of all four molecular plans that in unison allow the living cell to be created.

From the point of view of maintaining symmetry therefore, it can be said that the oneness of light in (1) seeks to maintain that oneness even as it expresses itself as quantum particles, or atoms, or molecular plans. The separate constructs at each level of matter and life need to operate or converge together to create a foundation on which further layers of matter and life can continue to express itself.

This implies that at some point, this hidden though operative oneness will express itself as a dynamic of ‘convergence’ even in each of (1)–(4) and will change the nature of (5). This then will in turn have effects on every layer of matter and life and suggest the emergence of new patterns, as subsequent sections will elaborate.

Equation (6), ‘Fivefold Information When Light Travels Infinitely Fast’, is a modified version of (1), where ‘Cv’ depicts the ‘convergence’ that has now explicitly expressed itself. Hence,

$$R_{C_{\infty}}: [\text{Pr}, \text{Po}, K, H, \text{Cv}]. \quad (6)$$

Subsequently, (2) is transformed into Eq. (7), ‘Transformation into Five Sets’, where S_{Cv} is the set of ‘oneness’. Hence,

$$R_{C_K}: [S_{\text{Pr}}, S_{\text{Po}}, S_K, S_H, S_{Cv}]. \quad (7)$$

Equation (3) transforms into Eq. (8), ‘Creation of Unique Fivefold Combinations’:

$$R_{C_N}: f[S_{\text{Pr}} \times S_{\text{Po}} \times S_K \times S_H \times S_{Cv}]. \quad (8)$$

In (8) hence, bases for a practically infinite number of unique *seeds* or functions, as in (3), continues to be specified.

Equation (4) transforms into Eq. (9), ‘Emergence of Quintergence’, where ‘Q’ refers to a new property—quintergence—that now operates in the cosmos:

$$R_{C_U}: [S, T, E, G, Q]. \quad (9)$$

Quintergence can be thought of as merging or converging of space, time, energy, and gravity with the foundational dynamic of oneness or convergence. It expresses itself as the ability of seeds to now more fully and quickly express the specific knowledge, power, presence, and harmony encapsulated by the seed so that in effect space, time, energy, gravity mingle to allow a different physics to arise.

In other words, the condition for quintergence to arise is when the dynamics of STEG allow the seed to transcend its status as a seed so that the intent that is seeking to express itself has a converged space–time–energy–gravity embryo within which to materialize. The materialization, hence, takes place in a way in which the meta-function presiding over the seed does not suffer marginalization due to the independence of space, time, energy, and gravity as they exist at R_{C_U} . In equation

form, quintergence may be expressed as a union (U) of space (S), time (T), energy (E), gravity (G), which subsequently allows materialization (Mat) of a seed ($f(x)$) formed at R_{C_N} to express its ideal light-form ($f(x)_{R_{C_N}}$) such that (\exists) STEG were all acting as they might at R_{C_N} .

$$\text{Quintergence} : \cup (S, T, E, G) \ni [f(x)_{R_{C_N}} \equiv Mat_{R_{C_U}}]. \tag{10}$$

Hence, the multi-layered fivefold light-based model is summarized by Eq. (11), ‘Multi-Layered Fivefold Light-Based Model’:

$$\left[\begin{array}{l} R_{C_\infty} : [Pr, Po, K, H, Cv] \\ (\downarrow R_{C_K} = f(R_{C_\infty})) \\ R_{C_K} : [S_{Pr}, S_{Po}, S_K, S_H, S_{Cv}] \\ (\downarrow R_{C_N} = f(R_{C_K})) \\ R_{C_N} : f(S_{Pr} \times S_{Po} \times S_K \times S_H \times S_{Cv}) \\ (\downarrow R_{C_U} = f(R_{C_N})) \\ R_{C_U} : [S, T, E, G, Q] \end{array} \right]_{\text{Light}} . \tag{11}$$

Equation (11) now summarizes the dynamics of the CAS that is the cosmos. This is none other than the fuller multi-dimensional dynamics of light so that which is implied by (1) expresses itself more fully by (11). This means that the material layer of existence in which light travels at the known speed of ‘c’ will substantially alter.

The use of quantum computation to bring potential into reality has far-reaching implications. Light’s fivefold nature will become the basis for deepening levels of complexity in matter and life. The electromagnetic (EM) spectrum level, the quantum level, and the layer of atoms are seen as comprising matter, while the layers of atoms and that of the cellular level are seen as comprising life. This will be further explained in the subsequent sections.

4 Sensing the Fivefold Electromagnetic Spectrum

At the first level of emergent complexity, the five characteristics of light are expressed through the EM Spectrum as its wave-range, energy-gradient, the speed of propagation, its potential for mass, and its ability to converge. These are a symmetrical continuation in the expression of light’s fivefold properties. In other words, sensing equipment attuned to the properties of knowledge, power, harmony, presence, and convergence should theoretically be able to sense the functionality implicit in the electromagnetic spectrum.

Here it will be suggested that the property of knowledge expresses itself as the range of waves implicit in the electromagnetic spectrum. Such ‘knowledge’ allows us to comprehend a variety of phenomena tied to the range of waves, from radio waves to gamma rays that can be identified based on the wavelength (λ) of light. This is summarized by Eq. (12): ‘The Knowledge Aspect of the Electromagnetic Spectrum’.

$$\text{Knowledge} \propto [f(\lambda)]. \quad (12)$$

The property of power expresses itself in the electromagnetic spectrum’s range of frequencies. It is known, for example, that power generated by light is not due to an increase in its intensity, but dues to an increase in frequency. This is expressed as Eq. (13), ‘The Power Aspect of the Electromagnetic Spectrum’, where ν is the frequency of the EM Spectrum:

$$\text{Power} \propto h\nu. \quad (13)$$

The principle of harmony, requiring a stable medium in which other phenomena can express themselves, must be related to the constant speed of light at U , c_U , which creates such a stabilizing medium. Equation (14), ‘The Harmony Aspect of the Electromagnetic Spectrum’, summarizes this idea:

$$\text{Harmony} \propto c_U. \quad (14)$$

Equation (15), ‘The Presence Aspect of the Electromagnetic Spectrum’, summarizes the masspotential by combining the equivalences of E and mc^2 , and the $h\nu$ and E , as in:

$$\text{Presence} \propto h\nu/c^2. \quad (15)$$

Equation (16), ‘Convergence Correlation in the Electromagnetic Spectrum’, summarizes the convergence of properties that now becomes possible so that a particular ν , λ , E combination may now also be a carrier of other secondary ν , λ , E combinations. This means that the primary ν , λ , E combination becomes more complex, and, in its essence, this is expressed as integrality (f):

$$\text{Covergence} \propto ff(\nu, \lambda, E). \quad (16)$$

Note that the knowledge aspect can be referred to as ‘wavearchetype’, the power aspect to ‘electro’, the presence aspect to ‘masspotential’, the harmony aspect to ‘magnetic’, and the convergence aspect to ‘pentic’. In this fivefold view therefore the electromagnetic spectrum can be referred to as a wavearchetype-electromagnetic-masspotential-pentic spectrum, and sensitivity to the fifth property of convergence will allow the sensing of related changes in the electromagnetic spectrum.

5 Sensing the Fivefold Structure of Quantum Particles

In the continuing journey of complexification emergent from light, the properties of light symmetrically express themselves in quantum particles. Existing fourfold symmetry is evident in the common fourfold categorization of quantum particles summarized by the Standard Model [12], and in this section, the possibility of a fivefold symmetry is further clarified.

The principle of knowledge, it can be seen, is related to quarks. But why? This can be intuited by considering that the identity and behavior of an atom are related to its atomic number. We know, for example, that an atom with atomic number of 47 will have a set of precise functions—malleability, ductility, superconductivity, resistance to corrosion, and so on—that uniquely identifies it as silver. Further, this element of silver will behave the same way regardless of space and time. But atomic number is a function of the number of protons in the nucleus, and further, all protons are comprised only of quarks. So we can intuit that quarks are a carrier of light’s property of knowledge. Equation (17), ‘Knowledge Aspect at the Quantum Particle Level’, expresses this idea:

$$\text{Knowledge} \propto f(\text{quarks}). \quad (17)$$

The principle of power or energy can be seen to be expressed by leptons. Considering electrons as a surrogate for the class of leptons, we know that the exchange of electrons at the atomic level results in the display of energy [13]. Intuitively we get a sense, therefore, that leptons must be a carrier of light’s property of power or energy. Equation (18), ‘Power Aspect at the Quantum Particle Level’, represents this by mathematical notation:

$$\text{Power|Energy} \propto f(\text{leptons}). \quad (18)$$

The principle of harmony can be seen to be related to the class of quantum particles referred to as bosons. Bosons are known to connect quantum particles together—and connection is related to the idea of harmony. Quarks, for example, require the boson known as a gluon to be bound together. This kind of connection is referred to as the strong nuclear force. The photon is connected to the distribution of the electromagnetic force. W and Z bosons are involved in the creation of lighter quarks and leptons from heavier quarks and leptons. The facilitation of these interactions is due to the presence of bosons that can therefore be surmised to be carriers of the property of harmony. Equation (19), ‘Harmony Aspect at the Quantum Particle Level’, summarizes this:

$$\text{Harmony} \propto f(\text{bosons}). \quad (19)$$

The principle of presence can be seen to be related to the Higgs boson. This is because any quantum particle with mass must interact with the Higgs-field via

the Higgs boson in order to acquire mass. The other fundamental particle discovered is responsible for granting quarks mass. Equation (20), ‘Presence Aspect at the Quantum Particle Level’, summarizes this relationship:

$$\text{Presence} \propto f(\text{Higgs_boson}). \quad (20)$$

Just as each of the other categories of quantum particles has multiple kinds of ‘sub’-particles, note that recent research at CERN [14] suggests that the Higgs boson family may also have multiple particles.

Equation (11) then predicts a fifth structure at the quantum particle level—in essence a fifth type of quantum particle, once the cosmos has reached a stage where there is a dynamic equilibrium expressed by quintergence (10). This is summarized by Eq. (21). ‘Convergence Aspect at the Quantum Particle Level’, as a new category of quantum particles—amorems:

$$\text{Convergence} \propto f(\text{Amorems}). \quad (21)$$

Sensitivity to the fifth property of convergence will allow the sensing of related changes in quantum particles.

6 Sensing the Fivefold Structure of the Periodic Table

All atoms in the periodic table can be classified as belonging to either the p-Group, d-Group, s-Group, or f-Group. This section will elaborate on the continuing fivefold symmetry of light with the possible emergence of a fifth grouping of atoms.

The principle of knowledge can be seen to be related to p-Group atoms. For example, the p-Group has the element carbon from which life and subsequently thinking life emerges. That is a broad claim that gets further qualified when it is seen that the element silicon, also in column 14 of the periodic table, sits just below and adjacent to it. This means that many properties are shared. But more importantly computing or ‘thinking’ machinery is known to have arisen from silicon. Further, the p-Group also contains many archetypes such as metals, metalloids, non-metals, halogens, and noble gas sub-groups. In this way of seeing the ‘knowledge’ of what exists elsewhere in the periodic table exists as archetypes in the p-Group. But further, the single probability cloud that exists around the nucleus in an s-Group atom, has become two on either side of the nucleus. The idea of inherent duality existing along several axes creates many more form-based switches than exists in the s-Group element, which in a manner of seeing attracts a larger number of archetypes or possible functions into the p-Group. This too enhances the idea that the p-Group is a carrier of the property of knowledge. Equation (22), ‘Knowledge Aspect at the Level of Atoms’, summarizes this relationship:

$$\text{Knowledge} \propto f(p_orbital). \quad (22)$$

The principle of presence can be seen to be related to the d-Group. First, the probability spaces around the nucleus have in this case increased to four. Four lobes can potentially be thought of as creating four end-points of what is known to be an inherently stable structure [15]—the tetrahedron. This idea seems to be reinforced by Crystal Field Theory [16]. But further, it is elements such as platinum, iron, cobalt, titanium, copper, and zinc, among others, known for corrosion resistance, strength, and hardness, that comprise elements in the d-Group. These elements are known to create the infrastructure that animates the solid world of cities and connections between cities around us. Equation (23), ‘Presence Aspect at the Level of Atoms’, summarizes this relationship:

$$\text{Presence} \propto f(d_{\text{orbital}}). \quad (23)$$

The principle of power can be seen to be related to the s-Group. Philosophically the single probability cloud around the nucleus, depicting the equal likelihood of an electron in an s-Group atom being anywhere around the nucleus, suggests that such atoms are in a sense pioneers, setting the stage for the emergence of other atoms in due course. This seems to be borne out by the fact that elements such as hydrogen and helium belong to this group. These are earlier atoms that are also the most prevalent comprising 98% of the universe [17]. But further, s-Group elements have been referred to as ‘violent world’ [18] due to the high degree of electropositivity which allows the easy release of electrons and creation of positive ions typical of alkali metals. It is also known that a star shines [19] due to nuclear fusion or the transmutation of hydrogen to helium. Equation (24), ‘Power Aspect at the Level of Atoms’, summarizes this relationship:

$$\text{Power} \propto f(s_{\text{orbital}}). \quad (24)$$

The principle of harmony can be seen to be related to the f-Group. Elements in this group have six probability lobes in seven different planes around the nucleus and can be seen as an experiment in establishing more bonds in smaller spaces. This becomes an experiment in collectivity, that is further reinforced by the fact that lanthanides are known to be stable even with their higher atomic numbers. Actinides, which also have higher atomic numbers are by contrast radioactive and can be thought of as showing conditions when collectivity needs to split into smaller collectivities. Equation (25), ‘Harmony Aspect at the Level of Atoms’, summarizes this relationship:

$$\text{Harmony} \propto f(f_{\text{orbital}}). \quad (25)$$

Once the cosmos has settled into a new equilibrium displaying the property of quintergence, the dynamics of atoms will complexify perhaps allowing a-Shell atoms to also form, as summarized by Eq. (26), ‘Convergence Aspect at the Level of Atoms’:

$$\text{Convergence} \propto f(a_{\text{orbital}}). \quad (26)$$

A property of such a-Group atoms is that the nature of chemical reactions will change. Restrictions to do with bonding of atoms to form molecules will change once a-Group atoms are added to the mix. Sensitivity to the fifth property of convergence will allow the sensing of related changes in atoms.

7 Sensing the Fivefold Structure of Molecular Plans in Cells

At the cellular level, all living cells—whether plant, animal, or human—are made up of nucleic acids, proteins, lipids, and polysaccharides [20].

Nucleic acids are responsible for encoding, storing, transmitting, and maintaining the hereditary information that is essential for keeping a cell alive. They act as the cell's librarians, containing instructions on how to create proteins and when they must be produced. Furthermore, nucleic acids are the vessels by which knowledge is passed on to future generations of cells. Therefore, nucleic acids can be regarded as a manifestation of knowledge at the cellular level, as expressed by Eq. (27): 'Knowledge Correlation at the Cellular Level'.

$$S_{K(\text{cell})} \ni [\text{Knowl.}, \text{Wisdom}, \text{Law Making}, \text{Spread of Knowl} \dots]. \quad (27)$$

This relationship may also be summarized as a simpler version by Eq. (28), 'Simple Version of Knowledge Correlation at the Cellular Level':

$$\text{Knowledge} \propto f(\text{nucleic acids}). \quad (28)$$

Proteins are essential for the functioning of cells. They can be found in every part of the cell, and further, come in a variety of shapes and sizes. Some proteins are designed to form specific shapes, such as tubes, rods, nets, hollow spheres, among others, while others act as motors using energy to move and flex [20]. Additionally, many proteins also act as catalysts to enhance chemical reactions and transfer and transform atoms as needed. With such a wide range of abilities, proteins are used to complete nearly all tasks in the cell, with estimates of up to 30,000 different kinds of proteins in the human cell.

Proteins can be seen as a manifestation of dedication and hard work at the cellular level, as demonstrated by Eq. (29): 'Presence Correlation at the Cellular Level'. They are known for their diligence and determination, contributing to the perfect functioning of cells. Therefore, it can be said that proteins are essential to provide a service.

$$S_{Pr(\text{cell})} \ni [\text{Service}, \text{Perfection}, \text{Diligence}, \text{Perseverance}, \dots]. \quad (29)$$

This relationship may also be summarized more simply by Eq. (30), 'Simple Version of Presence Correlation at the Cellular Level':

$$\text{Presence} \propto f(\text{proteins}). \quad (30)$$

By themselves Lipids are individual small molecules. However, when grouped together, they are able to form the largest structures of the cell. When placed in water, they easily aggregate into large sheets that are not only waterproof, but also create natural boundaries. These boundaries enable concentrated interactions to take place within the cell and allow for tasks, such as the nucleus and mitochondria, to be contained in lipid-defined compartments. Moreover, each cell is surrounded and contained by a lipid-defined boundary.

Lipids can be seen as the facilitators of relationship and harmony within the cell, promoting collaboration and specialization, and potentially demonstrating early forms of compassion and love. This idea of harmony is represented by Eq. (31), ‘Harmony Correlation at the Cellular Level’:

$$S_{H(\text{cell})} \ni [\text{Love, Compassion, Harmony, Relationship} \dots]. \quad (31)$$

This relationship may also be summarized more simply, by Eq. (32), ‘Simple Version of Harmony Correlation at the Cellular Level’:

$$\text{Harmony} \propto f(\text{lipids}). \quad (32)$$

Polysaccharides, in contrast to individual and small lipids, are long branched chains of sugar molecules covered with hydroxyl groups that come together to form storage containers. This makes them useful for storing the cell’s energy, as well as building some of the sturdiest structures found in nature. For example, the exoskeleton of an insect is made of a type of polysaccharide.

Polysaccharides are known to provide energy and strength, enabling cells to undertake new adventures. This is evidenced by Eq. (33), ‘Power Correlation at the Cellular Level’, which illustrates how polysaccharides are sources of power and courage at the cellular level.

$$S_{Po(\text{cell})} \ni [\text{Power, Courage, Adventure, Justice,} \dots]. \quad (33)$$

This relationship may also be summarized more simply by Eq. (34), ‘Simple Version of Power Correlation at the Cellular Level’:

$$\text{Power} \propto f(\text{polysaccharides}). \quad (34)$$

Anandam is proposed as a possible future molecular plan that allows cells to leverage the property of quintergence. As a result of this molecular plan, the cell will be able to take the shape or function of any of the molecular plans within it or combine these in unforeseen ways. As a result, properties such as convergence, shape-shifting, lightness, heaviness, and so on get materialized. This precipitation of convergence at the cellular level is summarized by Eq. (35), ‘Convergence Correlation at the Cellular Level’:

$$S_{Cv(\text{cell})} \ni [\text{Combination, ShapeShifting, Expansion, Lightness, ...}]. \quad (35)$$

This relationship may also be summarized more simply by Eq. (36), ‘Simple Version of Conversion Correlation at the Cellular Level’:

$$\text{Convergence} \propto f(\text{andandams}). \quad (36)$$

Sensitivity to the fifth property of convergence will allow the sensing of related changes in molecular plans.

8 Summary and Conclusion

The fourfold model of light prototypes cosmos as a CAS. In this model, all surfacings of matter and life are seen as fourfold constructs whose architecture is determined by the four properties of light that maintain their symmetry regardless of the layer in which, or how light shows up.

The iterative, computational nature of the model, based on a system of feedback loops, however, indicates a potential or future phase in which the fourfoldness in the highest layer of light or the source conceptual space is more easily reflected in the physical layer where materialization takes place. When this happens then the ‘system’ as it were morphs and pushes forward an implicit fifth aspect that changes the mathematics of the model itself, and subsequently all architectures emanating from light.

What this model is predicting, and what this paper is highlighting, is that sensitivity to the fifth construct, referred to as ‘convergence’, will allow sensing corresponding changes to the very nature of macro-parameters—such as space–time, energy, gravity—operative at the cosmic level, to recognize dynamics related to a fifth possible macro-parameter, that of quintergence. Quintergence is predicted to alter the physics of space, time, energy, gravity so that these act as a more unified embryo within which dynamics suggested by the antecedent layers in the model are more fully reflected at the material level.

As this happens, then sensitivity to corresponding changes in the physics of the electromagnetic spectrum, and of quantum particles increases. Sensing possible changes to the fourfoldness of the electromagnetic spectrum as it displays five-foldness becomes possible. Hence, with the propagation of convergence, the electromagnetic-wavearchetype-masspotential aspects are accompanied by a ‘pentic’ aspect that allows complexification of electromagnetic function in terms of simultaneous combination of wavelength, frequency, and energy dimensions. A model attuned to this will allow changes driven by emerging fivefoldness to be sensed.

The pressure of meta-function of light continuing to express itself materially is then also predicted to bring into being a fifth type of ‘amorem’ quantum particle by which the mathematics of quantum particle combination goes through a radical change. As a result, a host of new atoms is also predicted to come into being. A

model attuned to emerging fivefoldness at the quantum particle level will therefore also allow such changes to be sensed.

If the realm of macro-parameters operative at the level of cosmos—the electromagnetic spectrum and quantum particles—is considered the domain of physics, it is clear then that the change in the model of light suggests that there will be emergent changes in physics itself. But further, it also suggests as substantive, changes to the realms of chemistry and biology.

The change to chemistry follows the change in structure of the periodic table itself. It is predicted that a-Group atoms may be created, that serve a similar function as did amorem particles at the quantum particle level. Basically, it will allow the limits of chemical reaction to be transcended so that a whole new host of atom combinations will become possible, allowing new molecules to come into being.

Equally, there are going to be emerging changes to biology itself. This follows from the suggestion that a fifth type of anandam molecular plan may also begin to manifest in living cells. This will give cells extraordinary capabilities, levels ahead of the already extraordinary capabilities of cells. Attunement to this will allow emergent fivefold changes at the level of cells to also be sensed.

References

1. Malik P. IEEE, Author Page. <https://ieeexplore.ieee.org/author/37086022058>. Last accessed 24 Oct 2023
2. Malik P. Amazon, Author Page. https://www.amazon.com/Pravir-Malik/e/B002JVAEZE%3Fref=dbs_a_mng_rwt_scns_share. Last accessed 24 Oct 2023
3. Malik P (2022) The role of a light-based quantum computational model in the creation of an oscillating universe. In: 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, Las Vegas, NV, USA, pp 0953–0959
4. Malik P, Pretorius L (2019) An algorithm for the emergence of life based on a multi-layered symmetry-based model of light. In: 2019 IEEE 9th annual computing and communication workshop and conference (CCWC), pp 0034–0041. IEEE, Las Vegas, NV, USA
5. Malik P (2020) A light-based quantum-computational model of genetics. In: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). IEEE, Vancouver, pp 1–8
6. Einstein A (1995) *Relativity: the special and general theory*. Broadway Books, New York
7. Wilczek F (2016) *A beautiful question: finding nature's deep design*. Penguin Books, New York
8. Perkowitz S (2011) *Slow light*. Imperial College Press, London
9. Rovelli C (2016) *Reality is not what it seems*. Penguin Random House, New York
10. Malik P (2019) *The origin and possibilities of genetics*. Possibilities Publishing, San Francisco
11. Malik P (2021) The emergence of quaternary-based computational-strata from a symmetrical. *IJSSST* 22(1):Paper 4
12. The Standard Model. <https://home.cern/science/physics/standard-model>. Last accessed 24 Oct 2023
13. Arabatzis T (2006) *Representing electrons: a biological approach to theoretical entities*. University of Chicago, Chicago
14. Overbye D (2015) Physicists in Europe find tantalizing hints of a mysterious new particle. *New York Times*, New York

15. Fuller B (1982) Synergetics: explorations in the geometry of thinking. MacMillan Publishing Co., New York
16. Crystal Field Theory. [https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Supplemental_Modules_and_Websites_\(Inorganic_Chemistry\)/Crystal_Field_Theory/Crystal_Field_Theory](https://chem.libretexts.org/Bookshelves/Inorganic_Chemistry/Supplemental_Modules_and_Websites_(Inorganic_Chemistry)/Crystal_Field_Theory/Crystal_Field_Theory). Last accessed 24 Oct 2023
17. Heiserman D (1991) Exploring chemical elements and their compounds. McGraw-Hill, New York
18. Tweed M (2003) Essential elements: atoms, quarks, and the periodic table. Walker & Company, New York
19. Gray T (2009) The elements: a visual exploration of every known atom in the universe. Black Dog & Levental Publishers, New York
20. Goodsell D (2010) The machinery of life. Springer, New York




Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



VOTUM: Secure and Transparent E-Voting System



Joaquin Egocheaga , William Angulo , and Cesar Salas 

Abstract Elections are an essential part of citizens' rights, and they are also conducted in universities and colleges to ensure transparent selection of ideal authorities while preventing identity fraud and information loss among voters. It is worth noting that Internet voting has gained significant attention in recent years, with many organizations worldwide planning to experiment with and implement it. To address these challenges, we propose VOTUM, a free fraud e-voting system that incorporates two authentication methods: facial recognition and one-time password (OTP). Additionally, the system employs two cryptographic algorithms to encrypt voters' information throughout the voting process and generates a unique code to verify the successful casting of votes. VOTUM's design is creative, flexible, colorful, and animated, aiming to encourage students and professors to fulfill their civic duty by participating in elections. Through interviews conducted with 31 students and university professors, we achieved a 90% trust level and a 15% margin of error to assess satisfaction with transparency, trust, and user experience within the VOTUM system. The results indicated a satisfaction level of over 90%, showing the significant contribution of this research in enhancing trust and transparency in the voting processes of universities and colleges.

Keywords E-voting · Facial recognition · One-time password · Security voting · Amazon Rekognition · Secure e-voting · Transparency voting · Face recognition voting

1 Introduction

Voting is one democratic activity that must be transparent because it is the fundamental mechanism for people to choose their representatives [1]. In a democratic system, only eligible members can vote once, and no one can change, delete, or

J. Egocheaga · W. Angulo · C. Salas (✉)
Peruvian University of Applied Sciences, Lima, Peru
e-mail: cesar.salas@upc.edu.pe

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_8

duplicate the votes. The tally should exactly reflect the election results and the traditional voting system. In this article, it is proposed to treat the elections at the university level with a portable and convenient solution that allows users to cast their vote at any time, from their own computing device connected to the Internet. The aim is to digitize the processes of election registration, voting, verification, and vote counting during the electoral cycle [2].

The implementation of electronic voting faces the challenges of the modernization of the electoral process, the training of people, and their confidence in the process and the electoral authorities. The main advantages of electronic voting are security, speed, reliability, and the possibility of correction [3]. Furthermore, an electronic voting system must have the following security requirements, which are eligibility, uniqueness, privacy, and accuracy [4].

With that in mind, our paper proposes a scheme of an electronic voting system that seeks to reduce the cases of fraud in university elections, using facial recognition and one-time password (OTP) as a way to ensure voter authentication and encryption of information for the security of the vote within a non-face-to-face electronic voting. Also, a development will be carried out for experimentation at the functional level to validate the business logic proposed in this paper and test the effectiveness of our proposed factors to ensure a secure voting process, so that it can be scalable for a more elaborate development for elections in a more real environment.

To achieve the proposed goal, cryptography and biometric authentication have been suggested as potential solutions to address the security and transparency concerns. However, a more comprehensive approach is necessary to ensure the safeguarding of voter registrations and the voting process, while also maintaining voter privacy. “If a cryptographic scheme is secure despite being publicly known, it is much more reliable” [5]. Furthermore, biometric authentication “is expected to replace the role of conventional techniques as access control to provide security for valuable data in avoiding attacks or fraud” [6]. The remainder of this paper is structured as follows: Sect. 2 outlines the fundamental components of the proposed system. Section 3 delves into the design aspects of our e-voting system. Section 4 validates the hypothesis through experiments and presents the obtained results. Finally, Sect. 5 summarizes the conclusions drawn from the study and highlights potential avenues for future research.

The results showed a 90% acceptance rate by the students interviewed regarding the usability and data security offered by VOTUM during the voting process.

2 Related Works

In this section, we will review some work related to e-voting systems to analyze different ideas and considerations that other authors had. Also, we will review papers related to biometric methods and schemas that other authors have proposed.

For the development of an electoral process, it is essential that it be secure and complete in order to guarantee the anonymity of the vote and accurate accounting.

This is why Yining Liu proposes an electronic voting scheme based on the secret exchange k-anonymity which consists in the fact that the receipts of voters' votes are published when they are generated, so it is not possible to distinguish individually to whom they belong, but they can accurately verify the result [7]. Moreover, Oprea and other authors agree with this idea and mention that human counting is prone to have several errors as it is also more costly so they propose an electronic voting system architecture based on blockchain tables which allows that the recorded data cannot be altered and to avoid potential fraud so that it is in charge of voting at the university level in which it is evident that it is based on principles such as voting only once, end-to-end verifiable, and auditable [8]. Also according to Johari and other authors consider that elections are a right that every citizen has and in order to encourage voter suffrage they propose a secure electronic voting system implemented in the cloud using microservices with node technologies using Azure Service Fabric [9].

On the other hand, in the aspect of security in elections today is to guarantee the judgment and realness of the vote can be seen from the use of biometric methods which allows elections to be more transparent, so Masud Ahmad proposes a voting system which requires a record of biometric information of thumbs prior to the elections so that once they start these, voters can be authenticated and to cast their vote in full [10]. In addition, according to the article by Vasanthi and Seetharaman, facial recognition is considered a prominent method in biotechnology and is in high demand for security, so a method of facial recognition through biometric data is proposed, based on multivariate correlations using geometric points and visual characteristics of the face, such as color and texture, which are represented with arrays [11].

3 Main Contribution

In this section, the main concepts and the principal contribution of this paper will be presented.

3.1 Method

The main contribution of this paper consists of the faculty of voters who can be authenticated using the face biometric authentications provided by Amazon Rekognition SendGrid for OTP. Furthermore, to increase voter participation and improve election results in a way that addresses the challenges associated with traditional voting practices [12]. To achieve it, we use an attractive design for the mobile app to make voters feel comfortable while voting.

This paper VOTUM is built to support university electoral processes considering prerequisites such as protection, qualification, and unquestionable status. The proposed system aims to achieve secure digital voting without compromising its convenience. Designed for mobile devices to facilitate user participation, with

measures such as facial recognition identification to prevent the occurrence of multiple votes from the same individual.

Architecture Overview of Solution Architecture

First summarize the components, the roles of the users and staff members who are involved. VOTUM belongs to the category of remote electronic voting systems, this class of digital voting systems involves the use of PCs, laptops, tablets, and other off-the-shelf gadgets. The operating systems where web app VOTUM can be used are Windows, MacOS, and Linux. Furthermore, VOTUM mobile app can be used on Android and IOS devices.

This class of digital voting frameworks includes the use of phones and PCs. The suggested solution eradicates the requirement for these individuals to verify both the eligibility of users to vote and the accuracy of the voting operations.

Actors of System

There are two kinds of actors who have interactions with the VOTUM system, staff members and voters who can be students or professors. Every voter is previously registered in the system with their university code by the staff members.

Staff individuals can be classified into two bunches based on their parts and obligations inside particular election stages. To do so, they connected with the central framework and VOTUM administrations:

- Staff involved in preliminary operations are responsible for tasks associated with preparing and verifying the voter list, as well as creating digital voter cards.
- Staff directing all voting operations physically are entrusted with guaranteeing the correct conduct of the voting process, which incorporates confirming the proper working of the voting system.

Elements of System

The by and large framework incorporates a few computer program applications that run at the same time on numerous physical machines, as appeared in Fig. 1. In specific, the proposed design includes the utilization of an administration panel, voting device, central system, facial recognition service, mail service and Identity Access Management (IAM). The facial recognition cloud service used is Amazon Rekognition. It is a cloud service that provides a simple and easy-to-use API that allows you to quickly analyze any image or video file stored in Amazon S3 to determine the level of similarity between them [13]. The Mail Service used is SendGrid is a provider of SMTP services in the cloud that allows the sending of electronic mail that gives the facility of not needing own mail servers [14].

Design Requirements and Assumptions

In this section, the design requirements, and general and security assumptions of VOTUM are described.

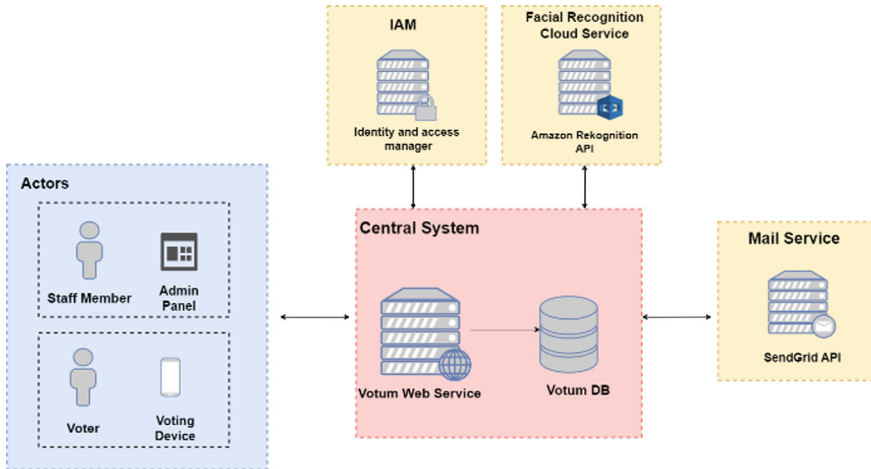


Fig. 1 VOTUM architecture overview

- **Privacy:** User’s privacy is guaranteed if it is not possible in any way to trace a vote back to the voter who cast it. To guarantee that, our solution meticulously separates the identification and voting phases, abstaining from the collection of any user-related information. During the whole voting process, no element of the system possesses knowledge of both the identity of voters and the content of their votes. The formal analysis that verifies VOTUM’s adherence to this principle demonstrates that each vote is entirely unlinked from the identity of the voter.
- **Authenticity:** The voter must provide their required authentication factors. Authorized voters only are provided with access to the voting page.
- **Fairness:** No entity can gain any knowledge about the partial ballot before the end of the polling.

Secure Data Storage Using Encryption

Our project applies to guarantee the security and anonymity of registered voter codes through encryption methods. During the phases prior to sending the vote, the voter’s code is encrypted with SHA256 in the database according to the vote, which allows it not to be decrypted. However, for the business operations of the voting process, in order to correctly perform the filtering, the voter’s code that is stored locally in the mobile device obtained when logging in is used, so that it can be encrypted in each business operation during the voting process and can be used to filter the votes as expected and at no time the original code is revealed. Once the voter has been authenticated and proceeds to submit his or her vote, the voter code that was used for filtering the votes is replaced by a new encryption of the original voter code in which hybrid encryption is used. The hybrid encryption consists of the encryption of a plain text, in this case the university student code using in first instance the DES

encryption algorithm. DES algorithm is an encryption algorithm consisting of a 64-bit block cipher with a 56-bit key [15]. When the university student code is encrypted, it is encrypted again using the RSA encryption algorithm. RSA algorithm consists of a three-step encryption algorithm, key generation, encryption, and encryption description. This requires a vast amount of 2048-bit prime numbers [16]. Finally, this last generated encrypted code is stored in the database to make the system recognize the owner of every vote when an election finishes.

The following steps describe how hybrid encryption is used.

- (a) Voter sends its vote to the server.
- (b) Server encrypts voter code by hybrid algorithm.
- (c) Database replace SHA-256 voter code with hybrid algorithm voter code.

Face Recognition Using Amazon Rekognition

Facial recognition is used to guarantee the authenticity and transparency of the vote through the voter's identity. The Amazon Rekognition service is used to achieve this goal. The following steps describe the facial authentication process in the VOTUM system.

1. The voter during the facial authentication phase has a picture of his or her face taken and sent to the VOTUM system. Likewise, the VOTUM system requests to the educational institution's server (Active Directory), the link to host the voter's image.
2. The VOTUM system retrieves the voter's image from wherever it is hosted.
3. VOTUM sends the two images to the Amazon Rekognition service API and receives the results of the comparison to determine if the vote should be authenticated.

VOTUM

To address the unique attributes of each election, our system functions through three specific phases, concerning to the operations conducted prior to, during, and following the voting process.

Pre-voting phase: In the phase prior to casting the vote, it is necessary to carry out other processes that are necessary for the voting phase to take place. The process of loading the voters belonging to the educational institution is essential, since, in this way, it is possible to know all the members of the institution who are entitled to vote.

1. The process of loading is done through a massive upload of an XLSX file. The other process prior to the voting phase is the creation of elections, which consists of the creation of the electoral processes with their respective information. Likewise, the people participating in the election as well as the parties involved are assigned.
2. XLSX files are sent to the web app server to get the data and start processing it with the business logic. Furthermore, data is sent to the database.
3. Database registers the new election with its participants and candidates.

Voting phase: During the voting phase, voters can express their preferences easily. In this process, the voter casts his favorite candidates and gets the credential via a secure manner such as face to face.

This process starts when the voter is connected to the system and has clicked on Login Button.

- Once a voter's identity has been verified, VOTUM no longer needs information about the user.
- In order to allow only authorized users to vote, the system verifies whether they are entitled to vote, and they have not previously voted.
- To ensure privacy, votes are encrypted after the voter submits his choice.
- During the voting stage, the voter can select the election of his or her preference. Subsequently, the voter may select the candidate of his choice and then confirm his choice.
- The authentication process initially prioritizes the use of facial authentication, in case the result is not successful, OTP authentication is used.
- The final step is that once the vote is authenticated, the voter must send the vote through a button on the interface.

Post-voting phase:

The post-vote casting phase is the audit phase, which consists of verifying the conformity of the votes based on a voting code that only voters have once they have cast their vote. This process seeks to guarantee that the votes cast are in conformity with the voter's choice.

The main contributions of this paper are as follows:

- Unlike most of the remote voting systems VOTUM applies facial recognition technology as a security measure and voter authentication to ensure the identity of the voter.
- VOTUM applies data encryption to ensure the security and anonymity of registered voter codes through encryption methods. During the phases of voting, the voter's code is encrypted with SHA256 and with the hybrid algorithm in different phases.

Hypothesis

VOTUM is expected to be a solution for university elections that provides all the necessary security and transparency properties to ensure user privacy and fairness of elections. In addition, it will be scalable to other voting levels.

4 Experiments

This section will discuss the experiments our project has undergone, as well as what is needed to replicate said experiments and a discussion of the results obtained after this process.

4.1 Experiments Protocol

In this subsection, details the configuration of the environment in which the experiments were performed. During a period of one week, a mock election was conducted to test how secure and transparent the proposed e-voting system is in the voting process. For which has been defined a population of 1,347,600 [17] considering the sample size was defined a sample space of 31, which are students between the ages of 18–22 years. The validations and tests of the system were made at the Peruvian University for Applied Sciences located in Lima, Perú. After each voting procedure, participants were given a survey consisting of 10 questions of two types closed and using Likert scale (1–5), which was designed to ask voters for their satisfaction in using the system and to encourage them to make suggestions related to the security of the system.

4.2 Results

In this subsection, the experiments carried out and the results obtained in each of these are detailed. For statistical purposes, asked students about their educational background, i.e., whether their studies related to scientific or humanities areas. The Likert scale is used to determine the level of satisfaction while using the mobile app.

One of the questions asked in the survey was whether the application adapts to the problem explained by the researcher, to evaluate the hypothesis variable about the security and transparency. A mean of 4.6 was obtained Also a confidence level of 90% was obtained with the asked question “Do you trust in voting,” to evaluate trust in electronic voting systems. A mean of 0.96 was obtained in the question of the survey was whether they considered the VOTUM application a secure application for electronic elections. Finally, 93.3% of the sample data found facial authentication and OTP authentication to be secure and 96.7% of the respondents agreed to recommend the application to other users.

4.3 Discussion

In this subsection, the results obtained in the previous section are detailed and discussed.

The purpose of the security variable is to determine how difficult it is to access the voter's data and how protected they are, as well as the voting data. On the other hand, the transparency metric intends to validate how unlikely it is that a person votes more than once and that the registered person is the one who votes and is not impersonated by another person. Considering these variables, clear evidence can be obtained through the statistical data that the interviewees are satisfied with the security of the application since an average score of 4.7 was obtained from 30 interviewees with a Likert scale. On the transparency side, the interviewees had a positive reaction regarding the, so it was obtained that 93.3% of the interviewees felt very secure with the authentication methods, so it can be affirmed that the transparency variable was achieved.

The purpose of validating the application based on the problems is to determine if it has been able to mitigate them. Based on the results, 93.4% of the sample population considers that the system was able to adapt to the problem of being able to carry out a secure and transparent election process. VOTUM can be considered that it does adapt to the problem.

5 Conclusions and Perspectives

VOTUM is a new electronic voting system suitable for any university election and based on, first, a suitable encryption algorithm in the voting management process; second, on the security and authentication of voting data and voter. Furthermore, the convenience of election procedures is heightened in terms of security and engagement. This is attributed to VOTUM being entirely open source, allowing security experts worldwide to assess the efficacy and resilience of this electronic voting system. Finally, VOTUM is successfully accepted by people because a high percentage of more than 90% of a survey of 31 people used the application and gave their approval in the fields of usability, security, and solution to the problem, so that they could effectively cast their vote. As future work, for VOTUM to be used in real elections, we have as future work the use of blockchain so that the registered data is immutable. In addition, another improvement that should be applied is the use of load balancer, evolving the system to microservices so that each service of the project is independent for both the administrator and voter. Finally, use a robust database manager to be able to support large data loads.

References

1. Ayala A, Daniel S, Vinzoneo M (2014) Los procesos electorales y las nuevas tecnologías. www.juridicas.unam.mx <http://biblio.juridicas.unam.mx>
2. Adeshina SA, Ojo A (2020) Factors for e-voting adoption - analysis of general elections in Nigeria. *Govern Inform Quart* 37(3). <https://doi.org/10.1016/j.giq.2017.09.006>
3. Seifert Bonifaz M (2014) Percepciones de los peruanos sobre el voto electrónico presencial. En: *Elecciones* 13:11–28. www.onpe.gob.pe
4. Issa H (2013) International conference on computer applications technology: ICCAT' 2013, 20–22 January: abstract proceedings: Sousse Tunisia. IEEE
5. Agate V, De Paola A, Ferraro P, Lo Re G, Morana M (2021) SecureBallot: a secure open source e-voting system. *J Netw Comput Appl*. <https://doi.org/10.1016/j.jnca.2021.103165>
6. Wati V, Kusriani K, Al Fatta H, Kapoor N (2021) Security of facial biometric authentication for attendance system. *Multimed Tools Appl* 80(15):23625–23646. <https://doi.org/10.1007/s11042-020-10246-4>
7. Liu Y (n.d.) Esquema de votación electrónica que utiliza el uso compartido de secretos y el anonimato K. red mundial. Retrieved 6 Oct 2022, from <https://sci-hub.se/https://doi.org/10.1007/s11280-018-0575-0>
8. Oprea SV, Bara A, Andreescu AI, Cristescu MP (2023) Conceptual architecture of a blockchain solution for e-voting in elections at the university level. *IEEE Access* 11:18461–18474. <https://doi.org/10.1109/ACCESS.2023.3247964>
9. Johari R, Kaur A, Hashim M, Rai PK, Gupta K (2022) SEVA: secure e-voting application in cyber physical system. *Cyber-Phys Syst* 8(1):1–31. <https://doi.org/10.1080/23335777.2020.1837250>
10. Ahmad M, Ur Rehman A, Ayub N, Khan MA, Hameed A, Yetgin H, Pakhtunkhwa K (2020) Security, usability, and biometric authentication scheme for electronic voting using multiple keys. *Res Art Int J Distrib Sens Netw* 7. <https://doi.org/10.1177/15501477220944025>
11. Vasanthi M, Seetharaman K (2022) Facial image recognition for biometric authentication systems using a combination of geometrical feature points and low-level visual features. *J King Saud Univ Comput Inform Sci* 34(7):4109–4121. <https://doi.org/10.1016/j.jksuci.2020.11.028>
12. Mahamat M, Adeshina SA, Arreytambe T, Institute of Electrical and Electronics Engineers (2014) Proceedings of the 11th international conference on electronics, computer and computation (ICECCO'14) : international conference, September 29–October 1, 2014 : Abuja, Nigeria
13. AWS (2023) Qué es Amazon Rekognition?. https://docs.aws.amazon.com/es_es/rekognition/latest/dg/what-is.html
14. SendGrid (2016) What is SendGrid? SendGrid (2023). <https://sendgrid.com/wp-content/uploads/2016/09/SendGrid-Implementation-Review.pdf>
15. Zeebaree SRM (2020) DES encryption and decryption algorithm implementation based on FPGA. *Indonesian J Electr Eng Comput Sci* 18(2):774–781. <https://doi.org/10.11591/ijeecs.v18.i2.pp774-781>
16. Yao F (2021) Hybrid encryption scheme for hospital financial data based on Noekeon Algorithm. *Secur Commun Netw*. <https://doi.org/10.1155/2021/7578752>
17. INEI (2021) INEI (20230). <https://m.inei.gob.pe/estadisticas/indice-tematico/university-tuition/>
18. Smith M, Miller S (2022) The ethical application of biometric facial recognition technology. *AI & Soc* 37(1):167–175. <https://doi.org/10.1007/s00146-021-01199-9>
19. Kumar M, Chand S, Katti CP (2020) A secure end-to-end verifiable internet-voting system using identity-based blind signature. *IEEE Syst J* 14(2):2032–2041. <https://doi.org/10.1109/JSYST.2019.2940474>

20. Llanos J, Coral W, Alarcon A, Cruz J, Ramirez J (2019) Electronic voting system for universities in Colombia. In: ICINCO 2019—proceedings of the 16th international conference on informatics in control, automation and robotics, vol 1, pp 325–332. <https://doi.org/10.5220/0007929103250332>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Simulation and Fault Diagnostics Using I–V and P–V Curve Tracing



Kabelo Mashiloane, Peet F. Le Roux, and Coneth G. Richards

Abstract Localization of problems continues to be very difficult, especially in large-scale photovoltaic (PV) systems. Especially for small-scale PV plants, the layout of PV systems significantly impacts the efficiency of detection systems. Due to faults occurring within PV arrays, this paper aims to highlight the value of fault detection in PV systems through I–V curve features. This is achieved by simulating models using MATLAB/Simulink of normal and faulty operations. Investigating faults in solar PV arrays is critical in improving PV systems' dependability, effectiveness, and safety. A quick and efficient way to determine the actual performance of solar PV modules or strings is to use the I–V curve. To guarantee a PV installation's operational dependability, fault detection is essential. Identifying and detecting faults, particularly in installations of solar systems, remains a major difficulty. The paper proposes an effective fault detection and identification method that uses PV array I–V curve analysis.

Keywords Fault detection · I–V curve · MATLAB/Simulink · PV systems · Reliability

1 Introduction

The need to include renewable energy in the electrical system is growing due to widespread worries about climate change. Because photovoltaic (PV) power generation is secure, dependable, silent, ecologically friendly, and unlimited in resource distribution. It has a high energy quality, quick construction time, and many benefits over conventional power generation modes. Several nations have seen increases in PV energy production due to cost reductions [1].

A PV power plant can also be constructed by joining many PV modules in a series or parallel configuration because PV is a scalable and modular technology. Any issue

K. Mashiloane (✉) · P. F. Le Roux · C. G. Richards
Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa
e-mail: kabelomashiloane@gmail.com

among the PV modules could impact how well the system functions after they are electrically connected [2, 3]. Analyzing, detecting, and protecting faults is crucial in solar PV systems. Solar PV systems' PV arrays, power conditioning units, batteries [4], wiring, and utility interconnections are still susceptible to failures or defects despite having no moving parts. In PV arrays [5], since PV modules are energized by sunlight during the daytime, it is difficult to shut them down during fault conditions.

The analysis is significant in controlling the quality assurance of the PV modules to ensure that they provide dependable power generation by doing consistent PV module analysis with the I–V curve. It helps spot internal and external problems such as damaged solar cells, shaded regions, elevated temperature stresses, or malfunctioning bypass diodes by comparing the measured I–V curve to the predicted curve.

While short-circuit current and open-circuit voltage measurements can be measured using digital multimeters and clamp meters, I–V curve tracing enables monitoring while the module is in use. Knowing how a PV module performs under load allows for non-invasive diagnostics, which eliminates expensive, time-consuming, and invasive repairs that reduce the performance of the PV system as a whole.

A method for detecting, classifying, and locating faults is proposed in this study. A top-down approach will be used based on I–V curve analysis to detect and find faults that may occur. Simulated faults demonstrate the method's ability to classify and identify faults clearly in PV systems. Unlike other techniques, I–V curve analysis helps understand how faults affect PV systems while simultaneously being flexible [6, 7]. The simulations will be done using MATLAB/Simulink, whereas many other researchers use ETAP software for their modelling [8].

2 Types of Faults in the PV Array

Faults can also be categorized as permanent or transient based on the length of the PV system [9]. A transient issue lasts only a short while, such as dust, dirt, or snow buildup on the PV module's surface. However, a long-lasting flaw (such as age, slack, or disconnected electrical wire) persists in the system. According to [9], these shadows are produced by adjacent objects (such as trees, buildings, and other objects) or clouds moving directly overhead. The main way these faults have been characterized is by their common traits. This section lists common defect kinds [10] (Figs. 1, 2 and Table 1).

3 Photovoltaic I–V Curve Analysis

A PV module, string, or array can operate at any position along the I–V curve (current versus voltage) depending on the ambient conditions. A curve that begins at the short-circuit current and finishes at the open-circuit voltage is shown in Fig. 3.

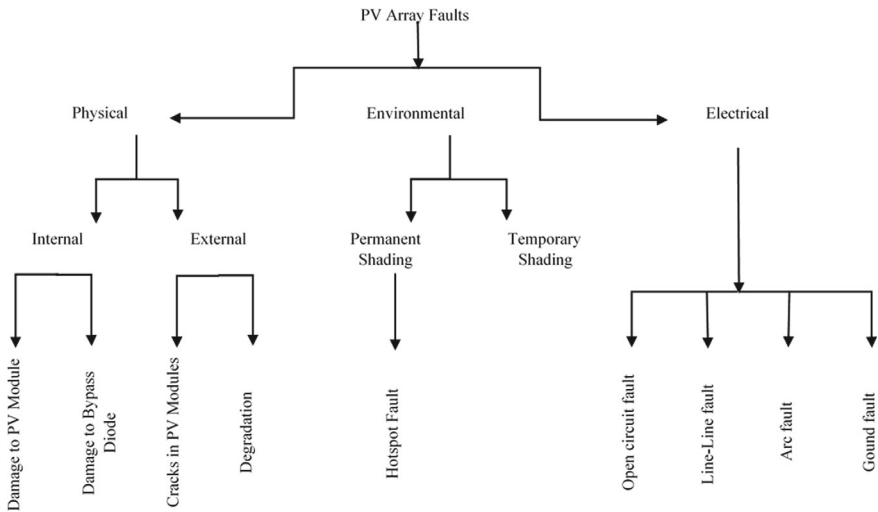


Fig. 1 Classification of faults

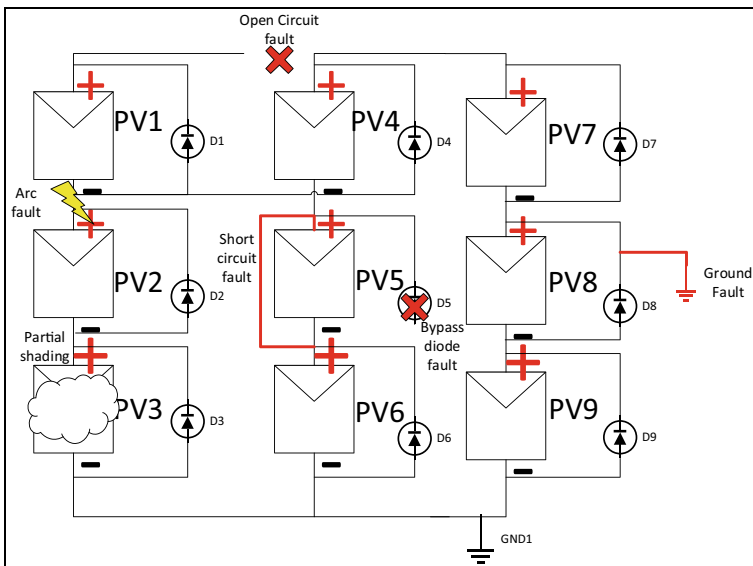


Fig. 2 Typical faults on PV system

The maximum power point at the I–V curve’s knee point is the operating point which produces the most output power [18]. As temperature and irradiance change, the inverter’s Maximum Power Point Tracking circuit (MPPT) is tasked with locating the maximum power point. Power against voltage (P–V) curves have zero at either end or reach their maximum at the I–V curve’s knee point. Any impediment that

Table 1 Reason for occurrence

Fault	Reason for occurrence
Hotspot fault	Occurs due to cell mismatch, excessive resistance, or cell deterioration. Partially shadowing [11]
Partial shadowing	Caused by obstructions to the sun irradiance in the form of moving clouds, trees, and other structures [12]
Line-Line fault	When a system’s two DC conductors short-circuits one another [13]
Ground fault	A PV system’s cables or metal components short-circuit accidentally [14]
Arc fault	The plasma discharges and the array may burn because of the extremely high temperature
PV module fault	This happens due to broken glass, disconnections, or incorrect connections [15, 16]
Diode faults	Overheating can be blamed for this bypass diode with a short-circuit or open-circuit [17]

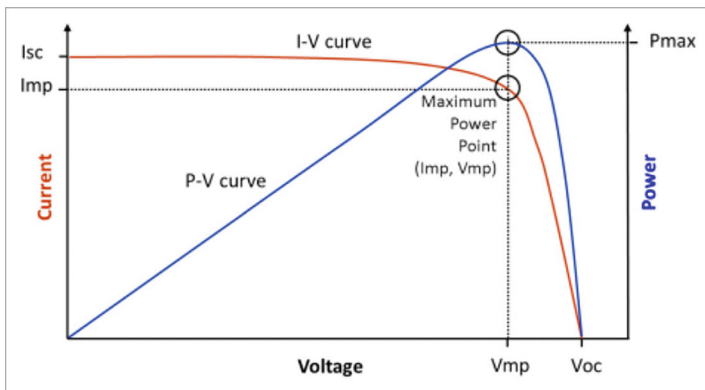


Fig. 3 I-V and P-V curve characteristic graphs

alters the I-V curve’s form, as depicted in Fig. 4, will lower peak power and lessen the array’s worth. Figure 5 illustrates how a mismatch impairment affects the output power.

The plots obtained for all conceivable points within a certain working range are the power voltage (P-V) and current voltage (I-V). There is a precise maximum power point (MPP) on each curve.

$$P_{mp} = I_{mp} * V_{mp} \tag{1}$$

From a characteristic curve’s maximum point, Imp and Vmp can be found. A PV cell’s open-circuit voltage, short-circuit current, fill factor, and efficiency are its additional characteristics.

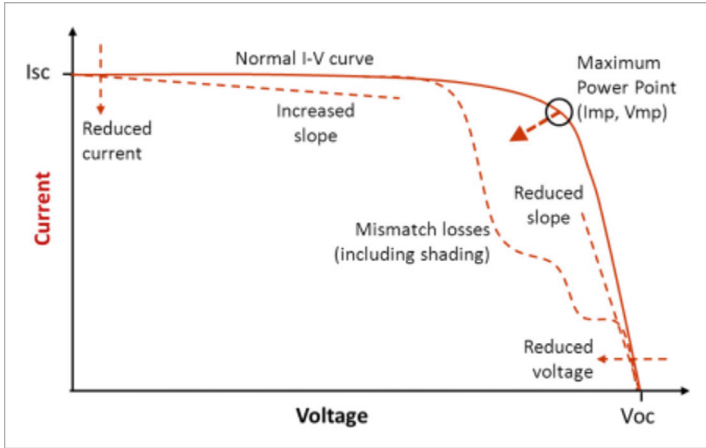


Fig. 4 The losses that can reduce the output of the PV array

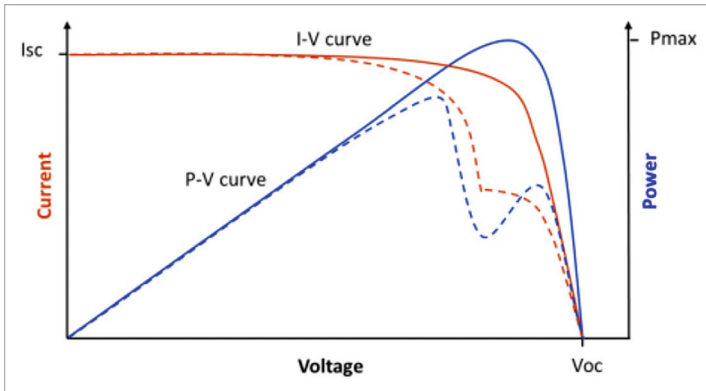


Fig. 5 The effects of some faults on the PV array

When the PV cell’s circuit is still open, and the current flowing through it equals zero, the “open-circuit voltage” is present.

$$V_{oc} = V(\text{at } I = 0) \tag{2}$$

The open-circuit voltage (V_{oc}) is the maximum voltage in a solar cell while the current is zero. The forward bias on the solar cell, symbolized by the open-circuit voltage, results from the PV cell’s preference for connection with the current created by Light.

$$V_{oc} = \frac{NKT}{q} \ln \left[\frac{I_L}{I_0} + 1 \right] \tag{3}$$

where N = ideality factor, I_L = light-generated current, T = temperature, I_0 = dark saturation current, K = Boltzmann constant, q = charge of the electron.

Temperature and open-circuit voltage (V_{OC}) are exactly related according to the equation above. As a result, V_{OC} increases directly as the temperature does. However, this did not occur because the saturation current increases quickly as the temperature rises. As a result, it is difficult to ascertain the effect of temperature on open-circuit voltage. As the saturation current varies with temperature, it gets smaller.

When zero voltage is discovered across the system, the maximum output current from a solar cell is derived as short-circuit current (I_{SC}).

$$I(\text{at } V = 0) = I_{sc} \quad (4)$$

The short-circuit current is formed and accumulated from light-generated carriers. In a perfect solar cell with no resistive loss mechanisms, there is no differentiation between the light-generated current (I_L) and the short-circuit current (I_{SC}). Short-circuit current is, therefore, the maximum current that can be recovered from a solar cell.

A solar cell's quality can be assessed by determining its Fill Factor (FF). The computation is carried out by dividing the greatest power by the theoretical power obtained by multiplying the short-circuit current by the open-circuit voltage.

$$FF = \frac{I_{mp} V_{mp}}{I_{sc} V_{oc}} = \frac{P_{max}}{I_{sc} V_{oc}} \quad (5)$$

The Fill Factor (FF) can calculate an organic solar cell's power conversion efficiency. Various variables may greatly influence the fill factor, and these variables interact intricately. Consequently, it is not easy to comprehend the idea of the FF. Based on the three main components of the PV cell equivalent circuit, the diode, shunt resistance, and series resistance, scientific progress in comprehending FF in organic solar cells should be evaluated.

Efficiency can be calculated by dividing the greatest output power by the input power.

$$\eta = \frac{P_m}{P_{in}} * 100 \quad (6)$$

A key factor in comparing a PV cell's performance to other solar cells is its efficiency. A solar cell's output and input energy can be compared to determine its efficiency. The operating temperature of the PV cell, the band spectrum, and the sunlight's intensity can all impact how efficient they are.

4 Simulation Models

Specifically, the Soltech 1STH-215-P polycrystalline module will be used in the simulation. The module’s specs are displayed in Table 2.

A series–parallel connection will be built for the simulation since it produces much electricity for big PV system plants.

The modules will be stringed together as a PV module to get the necessary voltage. After that, two additional strings will be joined in parallel to create the system’s required current level. The schematic diagram and simulation setup for the system model that will be simulated are depicted in the accompanying Fig. 6. By letting electricity flow around the damaged or shaded panel or string, the bypass diodes linked in parallel to each PV module help increase the system’s efficiency and dependability. Figure 6 illustrates the Simulink model with no fault; the PV system is a 3×3 system (i.e., consists of 3 arrays connected in parallel to the PV system with a string consisting of 3 PV modules connected in series). Table 3 provides the parameters of each string on the system, and Table 4 provides the parameters of the system’s total output.

4.1 Case Study One—Partial Shading

The simulation will compare the normal PV systems illustrated in Fig. 7 with Fig. 9 and Fig. 11 under partial shading faults. Figure 8 shows the schematic diagram of a system with two PV modules affected by partial shading, and Fig. 10 shows the schematic of four PV modules under shading conditions. This will show the output effect of partial shading on the PV system.

Table 2 Parameters of Soltech 1STH-215-P polycrystalline module

Parameters	Values
Maximum power (W)	213.15 W
Open-circuit voltage V_{oc} (V)	36.3 V
Voltage at maximum power point V_{mp} (V)	29 V
Temperature coefficient of V_{oc} (%/deg.C)	– 0.36099%/deg.C
Short-circuit current I_{sc} (A)	7.84 A
Current at maximum power point (A)	7.35 A
Temperature coefficient of I_{sc} (%/deg.C)	0.102%/deg.C
Cells per module (N_{cell})	60
Shunt resistance R_{sh} (ohm)	313.3991 Ω
Diode ideality factor	0.98117
Series resistance R_s (ohm)	0.39383 Ω

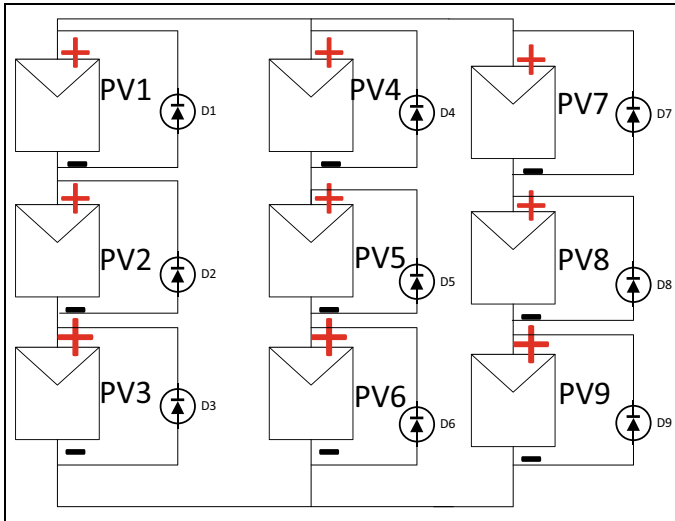


Fig. 6 Schematic diagram of modelled simulation with no fault

Table 3 Parameters of one string on the system (calculated)

Parameters	Values
Maximum power (W)	639.45 W
Open-circuit voltage V_{oc} (V)	108.9 V
Voltage at maximum power point V_{mp} (V)	87 V
Short-circuit current I_{sc} (A)	7.84 A
Current at maximum power point (A)	7.35 A
Power (W)	645.00 W
Efficiency (η)	99.14%
Fill factor FF	0.799

Table 4 Parameters of system output (calculated)

Parameters	Values
Maximum power (W)	1918.35 W
Open-circuit voltage V_{oc} (V)	108.9 V
Voltage at maximum power point V_{mp} (V)	87 V
Short-circuit current I_{sc} (A)	23.52 A
Current at maximum power point (A)	22.05 A
Power (W)	1935 W
Efficiency (η)	99.14%
Fill factor FF	0.799

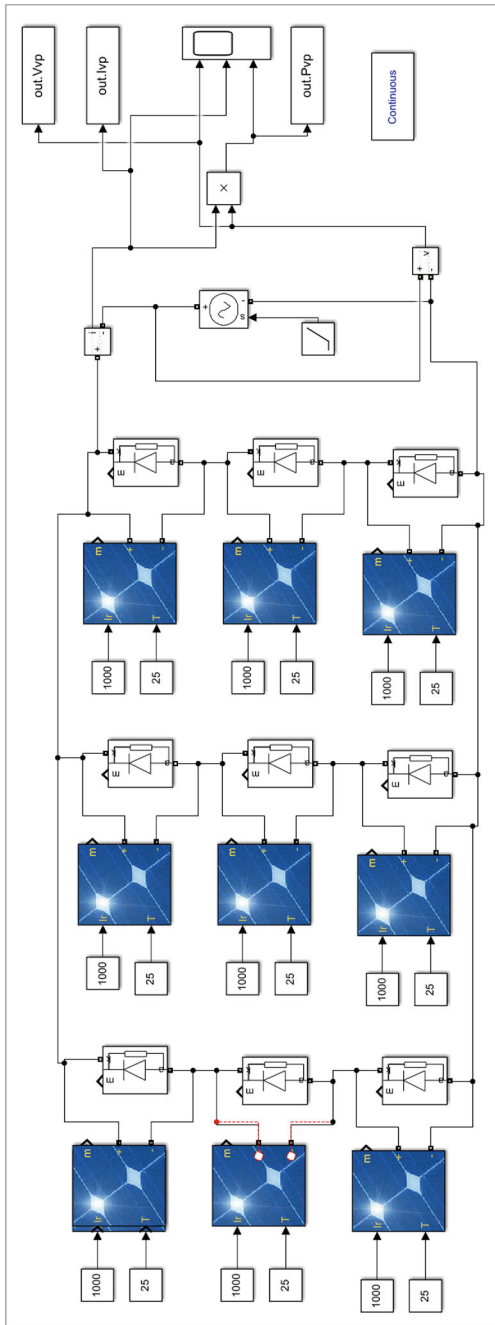


Fig. 7 Schematic diagram of modelled simulation with no fault

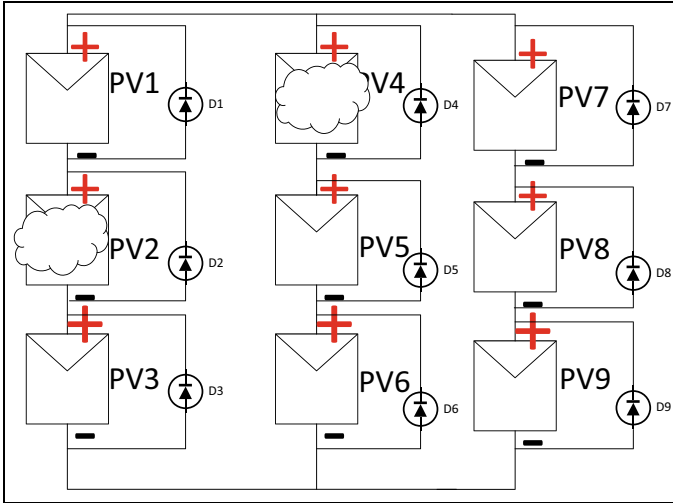


Fig. 8 Schematic diagram of two-shaded modules

Figure 12, the I–V curve, illustrates no fault as per the blue line plotted on the graph, indicating that all modules perform at the highest efficiency without shading. The P–V curve as seen in Fig. 13, the two-shaded modules, as denoted by the red line's curves on Fig. 13 of PV2 and PV4 as shown in Fig. 9, have low irradiance input. With the aid of bypass diodes, the modules got bypassed due to the low irradiance condition. However, as seen in Fig. 13, this created two maximum power points on the system and two knee points on the I–V curve in Fig. 12; thus, the power output has decreased. The four-shaded modules, as represented with the yellow line as shown in Fig. 13, have low irradiance input that shows significant losses. The plot shows that PV2 got bypassed due to its low irradiance, and in the second array, PV4 and PV5 got bypassed due to low irradiance, and PV6 got bypassed on the third array of Fig. 11 simulation model.

As seen in Fig. 13, the Maximum Power Point (MPP) is illustrated, indicating the PV modules were bypassed. This is due to partial parts of the surface of the PV modules being shaded, thus decreasing the power produced by the shaded modules. Should this be ignored, a hotspot will be created, causing a fault in the solar system.

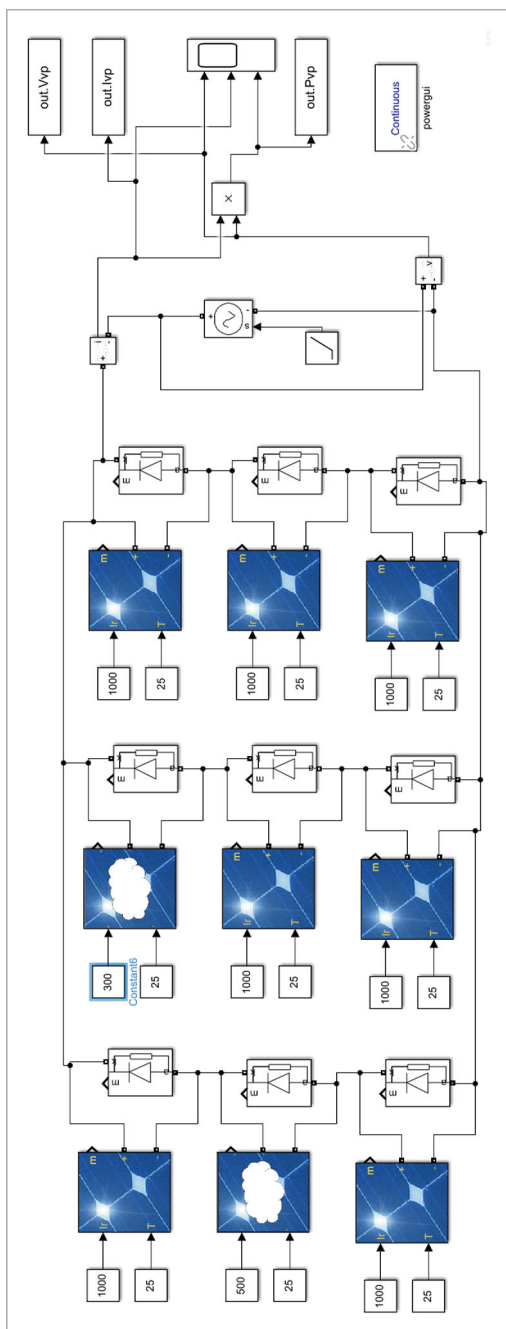


Fig. 9 Schematic diagram of modelled simulation with two-shaded modules

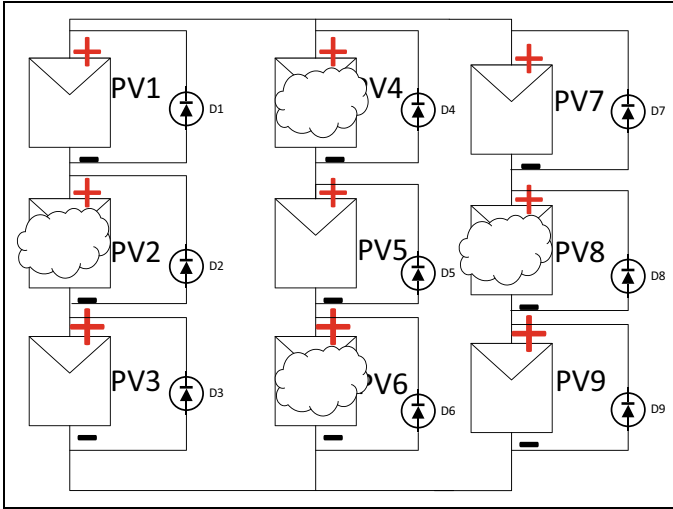


Fig. 10 Schematic diagram of four-shaded modules

4.2 Case Study Two—Module Mismatch

The simulation model in this case study will have figures where some of the modules will have different ratings, also known as a mismatch of PV modules, which will be simulated and compared. Table 5 will show the outputs for PV module 2 mismatch A, and then Table 5 will also show the specifications for PV modules 2 and 9 of mismatch B. They will be compared to a normal PV system with PV modules with the exact specifications.

The I–V characteristic curve in Fig. 14 with the electrical parameters of Table 2 in all modules of Fig. 7 simulation model under Standard Testing Conditions (STC), it can be observed that the maximum power produced, without losses, will be expected as shown in Figs. 14 and 15.

As indicated above, the mismatch A scenario refers to the PV2 module with different rating conditions, as shown in Table 5. For the remaining PV modules of Fig. 7, simulation model mismatch loss will be the result, as shown in Figs. 14 and 15, where the I–V and P–V curves are denoted in orange line curves.

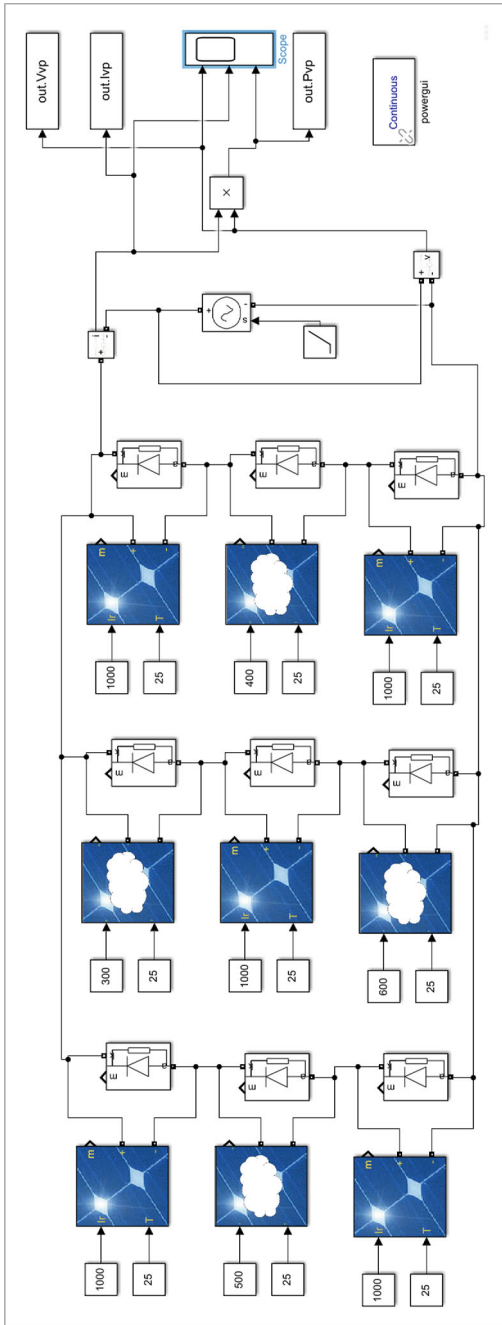


Fig. 11 Schematic diagram of modelled simulation with four-shaded modules

Fig. 12 I–V curve of partial shading results with no fault reference curve

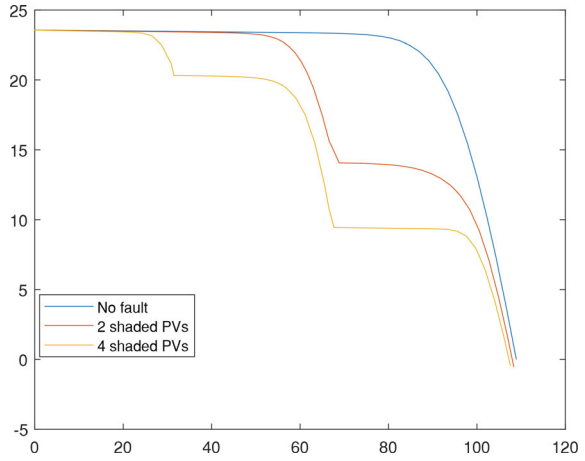


Fig. 13 P–V curve of partial shading results with no fault reference curve

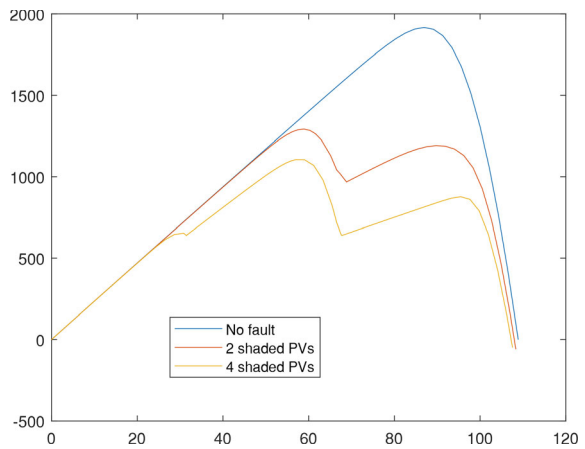


Table 5 Modules with lower output than their ratings

Parameters	Values (PV2)	Values (PV9)
Maximum power (W)	189.97 W	179.21 W
Open-circuit voltage V_{oc} (V)	35.20 V	34.2 V
Voltage at maximum power point V_{mp} (V)	27.10 V	26 V
Short-circuit current I_{sc} (A)	7.38 A	7.1 A
Current at maximum power point (A)	7.01 A	6.89 A
Cells per module (N_{cell})	60	60
Shunt resistance R_{sh} (ohm)	1085.67 Ω	5609.60 Ω
Series resistance R_s (ohm)	0.57 Ω	0.61 Ω

Fig. 14 I-V curve of mismatch results with no fault reference curve

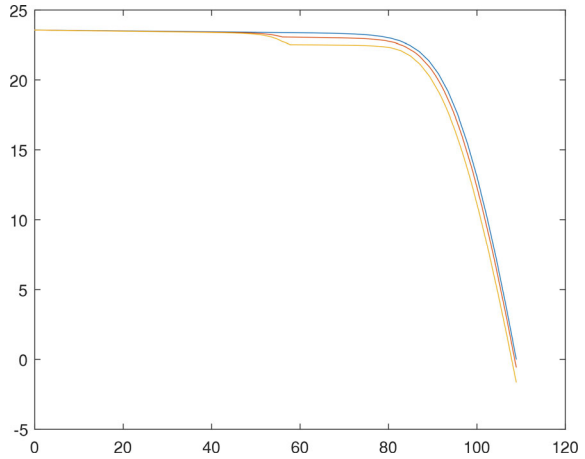
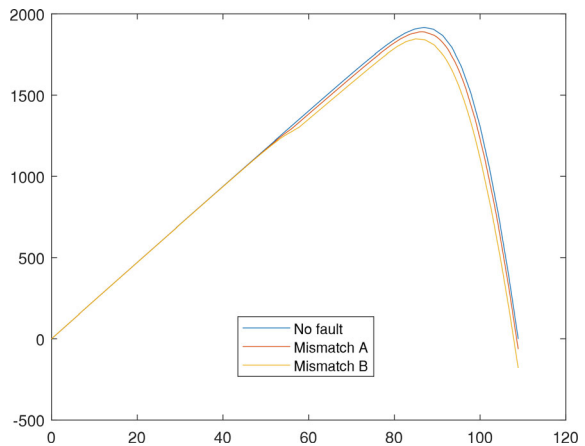


Fig. 15 P-V curve of mismatch results with no fault reference curve



Mismatch B, denoted with a yellow line in Figs. 14 and 15, shows an increase in mismatch losses with two mismatch modules on PV2 and PV9 with the rating characteristics of Table 2 than the rest of PV modules. The results show that when designing a PV system, the modules' ratings should be considered when designed. The short circuit (I_{SC}) and open circuit (V_{OC}) tests should be conducted on the PV modules to ensure they match the ratings of the nameplate under STC conditions.

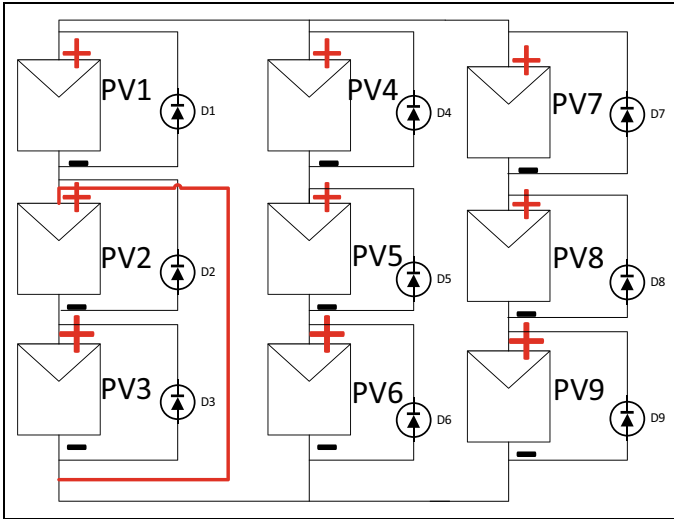


Fig. 16 Schematic diagram of a short-circuit in the system

4.3 Case Study Three—Short-Circuit Fault

The schematic diagram in Fig. 12 shows a short-circuit fault from the positive terminal of the PV2 module to the negative terminal of the PV3 module, and Fig. 13 shows the Simulink model of the diagram. Figure 15 is another possible scenario of a short-circuit from the PV2 module to between the PV5 and PV6 modules. The two potential short-circuit faults will be compared to a normal working PV system (Figs. 16, 17, 18 and 19).

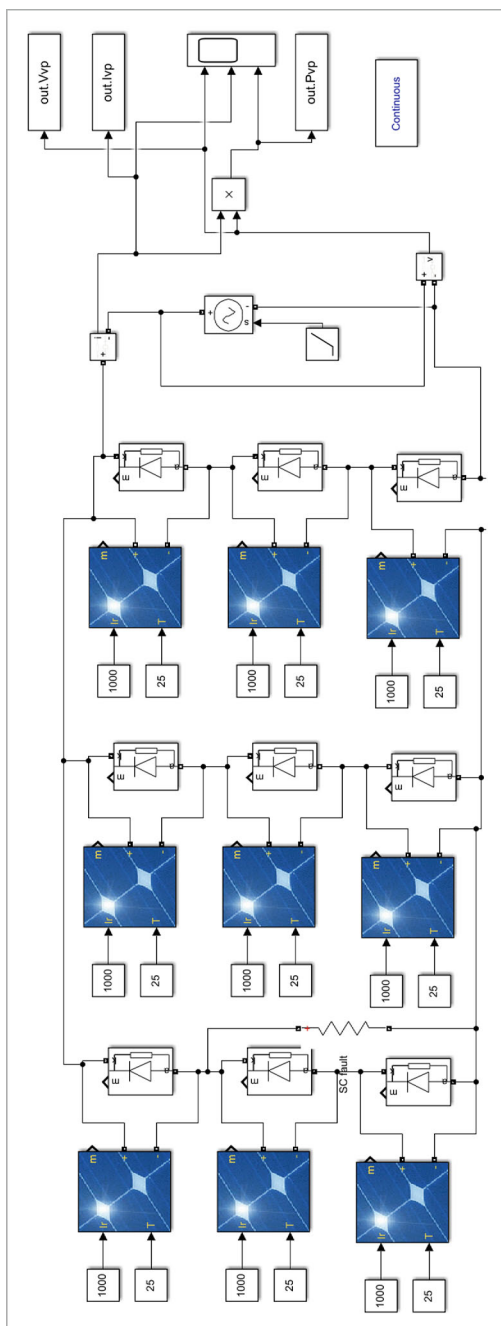


Fig. 17 Schematic diagram of modelled simulation with a short-circuit

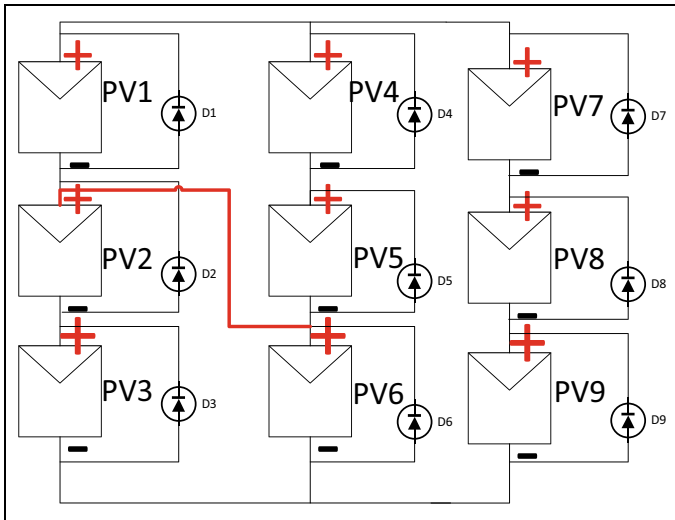


Fig. 18 Schematic diagram of a short-circuit in the system

The plots show that the PV2 and PV3 being on the same string, the output of the system significantly affects the output, I_{SC} is not affected where the V_{OC} is highly impacted, and due to this, the maximum output of the system will drop. With the short-circuit between two strings, as shown in Fig. 4.10. I_{sc} is also unaffected, but V_{oc} has reduced and untimely lowers the PV system's output. This indicates that a short-circuit does not affect I_{sc} but reduces V_{oc} and drops the power output of the Solar system. The occurrence of short circuits is primarily due to line-to-line faults (Figs. 20 and 21).

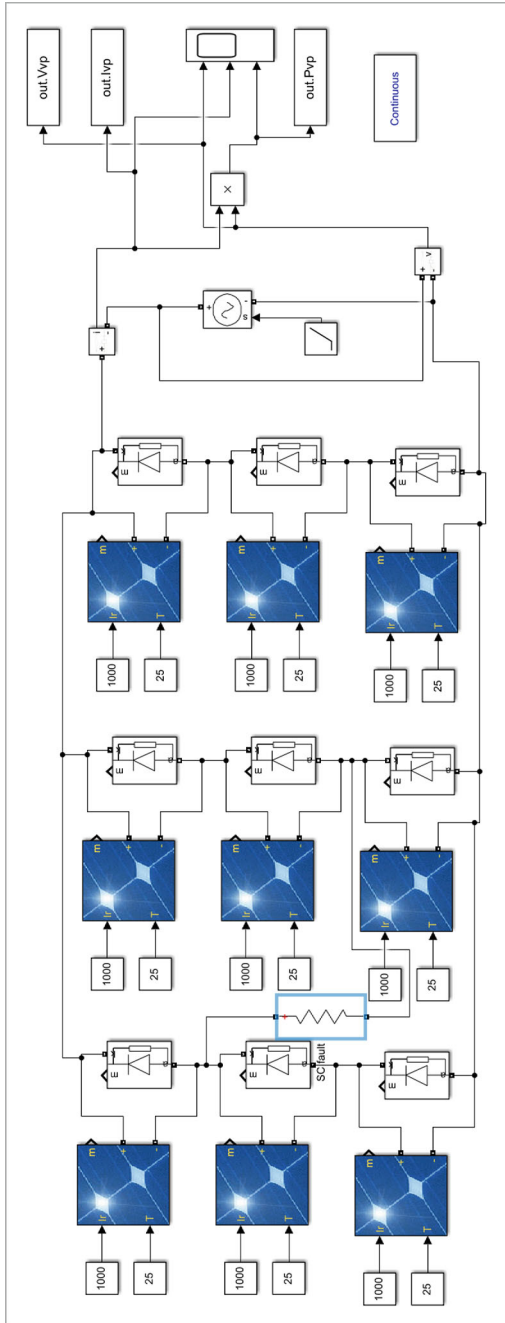


Fig. 19 Schematic diagram of modelled simulation with a short-circuit

Fig. 20 I–V curve of short circuit faults results with no fault reference curve

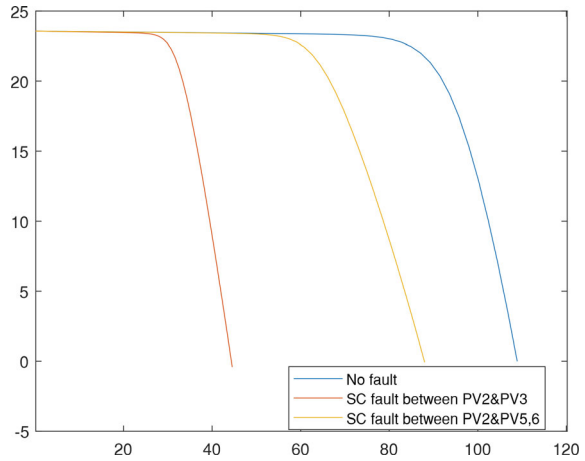
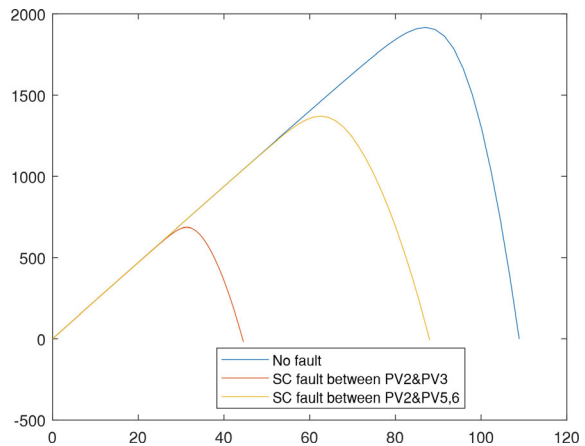


Fig. 21 P–V curve of short circuit faults results with no fault reference curve



5 Conclusion

Fault monitoring is essential to ensure that PV systems operate safely and dependably. In Solar PV systems monitoring, automatic fault identification and categorization remain a significant problem. This research proposed a diagnosis method based on examining features curve deviation. A model with no fault was simulated and compared to different fault model conditions of the PV system to identify the faulty modules. The I–V and P–V curves’ shapes offer essential details regarding possible causes of performance issues in the modules and provide information to determine the most likely reason for the PV modules’ under-performance. It will be essential for the analysis to be performed during installation, commissioning, or periodically on system inspection to check the system’s health and performance at specified parameters. I–V testing would not be sufficient to inspect a PV plant with numerous

PV strings on each module to identify faults, as it will be time-consuming, and the labor costs would be high. For future research, smart I–V and P–V curve analysis for automatic inspection of PV systems should be considered for a less time-consuming and efficient analysis.

References

1. Bakdi A et al (2021) Real-time fault detection in PV systems under MPPT using PMU and high-frequency multi-sensor data through online PCA-KDE-based multivariate KL divergence. *Int J Electr Power Energy Syst* 125:106457
2. Dhibi K et al (2020) Reduced kernel random forest technique for fault detection and classification in grid-tied PV systems. *IEEE J Photovolt* 10(6):1864–1871
3. Abubakar A, Almeida CFM, Gemignani M (2021) Review of artificial intelligence-based failure detection and diagnosis methods for solar photovoltaic systems. *Machines* 9(12):328
4. Makola CS, Le Roux PF, Jordaan JA (2023) Comparative analysis of lithium-ion and lead-acid as electrical energy storage systems in a grid-tied microgrid application. *Appl Sci* 13(5):3137
5. Cubukcu M, Akanalci A (2020) Real-time inspection and determination methods of faults on photovoltaic power systems by thermal imaging in Turkey. *Renew Energy* 147:1231–1238
6. Venkatakrishnan G et al (2023) Detection, location, and diagnosis of different faults in large solar PV system—a review. *Int J Low-Carbon Technol*, p ctad018
7. Fadhel S et al (2019) PV shading fault detection and classification based on IV curve using principal component analysis: application to isolated PV system. *Sol Energy* 179:1–10
8. Makola C, Le Roux P, Jordaan J (2021) Design, analysis, and operation of photovoltaic power in a microgrid with an EESS. In: 2021 IEEE PES/IAS Power Africa. IEEE
9. Madeti SR, Singh S (2017) A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Sol Energy* 158:161–185
10. Basnet B, Chun H, Bang J (2020) An intelligent fault detection model for fault detection in photovoltaic systems. *J Sens* 2020:1–11
11. Tina GM, Cosentino F, Ventura C (2016) Monitoring and diagnostics of photovoltaic power plants. In: *Renewable energy in the service of mankind, selected topics from the world renewable energy congress WREC 2014*. Springer, vol II
12. Stellbogen D (1993) Use of PV circuit simulation for fault detection in PV array fields. In: *Conference record of the twenty third IEEE photovoltaic specialists conference-1993 (Cat. No. 93CH3283-9)*. IEEE
13. Gokmen N et al (2012) Simple diagnostic approach for determining of faulted PV modules in string based PV arrays. *Sol Energy* 86(11):3364–3377
14. Hachana O, Tina GM, Hemsas KE (2016) PV array fault diagnostic technique for BIPV systems. *Energy Build* 126:263–274
15. Chao K-H, Ho S-H, Wang M-H (2008) Modeling and fault diagnosis of a photovoltaic system. *Electr Power Syst Res* 78(1):97–105
16. Kaplanis S, Kaplani E (2011) Energy performance and degradation over 20 years performance of BP c-Si PV modules. *Simul Model Pract Theory* 19(4):1201–1211
17. Chine W et al (2015) Fault diagnosis in photovoltaic arrays. In: 2015 international conference on clean electrical power (ICCEP). IEEE
18. Nelson JA (2010) Effects of cloud-induced photovoltaic power transients on power system protection

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Comparing Adopter, Tester, and Non-adopter of Collaborative Augmented Reality for Industrial Services



Maike Müller , Stefan Ohlig , Dirk Stegelmeyer , and Rakesh Mishra 

Abstract Collaborative augmented reality (CAR) is a remote collaboration technology that utilizes augmented reality (AR) to create a shared environment for distributed collaborators conducting physical tasks. CAR became commercially available a few years ago, and its industrial adoption was accelerated by the contact and travel restrictions imposed during the Covid-19 pandemic to provide industrial services. However, it seems that despite implementation, the technology is not fully embraced and used regularly. Therefore, the objective of this paper is to explore differences in the assessment of perceived benefits, opportunities, challenges, and barriers in implementing CAR among different adoption status groups (i.e., Adopters, Testers, and Non-adopters). To achieve this objective, we conducted a survey in the German capital equipment industry. With a sample size of 130 companies, our study is the first attempt to quantitatively explore CAR adoption in the capital equipment industry and it provides valuable insights into the reasons for potential hesitations in adopting CAR.

Keywords Augmented reality · Collaboration technology · Technology adoption · Technology implementation · Remote service · Capital equipment

1 Introduction

Many capital equipment manufacturers provide services such as equipment installation, maintenance, fault diagnosis, repair, and overhaul. A key customer requirement in industrial service delivery is short response times to equipment failures,

M. Müller (✉) · S. Ohlig · R. Mishra
School of Computing and Engineering, University of Huddersfield, Huddersfield, UK
e-mail: maike.mueller@hud.ac.uk

M. Müller · S. Ohlig · D. Stegelmeyer
Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences,
Frankfurt, Germany
e-mail: stegelmeyer@fb2.fra-uas.de

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_10

particularly those causing downtime [1, 2]. However, manufacturers are faced with a globally dispersed installed base, resulting in significant travel times to deliver these services. Moreover, product complexity often demands scarce specialized knowledge [2, 3]. Apart from that, delivering industrial services often involves complex value creation networks, with participants including at least the capital equipment manufacturers and their customers, but often also component suppliers, or other service providers [2]. Thus, efficient service delivery requires effective information sharing and remote collaborative problem-solving in inter-organizational settings. However, remote collaboration often relies on traditional communication by phone, messaging, and email [4, 5], which can lead to long resolution times and misunderstandings [6, 7].

To address these challenges, capital equipment companies have started to extend their remote technology toolbox with collaborative augmented reality (CAR), which has become commercially available a few years ago [8]. CAR allows multiple users such as remote experts and on-site technicians to share the same augmented environment, when they are not co-located [9]. This facilitates knowledge transfer and helps in executing physical service tasks, for example when technicians lack knowledge during on-site service interventions [10]. However, even though the contact and travel restrictions imposed during Covid-19 accelerated the adoption of CAR [11, 12], anecdotal evidence provided by our industry partners indicates that, despite its implementation, CAR is not fully embraced and used regularly. Therefore, wide-scale industrial diffusion seems yet to be realized [10, 13].

Therefore, the objective of this paper is to explore differences in the assessment of perceived benefits and opportunities, as well as challenges and barriers of implementing CAR among different adoption status groups (i.e., Adopters, Testers, and Non-adopters). To achieve this objective, we conducted a survey in the German capital equipment industry. Thereby, this paper contributes insights into why companies in the capital equipment industry might hesitate to adopt CAR.

2 Theoretical Background

2.1 Collaborative Augmented Reality

Augmented reality (AR) is defined as ‘any case in which an otherwise real environment is “augmented” by means of virtual (computer graphic) objects’ [14]. CAR is a specific type of AR, also known as ‘AR remote maintenance’, ‘mobile collaborative AR’, or ‘tele-maintenance’, among others [9]. CAR consists of two main features: a shared view and awareness cues when operated on handheld devices. When wearable AR devices such as head-mounted displays (HMDs) are utilized, hands-free communication is added to the scope of function.

A shared view allows remote experts and on-site technicians to see the same workspace when collaborating remotely. The primary goal of shared views is to enhance situational awareness and establish effective communication. This can result

in performance enhancements of 30–40% and allows remote experts to take corrective actions to reduce misunderstandings and errors [15]. Various approaches enable shared views, ranging from live video links [5, 16] to advanced systems combining AR and virtual reality modes [17, 18]. In industry, video links are commonly used.

The second main feature of CAR is awareness cues, which also aim to enhance situational awareness and effective communication by overlaying virtual content onto the view of the on-site technician. It is this feature that distinguishes CAR from a mobile teleconferencing system. Virtual content can encompass pre-made 2D elements such as cursors, arrows, and circles [19, 20], as well as more complex 3D objects like CAD representations of machine components [18, 21], or free-hand drawing tools [16, 19]. Engineering research is also developing advanced communication cues, such as projecting the remote expert's hand gestures and head/gaze direction into the view of the on-site technician [17, 18]. However, advanced awareness cues have not yet found practical application in the industry, and even complex cues are rarely used.

Hands-free communication involves the utilization of an HMD to enable on-site technicians to carry out physical tasks while receiving remote instructions. Given that field service technicians are highly mobile workers, researchers often prioritize the integration of HMDs in the development of CAR technology. However, in practice CAR is currently predominantly delivered through hand-held devices, since on-site technicians prefer smartphones and tables over HMDs [10].

2.2 *Related Work*

Most research into CAR has been centered around the development of prototypes within controlled laboratory environments [9, 10]. As expected, these studies offer valuable technical perspectives on the implementation of CAR. Yet, companies are also struggling with organizational aspects [22, 23]. Currently, there are only a handful of studies available that focus on CAR adoption [6, 24, 25]. Other studies have investigated adoption across a broad spectrum of industrial AR applications and use cases, without specifically concentrating on CAR aspects [13, 22, 26]. In contrast, our research is exclusively centered on AR for remote collaboration and inter-organizational service delivery. The groundwork for this paper is published in [8]. The study developed an adoption model based on a qualitative research design. To the best of the authors' knowledge, at present, no quantitative survey on CAR adoption is available.

3 Methodology

This paper aims to explore differences in the assessment of perceived benefits and opportunities, as well as challenges and barriers of implementing CAR among different adoption status groups. To this end, a questionnaire survey was conducted within the German capital equipment industry.

3.1 *Sampling and Data Collection*

Based on our previous research [8], which provided a list of benefits, opportunities, challenges, and barriers of adopting CAR in industrial services within the capital equipment industry, a questionnaire was developed. This questionnaire consists of 53 statements concerning CAR adoption (cf. Appendix 1), along with six additional questions to characterize the survey participants. Statements concerning CAR adoption were measured using five-point Likert-type items, with values ranging from ‘strongly disagree’ (1) to ‘strongly agree’ (5). To measure the characteristics of the survey participants’ categorical items were used.

The targeted population of this survey are companies operating in the capital equipment industry offering technical services to their customers. The most suitable survey respondents are service managers at the first-line or middle management levels, as they possess a comprehensive understanding of the strategic implications of implementing CAR, along with the technical, user, and customer-related aspects. Due to the hidden nature of the population, conventional random sampling was not feasible. Therefore, we primarily collected data during events organized by VDMA, the German Mechanical Engineering Industry Association, which provided access to our target respondents. This strategy reached 235 potential respondents at VDMA events. Further, we disseminated the questionnaire to an additional 95 individuals identified from companies advertising CAR support, reached via email or LinkedIn message. In total, we distributed 340 questionnaires between September 2022 and June 2023. The response rate was 39%, with 132 returned questionnaires, including two responses of consulting companies, which were excluded. Thus, the sample size of the survey is $N = 130$.

3.2 *Missing Data Analysis and Data Imputation*

Missing data poses a pervasive challenge since statistical analysis methods assume complete data [27]. Our analysis revealed 125 missing values, constituting 1.6% of the dataset. Missing values occurred in 34 cases, representing 25.8% of the total cases, and occurred across 50 different items, amounting to 83.3% of the total items. The percentage of missing values per item ranged from 0.8 to 6.1%.

Choosing a method to address missing data often involves selecting the best among undesirable options [28]. Yet, applied researchers often disregard recommended imputation methods endorsed by statisticians [28], even though recommended imputation methods offer vital advantages over other missing value treatments, such as sample size preservation and sustained statistical power [29].

We employed Multiple Imputation through the Multivariate Imputation by Chained Equations (MICE) algorithm [30] using the Predictive Mean Matching (PMM) method [31, 32] as detailed in [33]. All questionnaire items were included in the MICE-PMM model. Items with no missing values served as predictors, while those with over 3% missing values were solely imputed. We selected 50 imputations with 50 iterations and 5 donor cases, as recommended in [33]. The MICE-PMM algorithm was executed using IBM SPSS Statistics 28.0. We evaluated convergence by creating convergence plots for all imputed items, confirming the stability of the algorithm, with no systematic trends observed, and, as a result, no convergence issues present.

3.3 *Sample Profile*

Table 1 displays the proportionate breakdown of the sample. The results were pooled using the arithmetic average of the 50 imputed datasets. The vast majority (94.5%) of survey participants represent capital equipment companies, which encompass manufacturers of series machines, customized components or machines, component suppliers as well as plant engineering firms. Only 3.9% are software providers and 1.5% are trading or service companies. More than 80% of the companies within the sample are currently engaged in CAR, with 30.3% testing CAR, 20.1% in the process of implementing it, and 31.7% already operational users having the implementation completed. A small fraction of 4.7% has opted not to proceed with the adoption, and 13.1% have never considered adopting CAR. Consequently, the sample possesses valuable adoption experience to provide insights into CAR implementation. This also holds true for the respondents' management level and organizational function. A clear majority (76.1%) work in after-sales service units, and over two-thirds hold positions in either first-line (16.2%) or middle management (55.3%).

4 Results

To test for statistically significant differences in the distributions of items across the adoption status groups, we analyzed the data using the Kruskal–Wallis H test [34]. The Kruskal–Wallis H test is a nonparametric test used when the group differences to be tested come from different observations and assumptions of parametric testing (e.g., normality due to ordinal measures) are violated [35]. The test results were

Table 1 Sample profile ($N = 130$) as percentages (pooled data)

	Adopter ($N = 67.34$)	Tester ($N = 39.44$)	Non-adopter ($N = 23.18$)
<i>Adoption status</i>			
Never considered implementation	–	–	73.71
Decided against implementation	–	–	25.86
Tester (implementation decision outstanding)	–	100.00	–
Implementer (currently implementing)	38.78	–	–
Operational user (implementation completed)	61.22	–	–
<i>Firm size</i>			
Below 50	2.97	–	4.31
51–250	20.80	20.81	25.86
251–1000	30.16	37.31	48.28
1001–10,000	40.12	34.01	12.93
Above 10,000	5.94	7.87	8.62
<i>Business type</i>			
Capital equipment manufacturer	95.54	92.39	95.69
Software provider	2.97	7.61	–
Trading/service company	1.49	–	4.31
<i>Servitization level</i>			
Basic services	28.68	32.49	39.66
Intermediate services	66.86	65.23	60.34
Advanced services	4.46	2.54	–
<i>Participant's organizational function</i>			
After-sales service	83.66	72.08	61.21
IT/Digitalization	7.43	2.54	4.31
R&D	2.97	5.08	17.24
Manufacturing	1.49	7.61	8.62
Sales and marketing	2.97	7.61	4.31
Other	1.49	5.08	4.31

(continued)

Table 1 (continued)

	Adopter (<i>N</i> = 67.34)	Tester (<i>N</i> = 39.44)	Non-adopter (<i>N</i> = 23.18)
<i>Participant's management level</i>			
Operational employee	19.32	10.15	21.55
First-line management	11.89	20.30	21.55
Middle management	58.40	56.85	43.97
Top management	10.40	12.69	12.93

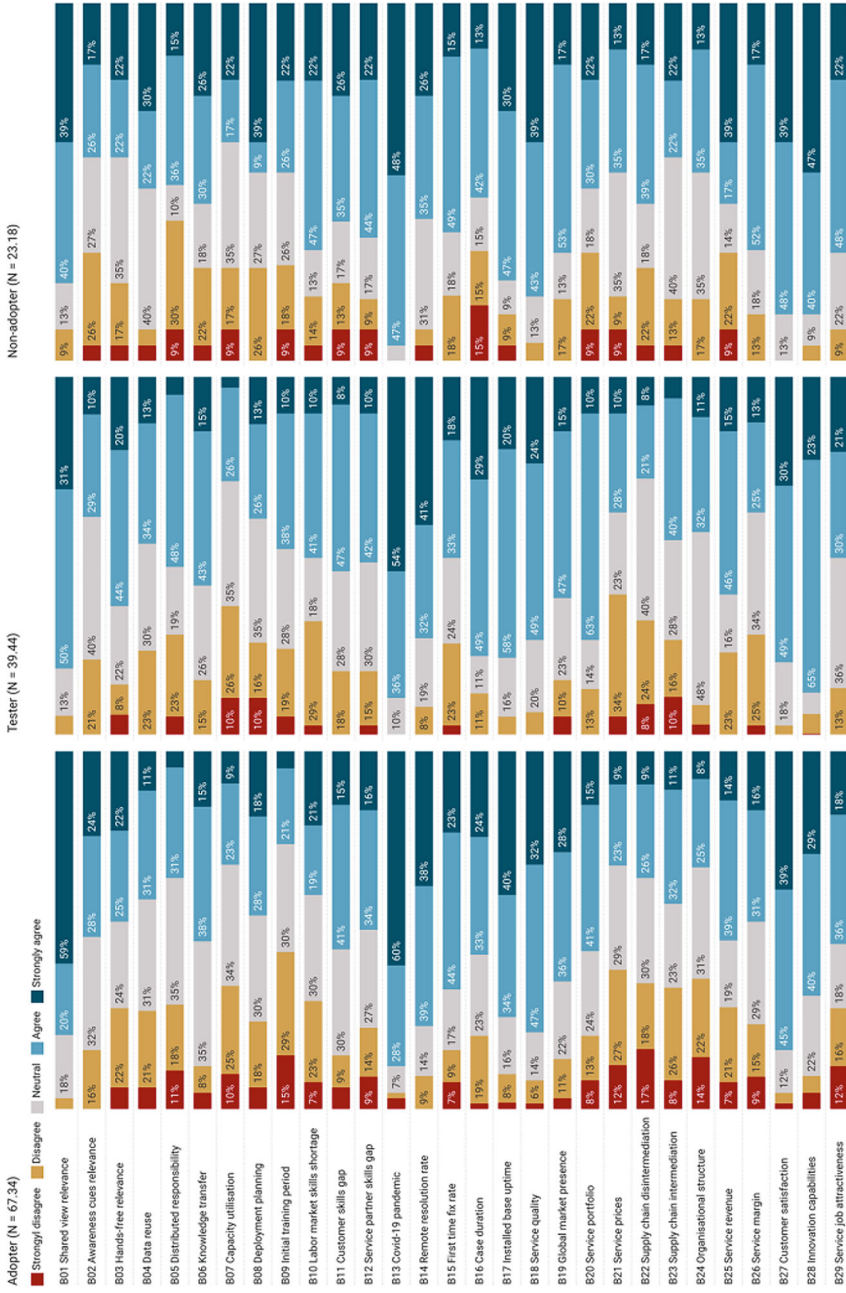
pooled, calculating the arithmetic average of mean ranks and the median of *H* statistics and *p* values. Given the exploratory nature of this survey, the significance level was set at 0.1. To understand the specific nature of the differences in the adoption status groups, we conducted a post-hoc analysis with a Bonferroni correction for multiple comparisons.

4.1 Benefits and Opportunities

Figure 1 presents stacked bar charts illustrating survey participants' frequency of agreement and disagreement with benefits and opportunities. The Kruskal–Wallis *H* test showed that there is a statistically significant difference in 4 out of 29 items representing benefits and opportunities associated with implementing CAR. Descriptive statistics for significant differences are available in Appendix 2.

The Kruskal–Wallis *H* test showed that there is a significant difference in the assessment of *B01 shared view relevance* between the adoption status groups, $H(2) = 4.623, p = 0.099$. The mean rank score was 60.15 for Non-adopters, 57.81 for Testers, and 71.81 for Adopters. However, pairwise comparisons with adjusted *p*-values showed no statistically significant differences when comparing Adopters to Testers ($p = 0.495, r = -0.194$), Adopters to Non-adopters ($p = 0.135, r = -0.146$), or Non-adopters to Testers ($p = 1.000, r = 0.034$). The high agreement rates (around 80%) regarding the importance of a shared view for collaboration between remote experts and on-site technicians were consistent across adoption status groups.

Statistically significant differences between the adoption status groups were also found in *B09 initial training period*, $H(2) = 8.384, p = 0.015$. The mean rank score was 75.63 for Non-adopters, 74.91 for Testers, and 56.47 for Adopters. Pairwise comparisons revealed significant differences between Adopters and Testers ($p = 0.037, r = 0.243$), as well as between Adopters and Non-adopters ($p = 0.082, r = 0.232$). However, no significant differences were found between Non-adopters and Testers ($p = 1.000, r = 0.008$). Testers and Non-adopters appear to overestimate the value of CAR in reducing the initial training periods for field service technician



Created with Datawrapper

Fig. 1 Frequency stacked bar charts (pooled results) for benefits and opportunities

recruits. In contrast, Adopters seem to have had different experiences. Only 26% of Adopters agree with the statement provided in the questionnaire, while 48% of Testers and 47% of Non-adopters agree.

In terms of the *B24 organizational structure*, we found a statistically significant difference between adoption status groups, $H(2) = 7.417, p = 0.024$, with a mean rank score of 73.90 for Non-adopters, 74.51 for Testers, and 57.30 for Adopters. However, there were no significant differences between Adopters and Non-adopters ($p = 0.169, r = 0.201$), as well as between Non-adopters and Testers ($p = 1.000, r = -0.013$). On the other hand, Adopters and Testers ($p = 0.049, r = 0.233$) significantly differ in their assessment of the potential of CAR in terms of redesigning the organization through centralization or decentralization of service knowledge. While Testers are optimistic that CAR could help to redesign the organization, Adopters more often disagree with this notion.

Statistically significant differences between the adoption status groups were also found in *B28 Innovation capabilities*, $H(2) = 4.626, p = 0.099$, with a mean rank score of 78.42 for Non-adopters, 66.39 for Testers, and 60.49 for Adopters. Pairwise comparisons revealed significant differences between Adopters and Non-adopters ($p = 0.095, r = 0.226$). However, Adopters and Testers ($p = 1.000, r = 0.080$), as well as Testers and Non-adopters ($p = 0.537, r = 0.170$), were not significantly different in their assessment of the perceived potential of CAR in demonstrating innovation capabilities toward customers. Even though the majority of Adopters (69%) agree on the potential of CAR to demonstrate innovation capabilities toward customers, we also observe more disagreement (9%) and neutral (22%) responses compared to Testers (88% agree, 6% neutral, and 6% disagree) and Non-adopters (87% agree, 9% neutral, and 4% disagree). With p -values of 0.137 and greater, the Kruskal–Wallis H test did not find statistically significant differences in the other benefits and opportunities analyzed.

4.2 Challenges and Barriers

Figure 2 presents stacked bar charts illustrating survey participants' frequency of agreement and disagreement with challenges and barriers. Just as with benefits and opportunities, the Kruskal–Wallis H test was used to test for statistically significant differences in the distributions of the adoption status groups. The same pooling procedure and significance level were used as described in the previous section. The Kruskal–Wallis H test showed that there is a statistically significant difference in 4 out of 24 items representing challenges and barriers associated with implementing CAR. Descriptive statistics for significant differences are available in Appendix 2.

The Kruskal–Wallis H test showed that there is a statistically significant difference in the assessment of *C07 data transmission* between the adoption status groups, $H(2) = 5.852, p = 0.053$. The mean rank score was 71.03 for Non-adopters, 74.36 for Testers, and 58.36 for Adopters. Pairwise comparisons with adjusted p -values showed no statistically significant differences when comparing Adopters to Non-adopters

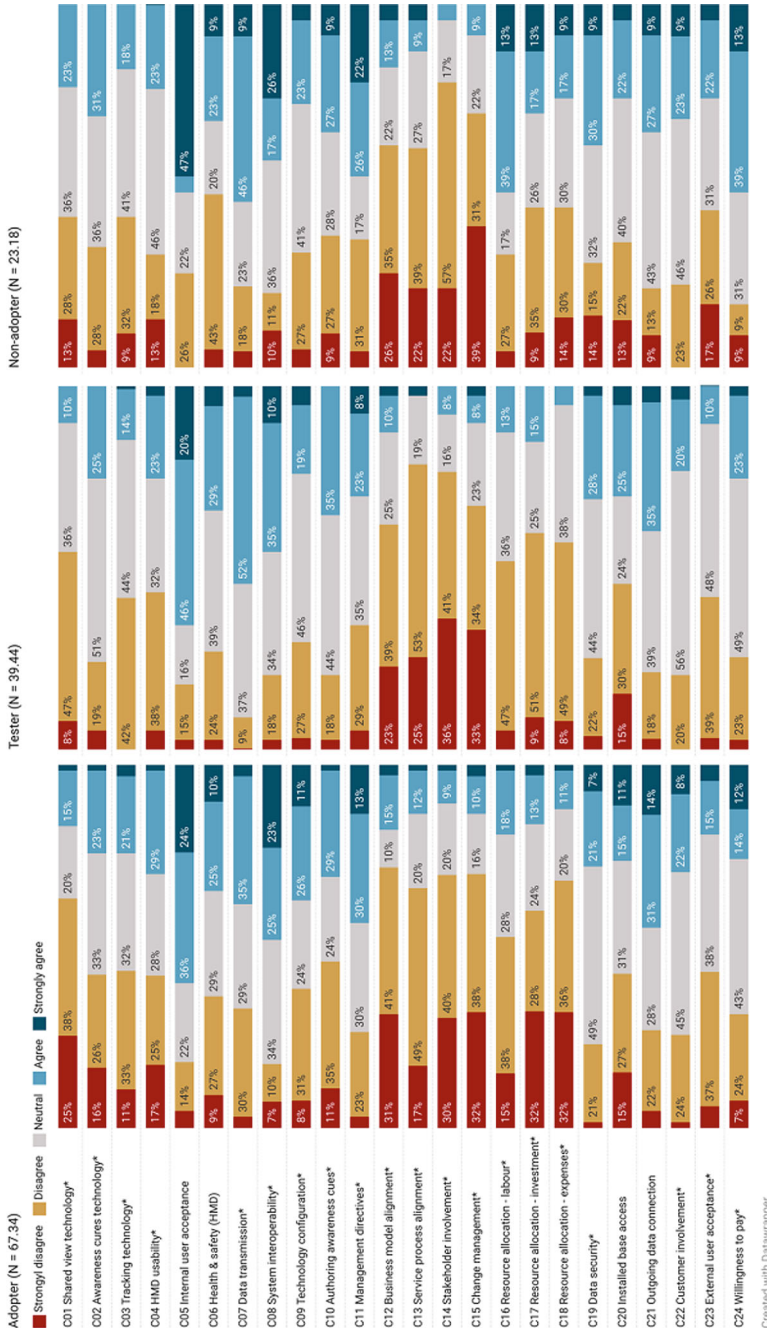


Fig. 2 Frequency stacked bar charts (pooled results) for challenges and barriers. Note: *Indicates reverse coded items

($p = 0.400$, $r = 0.145$), or Non-adopters to Testers ($p = 1.000$, $r = -0.047$). However, Adopters differ significantly from Testers ($p = 0.065$, $r = 0.241$). While 55% of Testers agree and 9% disagree that undisturbed transmission of data (e.g., video stream, 3D models) is an issue, only 38% of Adopters agree and 33% disagree. This suggests that Adopters might have overcome data transmission issues during implementation.

Statistically significant differences between the adoption status groups were also found in all three items (i.e., C16–C18) asking for the management's willingness to provide necessary resources to implement CAR. The mean rank score for *C16 labor resource allocation* ($H(2) = 8.898$, $p = 0.012$) was 45.57 for Non-adopters, 69.13 for Testers, and 70.19 for Adopters. Pairwise comparisons revealed significant differences between Adopters and Non-adopters ($p = 0.013$, $r = -0.301$), as well as between Testers and Non-adopters ($p = 0.037$, $r = -0.316$). However, no significant differences were found between Adopters and Testers ($p = 1.000$, $r = -0.016$). The majority of Non-adopters (52%) agree with the statement that their management would not be willing to provide the necessary resources in the form of employees for a potential implementation project. This is in sharp contrast to the frequency of agreement and disagreement among Adopters (19% agree, 52% disagree), and Testers (13% agree, 52% disagree).

In terms of the *C17 investment resource allocation*, we also found a statistically significant difference, $H(2) = 5.247$, $p = 0.076$, with a mean rank score of 52.03 for Non-adopters, 63.27 for Testers, and 71.39 for Adopters. However, there were no significant differences between Testers and Non-adopters ($p = 0.717$, $r = -0.149$), as well as between Adopters and Testers ($p = 0.559$, $r = -0.111$). Adopters and Non-adopters ($p = 0.076$, $r = 0.235$) on the other hand significantly differ in their assessment of their management's willingness to provide initial investments necessary to the implementation project. Even though it is less explicit compared with *C16 labor resource allocation*, Non-adopters still agree more often (30%) and disagree less often (44%) than Adopters (16% agree, 60% disagree) and Testers (15% agree, 60% disagree).

A similar picture emerges regarding *C18 expenses resource allocation*. We found a statistically significant difference, $H(2) = 7.164$, $p = 0.028$, with a mean rank score of 51.59 for Non-adopters, 60.58 for Testers, and 73.11 for Adopters. Adopters and Non-adopters ($p = 0.034$, $r = 0.267$) significantly differ in their assessment of their management's willingness to allocate budget for ongoing operating costs (e.g., software license fees) to the implementation project. However, there were no significant differences between Testers and Non-adopters ($p = 0.979$, $r = -0.124$), as well as between Adopters and Testers ($p = 0.110$, $r = -0.169$). While also not as explicit as with *C16 labor resource allocation*, Non-adopters still agree more often (25%) and disagree less often (44%) than Adopters (12% agree, 68% disagree) and Testers (5% agree, 57% disagree). With p -values of 0.143 and greater, the Kruskal–Wallis H test did not find statistically significant differences in the other challenges and barriers analyzed.

5 Discussion

Interestingly, also Non-adopters agree on the potential of CAR to improve important service KPIs (e.g., *B14 remote resolution rate*, *B15 first-time fix rate*, *B17 installed base uptime*, *B18 service quality*, and *B27 customer satisfaction*). While overall Testers seem more optimistic than Adopters, Non-adopters are especially skeptic in terms of challenges and barriers involved. Since no general disagreement with the benefits and opportunities of CAR is observable in Non-adopters, we conclude that the reason for holding back may be more due to the challenges and barriers involved.

Moreover, the survey results concerning the assessment of organizational challenges and barriers (i.e., C11–C24) emphasize that adopting CAR requires a considerable amount of organizational effort, not just overcoming technical implementation challenges (i.e., C01–C10). For example, even though no significant differences between adoption status groups were observed, the extent of agreement concerning a lack of *C11 management directives*, *C05 internal user acceptance*, and *C24 willingness to pay* for remote services is strikingly conspicuous among Non-adopters. While *C11 management directives* and *C05 internal user acceptance* seem common across all adoption status groups, it's only the Non-adopters who assess *C24 willingness to pay* for remote services as a major barrier. Furthermore, the perception that management would deny the resources required is unique to Non-adopters, and the result is statistically significant, too. Consequently, this factor might be crucial to the implementation decision.

All in all, we found only a few (8 out of 53) statistically significant differences in the assessment of benefits and opportunities, and challenges and barriers between adoption status groups. Nevertheless, these differences might determine whether companies adopt CAR in support of industrial service delivery or not. Therefore, withholding the implementation of CAR might ultimately be a mélange of perceived data transmission issues as a technical roadblock to CAR, a lack of management support in terms of resource allocation, especially in providing employees for a potential implementation project, and the inability to bill remote services successfully.

6 Conclusion

The objective of this paper was to explore differences in the assessment of perceived benefits and opportunities, as well as challenges and barriers of implementing CAR among different adoption status groups. To do so, we conducted a questionnaire survey ($N = 130$) within the German capital equipment industry. We analyzed 29 benefits and opportunities, as well as 24 challenges and barriers associated with the implementation of CAR, with respect to differences between adoption status groups. We found significant differences in the response behavior among Adopters, Testers, and Non-adopters regarding 4 (out of 29) benefits and opportunities (i.e., shared view

relevance, initial training period, organizational structure, innovation capabilities), and regarding 4 (out of 24) challenges and barriers (i.e., data transmission, and resource allocation in terms of labor, investment, and expenses). In doing so, this paper provides insights into why companies in the capital equipment industry might hesitate to implement CAR, even though they recognize the technology’s potential to improve service KPIs.

Appendix 1

CAR^a questionnaire (translated from German to English, statements shortened)

Item	Questionnaire statement
B01	A live view of the on-site workspace is important for collaboration
B02	AR annotations/virtual awareness cues are important for collaboration
B03	Hands-free communication is important for collaboration
B04	CAR helps us to collect information that can be reused
B05	CAR helps us to reduce the pressure on service employees during machine downtime
B06	CAR helps us to realize knowledge transfer
B07	CAR helps to balance variations of service capacity utilization
B08	CAR helps us to plan our service deployments through remote preparation sessions
B09	CAR helps us to reduce initial training periods of field service technician recruits
B10	CAR helps us to cope with the shortage of qualified technicians available on the labor market
B11	CAR helps us to cope with the low qualification levels of customer personnel
B12	CAR helps us to cope with the low qualification levels of subcontractors/third-party providers
B13	CAR helps us to provide services to our customers during Covid-19 travel restrictions
B14	CAR helps us to solve more service cases remotely
B15	CAR helps to increase our first-time fix rate
B16	CAR helps us to reduce the time our specialists spend on service cases
B17	CAR helps us to reduce customers’ downtime
B18	CAR helps us to increase the quality of service
B19	CAR helps us to cover markets in remote regions with a small and/or dispersed installed base
B20	CAR helps us to offer services that were previously not possible
B21	CAR helps us to reduce customers’ maintenance costs through lower service prices
B22	CAR helps us to eliminate service value chain stages by supporting customers directly
B23	CAR helps us to integrate subcontractors/third-party service providers
B24	CAR helps us to redesign the organization through de-/centralization of service knowledge

(continued)

(continued)

Item	Questionnaire statement
B25	CAR helps us to gain additional revenue with innovative service products
B26	CAR helps us to increase our service margin through savings in resources
B27	CAR helps us to improve customer satisfaction
B28	CAR helps us to demonstrate our innovation capabilities toward customers
B29	CAR helps us to improve the attractiveness of service careers
C01	Live view technologies are technically mature enough
C02	AR annotations/virtual awareness cues are technically mature enough
C03	The fixation of AR annotations/virtual cues to real objects works well
C04	HMDs are technically mature enough
C05	Internal user acceptance is a problem
C06	HMDs are problematic in terms of occupational health and safety
C07	CAR systems ensure the undisturbed transmission of data
C08	Integrating CAR with other systems is easy to implement
C09	Configuring commercial software and hardware components is easy to accomplish
C10	Creating AR annotations/virtual cues is easy to implement for remote experts
C11	Our management sets clear strategic goals for CAR implementation
C12	Adapting the business model and/or service contracts is required when implementing CAR
C13	Operational service processes need to be adapted during CAR implementation
C14	Stakeholders from different departments should be represented in the implementation team
C15	Change management is required to actively work toward internal user acceptance of CAR
C16	Our management is willing to provide the necessary employees for the implementation project
C17	Our management is willing to provide the necessary initial investments
C18	Our management is willing to provide the necessary budget for ongoing operating costs
C19	CAR systems ensure data security throughout the value creation network
C20	Our customers deny access to products when technicians are equipped with smart devices
C21	Our customers deny outgoing data connections for external smart devices
C22	Our customers contribute to create the prerequisites for CAR at their sites
C23	Our customers' machine operators/maintenance personnel accept CAR
C24	Our customers are willing to pay for CAR services

^a The original term used, 'Remote AR', was replaced with 'CAR' for stylistic consistency throughout this paper

Appendix 2

Descriptive statistics of items with statistically significant results in the Kruskal–Wallis test (pooled results)

Item	Adopter (<i>N</i> = 67.34)			Tester (<i>N</i> = 39.44)			Non-adopter (<i>N</i> = 23.18)		
	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
B01	5	4.356	0.88063	4	4.08	0.81364	4	4.086	0.94558
B09	3	2.717	1.10894	3	3.293	1.05811	3	3.338	1.26566
B24	3	2.895	1.16477	3	3.439	0.87041	3	3.432	0.94216
B28	4	3.833	1.04537	4	4.054	0.73502	4	4.302	0.8193
C07 ^a	3	3.057	0.95055	4	3.485	0.70406	4	3.369	1.03354
C16	4	3.465	1.00313	4	3.442	0.78567	2	2.705	1.14662
C17	4	3.725	1.14387	4	3.529	0.86376	3	3.101	1.20937
C18	4	3.865	1.0342	4	3.596	0.71431	3	3.231	1.17164

^a Reverse coding

References

- Toossi A, Lockett HL, Raja JZ, Martinez V (2013) Assessing the value dimensions of outsourced maintenance services. *J Qual Maint Eng.* <https://doi.org/10.1108/JQME-04-2013-0021>
- Marcon É, Marcon A, Ayala NF, Frank AG, Story V, Burton J, Raddats C, Zolkiewski J (2022) Capabilities supporting digital servitization: a multi-actor perspective. *Indus Market Manag.* <https://doi.org/10.1016/j.indmarman.2022.03.003>
- Herterich M, Peters C, Uebernickel F, Brenner W, Neff AA (2015) Mobile work support for field service: a literature review and directions for future research. In: *Wirtschaftsinformatik proceedings 2015*, vol 10
- Fernández del Amo I, Erkoynuncu J, Vrabčič R, Frayssinet R, Vazquez Reynel C, Roy R (2020) Structured authoring for AR-based communication to enhance efficiency in remote diagnosis for complex equipment. *Adv Eng Inform.* <https://doi.org/10.1016/j.aei.2020.101096>
- Vorraber W, Gasser J, Webb H, Neubacher D, Url P, Teti RDD (2020) Assessing augmented reality in production: remote-assisted maintenance with HoloLens. *Procedia CIRP.* <https://doi.org/10.1016/j.procir.2020.05.025>
- Jalo H, Pirkkalainen H, Torro O, Kärkkäinen H, Puhto J, Kankaanpää T (2018) How can collaborative augmented reality support operative work in the facility management industry? In: *Proceedings of the 10th international conference on knowledge management and information sharing*, pp 41–51. <https://doi.org/10.5220/0006889800410051>
- Jonsson K, Westergren UH, Holmström J (2008) Technologies for value creation: an exploration of remote diagnostics systems in the manufacturing industry. *Inform Sys J.* <https://doi.org/10.1111/j.1365-2575.2007.00267.x>
- Müller M, Stegelmeyer D, Mishra R (2023) Development of an augmented reality remote maintenance adoption model through qualitative analysis of success factors. *Oper Manag Res.* <https://doi.org/10.1007/s12063-023-00356-1>

9. Breitreuz D, Müller M, Stegelmeyer D, Mishra R (2022) Augmented reality remote maintenance in industry: a systematic literature review. In: Lecture notes in computer science. LNCS, vol 13446, pp 287–304. <https://doi.org/10.1007/978-3-031-15553-6>
10. Marques B, Silva S, Alves J, Rocha A, Dias P, Santos BS (2022) Remote collaboration in maintenance contexts using augmented reality: insights from a participatory process. *Int J Interact Design Manuf.* <https://doi.org/10.1007/s12008-021-00798-6>
11. Cavaleri J, Tolentino R, Swales B, Kirschbaum L (2021) Remote video collaboration during COVID-19. In: Proceedings of the 32nd annual SEMI advanced semiconductor manufacturing conference (ASMC). IEEE. <https://doi.org/10.1109/ASMC51741.2021.9435703>
12. Li X, Voorneveld M, de Koster R (2022) Business transformation in an age of turbulence—lessons learned from COVID-19. *Technol Forecast Soc Change.* <https://doi.org/10.1016/j.techfore.2021.121452>
13. Jalo H, Pirkkalainen H, Torro O, Pessot E, Zangiacomi A, Tepļjakov A (2022) Extended reality technologies in small and medium-sized European industrial companies: level of awareness, diffusion, and enablers of adoption. *Virt Real.* <https://doi.org/10.1007/s10055-022-00662-2>
14. Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. *IEICE Trans Inform Syst* 12:1321–1329
15. Gergle D, Kraut RE, Fussell SR (2013) Using visual information for grounding and awareness in collaborative tasks. *Hum Comput Interact.* <https://doi.org/10.1080/07370024.2012.678246>
16. Fang D, Xu H, Yang X, Bian M (2020) An augmented reality-based method for remote collaborative real-time assistance from a system perspective. *Mob Netw Applic.* <https://doi.org/10.1007/s11036-019-01244-4>
17. Piumsomboon T, Dey A, Ens B, Lee G, Billingham M (2019) The effects of sharing awareness cues in collaborative mixed reality. *Front Robot AI.* <https://doi.org/10.3389/frobt.2019.00005>
18. Wang P, Bai X, Billingham M, Zhang S, Wei S, Xu G, He W, Zhang X, Zhang J (2020) 3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-020-09731-7>
19. Adcock M, Gunn C (2015) Using projected light for mobile remote guidance. *Comput Support Coop Work.* <https://doi.org/10.1007/s10606-015-9237-2>
20. Bottecchia S, Cieutat J-M, Merlo C, Jessel J-P (2009) A new AR interaction paradigm for collaborative teleassistance system: the POA. *Int J Interact Des Manuf.* <https://doi.org/10.1007/s12008-008-0051-7>
21. Sanna A, Manuri F, Piumatti G, Paravati G, Lamberti F, Pezzolla P (2015) A flexible AR-based training system for industrial maintenance. In: Lecture notes in computer science. LNIP, vol 9254, pp 314–331. https://doi.org/10.1007/978-3-319-22888-4_23
22. Masood T, Egger J (2019) Augmented reality in support of Industry 4.0—implementation challenges and success factors. *Robot Comput Integr Manuf.* <https://doi.org/10.1016/j.rcim.2019.02.003>
23. Si2 Partners (2018) Augmented reality in service: ready for prime time? Management Report 2018. Technology in Service
24. Rapaccini M, Porcelli I, Espindola DB, Pereira CE (2014) Evaluating the use of mobile collaborative augmented reality within field service networks: the case of Océ Italia—Canon Group. *Prod Manuf Res.* <https://doi.org/10.1080/21693277.2014.943430>
25. Aquino S, Rapaccini M, Adrodegari F, Pezzotta G (2023) Augmented reality for industrial services provision: the factors influencing a successful adoption in manufacturing companies. *J Manuf Technol Manage.* <https://doi.org/10.1108/JMTM-02-2022-0077>
26. Porter ME, Heppelmann JE (2017) Why every organization needs an augmented reality strategy. *Harv Bus Rev* 95:46–57
27. Honaker J, King G, Blackwell M, Amelia II (2011) A program for missing data. *J Stat Softw* 45
28. Newman DA (2014) Missing data. *Organ Res Methods.* <https://doi.org/10.1177/1094428114548590>
29. Tsiriktsis N (2005) A review of techniques for treating missing data in OM survey research. *J Oper Manag.* <https://doi.org/10.1016/j.jom.2005.03.001>

30. van Buuren S, Groothuis-Oudshoorn K (2011) Mice: multivariate imputation by chained equations in R. *J Stat Softw* 45
31. Little RJA (1988) Missing-data adjustments in large surveys. *J Bus Econ Stat* 6:287–296
32. Rubin DB (1986) Statistical matching using file concatenation with adjusted weights and multiple imputations. *J Bus Econ Stat* 4:87–94
33. Heymans MW, Eekhout I (2019) Applied missing data analysis with SPSS and (R) studio. Heymans and Eekhout, Amsterdam
34. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. *J Am Stat Assoc.* <https://doi.org/10.1080/01621459.1952.10483441>
35. Field A (2013) *Discovering statistics using IBM SPSS statistics. And sex and drugs and rock 'n' roll*, 4th edn. MobileStudy. Sage, Los Angeles, London, New Delhi, Singapore, Washington DC

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Cyber Victimization: Tools Used to Combat Cybercrime and Victim Characteristics



Marc Dupuis  and Emiliya Jones 

Abstract Cyber victimization is explored through the lens of end users and the tools they use to combat cybercrime. These tools are important in mitigating a variety of threats to the confidentiality, integrity, and availability of information and associated systems for end users, whether through intentional criminal activity, accidents, or system/device malfunction. This is done by examining the characteristics of individuals and the degree to which they have been a victim, including various demographics and personality types. A large-scale survey was used to aid in this endeavor. Gender identification, household income, and education were all related to cybersecurity tool usage. Those that identified as male, had a higher reported household income, and/or were more educated, were more likely to use a variety of cybersecurity tools. Additionally, individuals with higher levels of neuroticism were less likely to use a number of cybersecurity tools. Implications and future directions are discussed.

Keywords Cyber victimization · Personality · Cybersecurity tools · Demographics

1 Introduction

The rapid expansion of technology use and integration has become self-evident through both public and personal spaces, including digitalization and the increased value of data. In parallel, the expansion of technology has uncovered unique opportunities for offenders to access unauthorized data, gain unauthorized systems control, and commit cybercrimes. Cybercrime is not as easily identified as traditional crime and can vary in severity, ranging from fraud and identity theft to threats and intimidation [1]. The concept of cybercrime is new within many societal, legal, political, and academic spaces, with the difficulties in understanding cybercrime stemming

M. Dupuis (✉) · E. Jones
University of Washington, Bothell, WA 98011, USA
e-mail: marcjd@uw.edu

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_11

from the simultaneous lack of consistency in terminology and cybercrime legislation across varying jurisdictions [2].

In this paper, cybercrime is a generalized term to describe both cyber-enabled crimes where the computer is utilized in a supporting role to assist with cybercrime execution, and cyber-dependent crimes identified as a result of computer availability [3]. As with traditional crime, the impact of victimization on users may vary based on numerous factors. The victimization experienced by end users may be explained through specific threats to the confidentiality, integrity, or availability of their information and systems, as well as the potential tools and other mitigations available to help thwart those threats [4].

This paper is organized as follows. The background explores cyber-dependent and cyber-enabled crime through the CIA cybersecurity pillars [1]. The background also investigates existing assumptions and research on the impact of cybercrime victimization on end users, including personality and behavior traits. Cybersecurity tool use and non-use are also investigated to examine victimization experience further. Next, the problem is identified, along with the hypotheses made based on existing research. As part of the study, the survey-based method deployed is described, along with the gathered data and analysis. Through these efforts, we seek to address the following five research questions:

- RQ1: How might cybersecurity victimization relate to the use or non-use of cybersecurity tools?
- RQ2: What is the role of demographics in the use and non-use of cybersecurity tools?
- RQ3: What is the role of demographics in the types of cybersecurity victimization experienced by end users?
- RQ4: To what extent does personality help explain the victimization experienced by end users?
- RQ5: To what extent does personality help explain the use and non-use of cybersecurity tools by end users?

Given the exploratory nature of this research, we will be examining how the data from the survey provides insight into these five research questions rather than specific hypotheses.

2 Background

2.1 Types of Cybercrimes

Cybercrime encompasses a range of criminal activities conducted over the internet or within a computer or network system by taking advantage of vulnerabilities in the system or infrastructure. Cybercrime may be categorized as cyber-dependent or cyber-enabled but can be executed in parallel [5]. Cybercrime impacts one or more

of the three CIA objectives of cybersecurity, which include confidentiality, integrity, and availability. Confidentiality ensures that data are not disclosed or available to unauthorized third parties. Integrity refers to the accuracy of the stored data, in which the data are trustworthy and has not been modified or altered by an unauthorized party. Availability, as per the CIA objectives, refers to data being accessible and usable upon the request of an authorized entity [1]. The CIA triad presents critical baselines for maintaining data and protecting systems against cybercrimes and threats. When a cybercrime occurs, the organization or system has likely been unsuccessful in properly enforcing one or more of the identified pillars of the CIA triad. Cyber-dependent and cyber-enabled crimes impact data interception, modification, theft, network crime, access crime, and other content-related cybercrimes [5].

Cyber-Dependent Crime. Cyber-dependent crimes can be identified as either unauthorized intrusions into computer networks (e.g., hacking) or cause disruption of system functionality or network space (e.g., distributed denial of service attacks (DDoS), viruses) [5]. Both involve the execution of the identified crime as dependent on the use of technology, the internet, or other forms of technological communication. Cyber-dependent crimes often derive from malware or malicious software developed to negatively impact end users, which may include negatively impacting system operations [6]. Malware can impact system operations and be leveraged to gain unauthorized access to personal data. For instance, viruses and worms are some of the more commonly known types of malware that can self-replicate, increasing the crime's scope. Viruses and worms damage or delete hardware, software, or other accessible files by the offender. Viruses are considered cyber-dependent crimes that require a host on the system and human intervention or action to open and ultimately run the infected file, disk, or other host. On the other hand, worms can be autonomous and do not require human initiation. Another example of cybercrime is trojans, a form of malware that appears to be a legitimate system or program on a device but ultimately provides illegal access and performs illicit functions unknown to end users, such as data corruption and theft.

Similarly, spyware is a type of malware intended to gather sensitive or personal data from infected systems with varying software types, with the primary objective being to gather sensitive data. For instance, key-logging software is a subtype of spyware that captures a user's keystrokes to enable access to data entry on the device, including access to passwords or financial data. Other examples of spyware include free or unwanted software known as adware and spyware that captures screenshots of victim's systems [5]. There are a variety of cyber-dependent crimes that can vary in scope and impact to end users.

Cyber-Enabled Crime. Cyber-enabled crimes, such as online fraud, digital piracy, and cyberbullying, are considered traditional crimes that are increased in severity or ease for the offender using computer networks, systems, or other technology. The most common cyber-enabled crimes derive from fraud or theft of personal data, financial data, or intellectual property [5]. The victims of cyber-enabled crimes are not limited to large corporations or businesses but can be individuals of the general

public. Although different motives have been discovered and analyzed to understand cybercrime incentives, including hacker classifications and typologies, one of the most typical incentives across varying-skilled hackers is financial gain [7]. For instance, electronic financial fraud, including online banking, internet-enabled card-not-present (CNP) fraud, and e-commerce fraud, are forms of cyber-enabled crime that could be executed without the use of the internet or technology but increase impact with the use of technology. In these instances, the business and customer may both be negatively impacted.

Similarly, fraudulent sales through online sites or auctions, as well as the sale of counterfeit products, are deemed as retail misinterpretations that may impact both businesses and consumers. Pharming is a similar cyber-enabled fraud method that redirects users to fake websites, often from phishing scams or mass-marketing fraud. Phishing or spear-phishing attempts for cyber-enabled fraud are often emails or other forms of communication used to gain access to personal or corporate information from users, such as passwords or financial data, and may be more targeted with spear-phishing to gain information from a particular target or audience by tailoring communication based on gathered information of the audience [5]. Offenders can leverage other content-related methods with cyber-enabled crime, such as cyber-sex or relationships, cyber threats, or cyber defamation, to access to end-user data by deceiving victims through emotional, authoritative, or criminal persuasion [8].

2.2 Impact of Cybercrime on End Users

Cybercrime and attacks have become more prevalent worldwide and as commonplace as technology and the internet, with end users as the primary targets. The impact of cybercrime is dependent on the audience and can vary drastically based on the data accessed, altered, or lack of availability, as aligned with the CIA triad objectives of the cybersecurity industry. The impact of cybercrime is not mutually exclusive, regardless of intent, and may impact businesses, communities, or individual users. Research has identified that users are more likely to respond to cybercrime's harmful effects and residual impact than individual crime. For instance, a malware cyberattack may impact an organization or critical business process that provides a resource to a community, such as water, energy or electricity, or a product or service. In these instances, the impact of the malware is reflected in the community and their loss of a resource, as opposed to the severity of the technical capabilities of cybercrime. Individuals may not necessarily be concerned or view the severity of potential impact based on the cybercrime method, but rather the impact they individually experience.

Research has identified that impact may vary between social and psychological impact, also known as emotional and behavioral, as well as physical and financial or material impact [9]. Social impact refers to the effect of cybercrime on a community or individual that causes disruption within daily life and increased loss of confidence or negative emotions toward the use of technology or cyberspace [10]. Psychological

impact may overlap with social impact concerning individual anxiety, worry, anger, or other emotional impact [9].

Similarly, physical impact may be intertwined with emotional impact. However, it is essential to note that it is not as likely to be directly physically impacted by cybercrime but rather indirectly impacted through victimization. Instances of indirect physical impact include but are not limited to, sleep deprivation, headaches, and weight loss, as per the victimization experience of end-users. The financial impact may refer to loss of finances from cybercrime, such as phishing or ransomware, and results in monetary loss. Although indirect costs may also be attributed to cybercrime impacts, such as the cost of time and resources to manage the effects of cybercrime or the loss of income due to the inability to work, it may also be considered when assessing cybercrime [10]. The impact of cybercrime and the victimization experienced by end users can vary on preexisting factors or directly as a result of the cybercrime.

Business Impact. Cybercrime continues to grow in frequency, severity, and negative impact on businesses. The impact of cybercrime on businesses can result in unauthorized access to business or customer personal or financial data, denial of service to company systems or resources, or monitoring of company activity on networks or systems, among other cyber-dependent or cyber-enabled crimes [8]. The expansion of information technology and the simultaneous increase in cybercrime have affected economic sectors, including the vulnerability of critical infrastructures, such as businesses and e-commerce, with security breaches and data theft. Not only have cybercrimes increased in frequency, but the complexity and severity of cybercrimes have increased significantly. Alternatively, cybercrimes may be undetected and confidential data may be exposed to unauthorized parties.

The impact of cybercrime is not limited to businesses and end users but has potential effects on information systems, reputation, finance, and stakeholder loss of confidence. Evaluating the impact of cybercrime on businesses is complex, as the most immediate quantitative evaluation stems from financial and monetary loss but may include loss in market value and reputation from consumers and stakeholders [11].

The impact of cybercrime is not only quantified by the monetary and financial losses of the business but may influence the preliminary cost of protection of the business. The costs associated with vulnerability and detection management, updating internal procedures, and purchasing secure software and hardware have developed a complex space for businesses to invest upfront protection costs while testing and monitoring the systems against emerging cybercrimes. Inadvertently, these increased costs for the business to operate in a secure space and protect both business and customer data are often passed on to the customer through increased prices of their products or services.

As a result, customers of businesses may become influenced or participate in a subculture of cyber-activists that may either protest physically by chaining themselves to buildings or participate in cyberattacks that typically perpetrate systems through a denial-of-service attack in retaliation. For instance, in December 2010,

PayPal was attacked by dozens of people claiming to be part of Anonymous, a well-known hacker group. The denial-of-service attack did not result in a complete shut-down of PayPal. However, this cyber-enabled crime may impact short and long-term revenue, especially for businesses relying predominantly on sales [8].

The impact of cybercrime on businesses is not always evident through monetary loss but can accrue reputational damage for the business. Based on data collected between 2002 and 2018 from 45 different companies, Oxford Academic identified that the largest and most significant breaches are associated with a 5–9% decline in reputation intangible capital following a data breach [12]. For this reason, businesses must consider precautionary mitigation measures along with incident response management and planning.

User Victimization Impact. End users may be impacted by cybercrime in various ways, depending on the type of crime, severity, and residual impact. Research has indicated that users experience negative emotions, including anger, annoyance, frustration, and feelings of being deceived or cheated, which appear to be the most common reactions to cybercrime, with the potential to develop into longer-term psychological effects [13]. Most theories and studies on the victimization of crime focus on traditional crime.

However, with cyberspace continuously expanding, it becomes vital to understand the impact and victimization of cybercrime victims and their experiences based on the unique characteristics of cybercrime. For instance, the physical distance between a victim and a cyber offender and the use of technology may alter how individuals are impacted, as opposed to assumptions from traditional crimes.

Theories such as the shattered assumptions theory (SAT) are applied to assess the victimization impact of traditional crimes. However, the extent to which the existing traditionally focused crime theories apply to cybercrimes has yet to be defined. SAT in the scope of cybercrime implies that victimization has led to the impairment of basic, optimistic assumptions people internalized about themselves and the world around them, such as being invulnerable and having autonomy of self and choice, as well as the world being controllable and comprehensible. Based on this theory, destroying the identified assumptions regarding technology and cyber communication may result in psychological, social, and behavioral effects. The implied effects of this theory may apply to individuals who have experienced cybercrime victimization and may be linked to the unique characteristics of cybercrime and reduce the sense of control an individual may have as a result [10].

The unique characteristics of cybercrime, such as the type of crime, can influence end users' assumptions of a system, technology, or the internet and result in reactions that range from anxiety to feelings of helplessness and even physical symptoms such as sleep disturbance [10]. Similarly, routine activity theory (RAT) has been applied to the analysis of cybercrime victimization to understand how a user's lifestyle or daily routine and activities may create opportunities for users to be victims of cybercrime [14]. According to this theory, crime results when a motivated offender, a suitable target, and the lack of a capable guardian are present simultaneously [10]. Differing from the SAT theory, the RAT theory implies that although it is essential to understand

the offender's motivation and the circumstances in which the crime is committed, it is more important to understand the circumstances that produce crime and, as a result, prevent the crime [14]. Thus, examining cybercrime and the victimization impact of users is complex by nature, and existing theories applied to traditional crime may be leveraged.

2.3 End-User Personality Impact on Cybercrime Victimization

In information and cybersecurity compliance, personality may predict behaviors associated with an individual's intentions, but individual behavior may not always correlate with intent. The "Big Five" personality model investigated the relation of personality traits to individual behavior. This model has been previously applied in studies to investigate the likelihood of cybersecurity victimization. However, to improve analysis, there needs to be more consensus on the most relevant factors to determine personalities with victimization experience and compliance.

The model comprises five personality factors: conscientiousness, openness, agreeableness, neuroticism, and extraversion. Conscientiousness refers to the impulse control behaviors that assist individuals with task and goal completion (e.g., planning, organizing, and delaying gratification). Openness refers to the extent of originality and complexness of an individual's mind. Agreeableness refers to an individual's attitude toward others, including trust. Similarly, extraversion is a trait that identifies an individual as having a sociable and energetic approach to others and the world. On the other hand, neuroticism includes anxiety and sadness, often referred to as a contrast between emotional stability and extraversion.

One analysis of the relationship between the Big Five model traits and cybersecurity dimensions (secure behavior, self-efficacy, and privacy attitudes) discovered that conscientiousness was linked to individuals who tend to engage in secure behaviors regarding technology and online behavior or habits. Openness was associated with positive self-efficacy, while neuroticism was negatively associated with self-efficacy. Concerning privacy attitudes and security concerns, conscientiousness, neuroticism, and extraversion were all positively associated traits [15]. At the same time, other research has indicated agreeableness as the only positive associated trait with privacy and information concerns [16]. It is essential to recognize that other factors may influence the relationship between personality and cybersecurity, including age, gender, and other individual-specific demographic factors [15].

Negative personality traits have also been explored to understand better the correlation between personality and cybercrime victimization. For instance, Machiavellianism, narcissism, and psychopathy have been noted as negative behavioral traits pertaining to cybersecurity and compliance. Machiavellianism refers to behavior that focuses on self-interest and often involves manipulation, deceit, and exploitation

of circumstances or individuals for self-interested purposes that often negate self-discipline and compliance. Narcissism in cybersecurity refers to individuals who exhibit negative behaviors of self-entitlement, resistance to criticism or feedback and strive for either attention or admiration from peers.

However, in cybersecurity, narcissism may influence non-compliance with security policies or best practices, similar to psychopathy, in which the behavioral traits highlight disregard of guidelines or other individuals. These behavioral traits are typically associated with insider threats within organizations [17]. However, the identified negative traits may be further explored to understand the victimizations of end users who exhibit the traits and may not exhibit the assumed adverse reactions or personality traits, as described by the Big Five model.

2.4 Cybersecurity Tool Use and Non-Use

End users may manage their use of technology and digital assets by increasing security using available cybersecurity tools, including password managers, multi-factor authentication, password complexity requirements, and other mitigations. For instance, a common perception of internet security and privacy stems from using a virtual private network (VPN) that creates an encrypted connection over an unsecured network. VPNs are primarily used for protecting users against cybercrime on low-security networks, such as publicly available networks, but are also frequently used in organizations. A VPN is also less expensive than a private wide area network (WAN), which may increase user appeal. Research has identified that VPNs have increased on a broader scale than organization and government use, as more users are using VPNs to safeguard their data and ensure data privacy. It has been identified that India, Vietnam, Thailand, and China are the top contributing countries to the use of VPNs, intending to hide their online activities from government censoring and viewing. The particular countries identified have stricter censorship rules, which appear to increase the use of VPNs [18]. It appears that factors such as geographical location may influence the use or non-use of available cybersecurity tools.

Another commonly used cybersecurity tool is password managers, which intend to store and handle user's passwords for different solutions and applications. Scaring individuals into creating strong and effective passwords is one possible approach, but it may not only be ineffective, it may also cause unintended negative consequences [19, 20]. From a practical perspective, password managers are intended to assist with storing different vendor-managed credentials for different services. However, due to poor usability and limited user experience, users may not find these tools helpful or as secure as intended. For these reasons, password managers unavoidably have become desired targets of cyber criminals. To counter potential authentication breaches, alternatives such as multi-level authentication such as two-factor authentication have become available for users, either as requirements or optional layers of security. There are many options for two-factor authentication, such as time-based

one-time passwords (e.g., email, SMS) or authenticator applications, pin codes, or image-gestured passwords [21].

With multi-factor authentication, it becomes more difficult for a cybercriminal to complete a remote attack, regardless of access to passwords and account information, since the second authentication factor is not designed to be local to the solution. Despite the benefits of multi-factor authentication, the potential impact on user experience, such as speed and productivity, may be negatively viewed and not deemed usable. Research has indicated that the availability of a second-factor device, depending on the initial device, may be a deficit to the usability of two-factor authentication. Also, research indicates that users may not view their data as valuable, and it becomes hard for them to justify using cybersecurity tools as the likelihood and impact are subjective [22].

2.5 Problem

Extensive research and data values provided by organizations, government agencies, and other data have shown a massive shift in the use of technology and its impact on operational use. The increased use of technology has developed a dependency on digitalization and the Internet-of-Things (IoT). These dependencies have increased malicious cybercrimes, such as malware attacks, data breaches, denial of service incidents, social engineering, phishing, and overall unauthorized access to data that may be altered or unavailable. Research and data provided by U.S.-based news agencies identify that one in five online consumers in the United States have been victims of cybercrime within the last two years.

Likewise, 23 percent of people worldwide will fall for spear-phishing attacks. At the same time, web pages are infected on average every 4.5 s [8]. Cybercrime statistics identify that according to the FBI's records of internet crime, with 800,944 complaints registered in 2022, predictions of increased breaches and the cost of impact from breaches will continue to increase. The identified complaints and analysis show that 80% of the reported cybercrimes are linked to phishing attacks [23].

With the increase in cybercrime, there has also been an increase in the identifiable types of cybercrime. For instance, in 2010, research noted that less than 50 million identified unique malware executables are known within the security industry. By 2012, the original quantity had doubled to 100 million, with an 80% increase in 2019 to 900 malicious executables. Also, cybercrime is associated with damaging monetary loss and costs to remediate impact and fund preemptive protection from increased cybercrime. Research estimates that a data breach, on average, costs 400 billion USD to the global economy [24]. Thus, it becomes critical that organizations not only adopt and implement strong cybersecurity standards, policies, and technical mitigations, but the end-user experience must be assessed for preventative measures. Increased understanding of end-user victimization experiences and their personalities toward cybercrime and compliance may assist with preventative measures along with incident response management.

Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation, etc. does not have an indent, either.

3 Methods

The current study uses a survey developed and hosted on the Qualtrics platform with participants recruited through the crowdsourcing platform Prolific. Crowdsourcing platforms like Prolific provide researchers with an opportunity to obtain a more diverse sample of participants than students and generally have high quality so long as various quality control measures are implemented [25–27].

Prior to data collection, ethics approval was sought and obtained through the Institutional Review Board. It was determined that the study qualified for exempt status and thus did not merit a full review given the low-risk nature generally associated with survey research.

3.1 Compensation

Participants were compensated \$3.00 for their participation in the study. Most participants (79.9%) indicated that the compensation received was either comparable to other projects (74.6%) or easier for the money (5.3%), while some (20.1%) indicated that more effort was required for the compensation received. The median completion time was approximately 15 min, which resulted in an hourly rate of about \$12.

3.2 Quality Control

There were 504 participants that completed the survey with 473 of those participants passing all eight quality control questions, which resulted in a rejection rate of 6.15%. The quality control questions consisted of questions with obvious answers, repeated gender identification and ethnicity questions at the beginning and end of the survey, and questions that told them explicitly which answer to select. Additionally, participants had to have already completed at least 1000 prior assignments with an approval rating of at least 98%. Since the survey was constructed using the English language, participants were also required to be fluent in English.

3.3 Demographics

All participants that completed the survey were residents of the United States with 47.6% indicating they lived in a suburban community, followed by 37.4% living in the city or urban community and 15.0% living in a rural community. Participants varied in age between 18 and 70–74 with 23% of participants between 18 and 29, 35.7% of participants between 30 and 39, 18.6% of participants between 40 and 49, with the remaining 22.6% of participants over the age of 50.

A majority of the participants identified as male (57.1%), followed by female (40.2%), non-binary (2.5%), with 0.2% of participants preferring not to provide their gender identification. Most participants identified as White (64.3%), followed by Black/African American (12.5%), Asian/Pacific Islander (9.9%), Hispanic/LatinX (8.5%), Other/Multi-Racial (4.0%), with the remaining 0.8% identifying as Native American / Alaskan Native. The participants were generally well-educated with 54.8% having earned a Bachelor's degree or higher.

3.4 Materials

Previously developed and validated instruments were used when possible. For example, the Big Five Inventory was used to assess the personality characteristics of participants [28–30]. Additionally, the identification of various cybersecurity tools and reasons for their use and non-use was adapted from other research [31]. Five specific tools/mitigations were examined in the context of the current study: anti-malware software, password manager, backup solution, virtual private network (VPN), and two-factor authentication. Additionally, other than two-factor authentication, these tools were differentiated between their use on mobile devices and tablets versus a laptop/desktop computer.

4 Results and Analysis

All statistical tests performed in this section were done using SPSS, version 19.

4.1 Cyber Victimization and Cybersecurity Tools

We examined cyber victimization through two different lenses. In the first lens, we asked participants whether they have experienced specific outcomes that may have been the result of cybercrime (see Appendix I) or another cyber incident. This was important since the nomenclature often used by cybersecurity experts may not be

as easily understood by the average non-technical end user. Next, we asked participants about whether or not they had experienced specific cybersecurity incidents (see Appendix II). For this series of questions, we did use regular cybersecurity terms.

Perhaps not too surprisingly, there was a strong positive Pearson correlation between the summed totals of these two different sets of questions, $r(331) = 0.523$, $p = 0.000$. Interestingly and perhaps in further support of our delineation between specific incident outcomes as compared to specific cybersecurity incidents, the use of cybersecurity tools was only related to the former and not the latter. There was a positive Pearson correlation between those that had experienced the outcomes that often result from a cybersecurity incident and the number of tools used, $r(372) = 0.191$, $p = 0.000$.

When we examined the specific tools used more closely, we found that those that use two-factor authentication were less likely to have experienced a cybersecurity incident, $r(386) = -0.155$, $p = 0.002$. This was not found when looking at two-factor authentication and potential cyber incident outcomes. However, what was interesting is the role cybersecurity tool use on mobile devices had with the propensity for participants to have experienced a cyber incident outcome. This included the use of anti-malware software, $r(377) = -0.133$, $p = 0.009$; a password manager, $r(372) = -0.131$, $p = 0.011$, and a VPN, $r(372) = -0.125$, $p = 0.015$. In all cases, the use of these tools on their mobile devices resulted in a lower likelihood that they experienced a negative outcome generally associated with a cybersecurity incident.

4.2 *Demographics and Cybersecurity Tools*

In this section, we examine the relationship various demographics may have with cybersecurity tool usage. This includes gender identification, education, age, and household income.

Gender Identification. We conducted an independent samples t-test to assess whether there was a statistically significant difference between those that identified as male ($N = 266$) versus those that identified as female ($N = 187$) with respect to cybersecurity tool usage. The test suggests that those that identify as male use a greater number of cybersecurity tools ($m = 4.73$) versus their female counterparts ($m = 4.02$). The t-statistic was -3.522 , with $df = 439.294$ ($p < 0.001$).

Beyond the above independent samples t-test, we also conducted a chi-square test to assess the relationship between gender identification and the use of specific cybersecurity tools. None of the cells in the analysis have an expected count of less than five. Table 1 has these results.

Participants that identified as male had higher than expected counts with respect to their use of anti-malware software and VPNs on both mobile and PC type devices, as well as greater use of two-factor authentication.

Education. We conducted an independent samples t-test to assess whether there was a statistically significant difference between those that had earned a Bachelor's degree or higher ($N = 253$) versus those that had not ($N = 213$) with respect to

Table 1 Gender identification and cybersecurity tool usage

Tool	Correlation (Chi-square)	Sig. (2-tailed)
<i>Anti-malware (M)</i>	6.967	0.008
<i>Anti-malware (PC)</i>	10.055	0.002
Password manager (M)	0.712	0.399
Password manager (PC)	0.855	0.355
Backup solution (M)	0.645	0.422
Backup solution (PC)	0.571	0.450
<i>VPN (M)</i>	6.539	0.011
<i>VPN (PC)</i>	17.974	0.000
<i>Two-factor authentication</i>	9.076	0.003

Bolditalics and asterisk shows statistically significant result

cybersecurity tool usage. The test suggests that those that have higher levels of education use a greater number of cybersecurity tools ($m = 4.65$) versus those with lower levels of education ($m = 4.20$). The t-statistic was -2.205 , with $df = 462.113$ ($p = 0.028$).

In addition to the recategorization of education as a dichotomous variable (Bachelor’s degree or higher), we examined it in its original ordinal form. Our goal was to determine the relationship between education and whether or not someone chooses to use a specific tool. Therefore, we use a rank biserial test to examine these relationships. The results may be found in Table 2.

An individual’s education does seem to play a role in their use of a backup solution, whether for their mobile device or PC, as well as their use of a VPN on their mobile device. Those with higher levels of education are more likely to use such tools.

Table 2 Education and cybersecurity tool usage

Tool	Correlation (Spearman)	Sig. (2-tailed)
Anti-malware (M)	- 0.074	0.108
Anti-malware (PC)	- 0.013	0.777
Password manager (M)	- 0.006	0.891
Password manager (PC)	- 0.010	0.826
<i>Backup solution (M)</i>	- 0.114*	0.013
<i>Backup solution (PC)</i>	- 0.168**	0.000
<i>VPN (M)</i>	- 0.1 × 03*	0.026
VPN (PC)	- 0.055	0.233
Two-factor authentication	0.002	0.965

Bolditalics and asterisk shows statistically significant result

Age. We conducted multiple independent samples t-tests to determine if there was an age delineation that resulted in higher versus lower cybersecurity tool usage. This included examining those 18–34 versus 35 and over, 18–39 versus 40 and over, and finally those between 18 and 49 versus those 50 and over. In all cases, a statistically significant relationship was not found. Thus, the number of cybersecurity tools used does not appear to be related to age.

In addition to the recategorization of age as a dichotomous variable, we examined it in its original ordinal form. Our goal was to determine the relationship between age and whether or not someone chooses to use a specific tool. Therefore, we use a rank biserial test to examine these relationships. The results may be found in Table 3.

An individual's age was related to their use of both anti-malware software on their mobile device and their use of two-factor authentication. Those older in age were more likely to use anti-malware software on their mobile device, while at the same time being less likely to use two-factor authentication.

Income. We conducted an independent samples t-test to assess whether there was a statistically significant difference between those that have a household income under \$100,000 per year ($N = 351$) versus those that have a household income greater than \$100,000 ($N = 115$) with respect to cybersecurity tool usage. The test suggests that those that have higher levels of household income use a greater number of cybersecurity tools ($m = 5.05$) versus those with lower levels of household income ($m = 4.25$). The t-statistic was -3.218 , with $df = 174.632$ ($p = 0.002$).

In addition to the recategorization of income as a dichotomous variable, we examined it in its original ordinal form. Our goal was to determine the relationship between income and whether or not someone chooses to use a specific tool. Therefore, we use a rank biserial test to examine these relationships. The results are found in Table 4.

Income appears to be the largest demographic factor driving cybersecurity tool usage. Those higher in income are more likely to use a variety of cybersecurity tools, including a password manager, backup solutions, and VPN on both their mobile and

Table 3 Age and cybersecurity tool usage

Tool	Correlation (Spearman)	Sig. (2-tailed)
<i>Anti-malware (M)</i>	<i>- 0.146**</i>	<i>0.001</i>
Anti-malware (PC)	- 0.072	0.120
Password manager (M)	0.049	0.287
Password manager (PC)	0.035	0.454
Backup solution (M)	0.028	0.540
Backup solution (PC)	- 0.024	0.602
VPN (M)	0.014	0.766
VPN (PC)	0.021	0.646
<i>Two-factor authentication</i>	<i>0.152**</i>	<i>0.001</i>

Bolditalics and asterisk shows statistically significant result

Table 4 Income and cybersecurity tool usage

Tool	Correlation (Spearman)	Sig. (2-tailed)
Anti-malware (M)	0.034	0.466
Anti-malware (PC)	0.005	0.909
Password manager (M)	- 0.156**	0.001
Password manager (PC)	- 0.151*	0.001
Backup solution (M)	- 0.134**	0.003
Backup solution (PC)	- 0.166**	0.000
VPN (M)	- 0.149**	0.001
VPN (PC)	- 0.123**	0.008
Two-factor authentication	- 0.057	0.214

Bolditalics and asterisk shows statistically significant result

PC type devices. This suggests that income may present a barrier to higher levels of cybersecurity tool usage among participants.

Demographics and Cyber Victimization. In this section, we discuss how cyber victimization may be related to various demographics, such as gender identification, education, age, and household income.

To examine the relationship between gender identification and cyber victimization, we conducted an independent samples t-test for both cyber incident outcomes cybersecurity incidents. Our goal was to assess whether there was a statistically significant difference between those that identified as male versus those that identified as female with respect to their self-reports of having experienced outcomes likely related to cybersecurity incidents and cybersecurity incidents themselves.

The test suggests that those that identify as female have experienced fewer outcomes related to cybersecurity incidents ($m = 2.73$) versus their male counterparts ($m = 3.26$). The t-statistic was $- 2.733$, with $df = 367$ ($p = 0.007$). Likewise, a similar finding exists for actual cybersecurity incidents. The independent samples t-test suggests that those that identify as female have experienced fewer actual cybersecurity incidents ($m = 1.80$) versus their male counterparts ($m = 2.20$). The t-statistic was $- 2.316$, with $df = 374$ ($p = 0.021$).

For education, age, and household income, we conducted a Spearman’s correlation coefficient test to see what relationships, if any, may exist between these demographics and cyber victimization. The only statistically significant finding was between potential cyber incident outcomes and education, $r(377) = 0.146$, $p = 0.004$. This suggests that those with higher levels of education are more likely to have experienced outcomes that may be attributed to a cyber incident.

Personality and Cyber Victimization. To examine whether there were any statistically significant relationships between cyber victimization and personality types, we conducted a Pearson correlation coefficient test. No statistically significant relationships were found.

Personality and Cybersecurity Tools. Similar to the above, we conducted a series of point-biserial correlation coefficient tests to determine if there was a relationship between the cybersecurity tools individuals' use and their personality. Table 5 has the results of the relationship between extraversion and cybersecurity tool usage.

Individuals with higher levels of extraversion were more likely to use anti-malware software on their mobile device, use a password manager or VPN on either their mobile device or PC, or use a backup solution for their PC. Their outgoing nature may increase their comfort level with trying a variety of cybersecurity tools. Likewise, their personality that may involve greater interactions with others may benefit even more from such cybersecurity tool usage.

Next, we examine the personality type agreeableness. Table 6 has the results of the relationship between agreeableness and cybersecurity tool usage.

Table 5 Extraversion and cybersecurity tool usage

Tool	Correlation (point-biserial)	Sig. (2-tailed)
<i>Anti-malware (M)</i>	– 0.095*	0.040
Anti-malware (PC)	– 0.065	0.161
<i>Password manager (M)</i>	– 0.121**	0.008
<i>Password manager (PC)</i>	– 0.099*	0.032
Backup solution (M)	– 0.103	0.025
<i>Backup solution (PC)</i>	– 0.172**	0.000
<i>VPN (M)</i>	– 0.133**	0.004
<i>VPN (PC)</i>	– 0.092*	0.045
Two-factor authentication	0.024	0.599

Bolditalics and asterisk shows statistically significant result

Table 6 Agreeableness and cybersecurity tool usage

Tool	Correlation (point-biserial)	Sig. (2-tailed)
Anti-malware (M)	– 0.045	0.327
Anti-malware (PC)	– 0.090	0.051
Password manager (M)	– 0.085	0.066
<i>Password manager (PC)</i>	– 0.135**	0.003
<i>Backup solution (M)</i>	– 0.126**	0.006
<i>Backup solution (PC)</i>	– 0.105	0.023
VPN (M)	– 0.029	0.526
VPN (PC)	0.006	0.888
Two-factor authentication	– 0.019	0.677

Bolditalics and asterisk shows statistically significant result

The results suggest that individuals that are more agreeable are more likely to use a password manager on their PC as well as employ a backup solution on both their mobile device and PC.

Next, we will examine individuals with higher levels of conscientiousness and how it may be related to cybersecurity tool usage. Table 7 has these results.

Individuals with higher levels of conscientiousness are more likely to use anti-malware software, a backup solution, and VPN on both their mobile device and PC. Generally speaking, individuals with higher levels of conscientiousness are more thoughtful and deliberate in the actions they take and decisions they make. This is perhaps reflected in their high use of a variety of cybersecurity tools.

Next, we will examine the personality type of neuroticism. Table 8 has the results.

In stark contrast to conscientiousness, those with higher levels of neuroticism are less likely to use a variety of cybersecurity tools, including anti-malware software or a VPN for their mobile device and PC, as well as a backup solution for their PC.

Table 7 Conscientiousness and cybersecurity tool usage

Tool	Correlation (point-biserial)	Sig. (2-tailed)
<i>Anti-malware (M)</i>	-0.158**	0.001
<i>Anti-malware (PC)</i>	-0.122**	0.008
Password manager (M)	-0.060	0.196
Password manager (PC)	-0.057	0.218
<i>Backup solution (M)</i>	-0.124**	0.007
<i>Backup solution (PC)</i>	-0.195**	0.000
<i>VPN (M)</i>	-0.137**	0.003
<i>VPN (PC)</i>	-0.121**	0.009
Two-factor authentication	0.007	0.873

Bolditalics and asterisk shows statistically significant result

Table 8 Neuroticism and cybersecurity tool usage

Tool	Correlation (point-biserial)	Sig. (2-tailed)
<i>Anti-malware (M)</i>	0.118*	0.010
<i>Anti-malware (PC)</i>	0.151**	0.001
Password manager (M)	0.076	0.100
Password manager (PC)	0.054	0.238
Backup solution (M)	0.048	0.298
<i>Backup solution (PC)</i>	0.164**	0.000
<i>VPN (M)</i>	0.147**	0.001
<i>VPN (PC)</i>	0.128**	0.005
Two-factor authentication	-0.039	0.403

Bolditalics and asterisk shows statistically significant result

Table 9 Openness and cybersecurity tool usage

Tool	Correlation (point-biserial)	Sig. (2-tailed)
Anti-malware (M)	-0.058	0.211
<i>Anti-malware (PC)</i>	<i>-0.162**</i>	<i>0.000</i>
Password manager (M)	0.027	0.563
Password manager (PC)	0.045	0.330
Backup solution (M)	-0.055	0.234
Backup solution (PC)	-0.077	0.096
VPN (M)	0.000	0.999
VPN (PC)	-0.046	0.314
Two-factor authentication	0.022	0.637

Bolditalics and asterisk shows statistically significant result

This is not surprising given the challenge that high levels of neuroticism often pose for individuals in decision-making. Anxiety and other negative emotions are often prevalent for those with high levels of neuroticism, which may significantly impair their ability to make a decision, especially one that may be in their best interest.

Finally, we examine the personality type openness. The results for openness and cybersecurity tool usage may be found in Table 9.

The only significant finding for the personality type of openness was with respect to the use of anti-malware software on one's PC. Those with higher levels of openness are more likely to use this cybersecurity tool.

5 Discussion

This paper examined five research questions in an exploratory manner to further provide insight on cyber victimization experienced by end users. This was done by conducting a large-scale survey using the Prolific crowdsourcing platform and the Qualtrics survey platform.

Several findings are worth noting. The greater use of cybersecurity tools, especially on mobile devices, was related to fewer outcomes that could stem from a cybersecurity incident. Additionally, individuals that identify as male are more likely to use some cybersecurity tools. This includes anti-malware software and VPN on both their mobile device and PC. Additionally, they are more likely to use two-factor authentication. Interestingly, those that identified as female reported fewer instances of both cybersecurity incidents and outcomes often related to cybersecurity incidents. Is this a function of how the devices are used or perhaps what devices are used?

Some similar disparities were observed for both education and household income. Those with higher levels of education are more likely to use a backup solution, whether for their mobile device or PC, as well as use a VPN for their mobile device.

Income had the most pronounced disparities. Those with higher household incomes were more likely to use a variety of cybersecurity tools—password manager, backup solution, and VPN for both their mobile device and PC. This may suggest that income disparities are also cybersecurity disparities.

Personality seemed to help explain to a certain extent both the use and non-use of a variety of cybersecurity tools. While any relationship between personality and cybersecurity tool usage was generally positive when it did exist, this was not true for the personality type of neuroticism. Those with higher levels of neuroticism are particularly disadvantaged in being able to use cybersecurity tools to help make their cybersecurity posture more secure.

5.1 *Limitations*

It is important to note some of the limitations inherent in this research. First, a survey was used and a crowd-sourced platform employed to recruit participants. While most participants indicated that compensation was commensurate with similar efforts in other studies, the participants nonetheless did have a financial motive to complete the survey as quickly as possible. A multitude of quality control questions and measures were developed to mitigate this issue with the results from 31 participants being discarded.

Second, a single research method was used in this study. Thus, common method bias is a concern [32, 33]. In addition to the multiple quality control procedures employed, the participants are in effect anonymous to the authors. Therefore, we do not believe common method bias is a significant concern in the current study, but it cannot be ruled out completely.

Third, data was collected from research participants at a single point in time. It is possible that the responses captured at that point in time do not fully reflect their attitudes, opinions, and beliefs. While this is unlikely to play a large role in negatively impacting the results in the aggregate, it is possible that it could impact the results.

Finally, this study was exploratory in nature. We performed various statistical tests to determine if statistically significant relationships were supported or not. And to the extent that such relationships were found, causation cannot be proven or otherwise demonstrated. Thus, it is possible that the relationships found may be related to another variable not accounted for in the current study [34].

6 Conclusion

As cybersecurity experts, organizations, educational institutions, and governments ponder how to help make end users more cyber secure, it is important to understand that the use and non-use of cybersecurity tools is multi-faceted. It is doubtful that a one-size-fits-all solution will be effective for everyone. If that can be acknowledged

by such entities, then it is likely they will encounter less frustration related to false hopes, but also achieve greater success for themselves and end users alike.

6.1 Future Research

Future research should focus on better understanding why the differences noted herein exist and how different populations may be best served through various interventions. This type of research would benefit from additional methodologies (e.g., focus groups, interviews, experiments). The disparities observed in the current study are disconcerting in many ways. It will only be through additional research that we may begin to make headway in addressing such disparities.

References

1. Hawdon J (2021) Cybercrime: victimization, perpetration, and techniques. *Am J Crim Justice* 46(6):837–842. <https://doi.org/10.1007/s12103-021-09652-7>
2. Phillips K, Davidson JC, Farr RR, Burkhardt C, Caneppele S, Aiken MP (2022) Conceptualizing cybercrime: definitions, typologies and taxonomies. *Forensic Sci* 2(2), Art. no. 2. <https://doi.org/10.3390/forensicsci2020028>
3. Bossler AM, Berenblum T (2019) Introduction: new directions in cybercrime research. *J Crime Just* 42(5):495–499. <https://doi.org/10.1080/0735648X.2019.1692426>
4. Dupuis M, Crossler R, Endicott-Popovsky B (2016) Measuring the human factor in information security and privacy. In: *The 49th Hawaii international conference on system sciences (HICSS)*, IEEE, Kauai, Hawaii
5. McGuire DM (2013) Cyber crime: a review of the evidence. Summary of key findings and implications. Home Office Res Rep 75:1–35
6. Razak MFA, Anuar NB, Salleh R, Firdaus R (2016) The rise of malware. *J Netw Comput Appl* 75(C):58–76. <https://doi.org/10.1016/j.jnca.2016.08.022>
7. Chng S, Lu HY, Kumar A, Yau D (2022) Hacker types, motivations and strategies: a comprehensive framework. *Comput Human Behav Rep* 5:100167. <https://doi.org/10.1016/j.chbr.2022.100167>
8. Das S, Nayak T (2013) Impact of cyber crime: issues and challenges. *Int J Eng Sci* 6(2):142–153
9. Bada M, Nurse JRC (2020) The social and psychological impact of cyberattacks. In: *Emerging cyber threats and cognitive vulnerabilities*, Elsevier, pp 73–92. <https://doi.org/10.1016/B978-0-12-816203-3.00004-6>
10. Borwell J, Jansen J, Stol W (2022) The psychological and financial impact of cybercrime victimization: a novel application of the shattered assumptions theory. *Soc Sci Comput Rev* 40(4):933–954. <https://doi.org/10.1177/0894439320983828>
11. Arcuri MC, Brogi M, Gandolfi G (2017) How does cyber crime affect firms? The effect of information security breaches on stock returns
12. Makridis CA (2021) Do data breaches damage reputation? Evidence from 45 companies between 2002 and 2018. *J Cybersec* 7(1): tyab021. <https://doi.org/10.1093/cybsec/tyab021>
13. Budimir S, Fontaine JRJ, Huijts NMA, Haans A, Loukas G, Roesch EB (2021) Emotional reactions to cybersecurity breach situations: scenario-based survey study. *J Med Internet Res* 23(5):e24879. <https://doi.org/10.2196/24879>
14. Ngo FT, Paternoster R (2011) Cybercrime victimization: an examination of individual and situational level factors. 5(1)

15. Lévesque FL, Fernandez JM, Batchelder D (2017) Age and gender as independent risk factors for malware victimization. Presented at the electronic visualisation and the arts (EVA 2017). <https://doi.org/10.14236/ewic/HCI2017.48>
16. Shappie AT, Dawson CA, Debb SM (2020) Personality as a predictor of cybersecurity behavior. *Psychol Popular Media* 9(4):475–480. <https://doi.org/10.1037/ppm0000247>
17. The Influence of Employee Personality on Information Security. ISACA. Accessed: Sep. 04, 2023. [Online]. Available: <https://www.isaca.org/resources/isaca-journal/issues/2021/volume-5/the-influence-of-employee-personality-on-information-security>
18. Pavlicek A (2018) Internet security and privacy in VPN 9(4)
19. Dupuis M, Jennings A, Renaud K (2021) Scaring people is not enough: an examination of fear appeals within the context of promoting good password hygiene. In: Proceedings of the 22st annual conference on information technology education, SnowBird UT USA: ACM, pp 35–40. <https://doi.org/10.1145/3450329.3476862>
20. Dupuis M, Renaud K, Jennings A (2022) Fear might motivate secure password choices in the short term, but at what cost? In: Proceedings of the 55th Hawaii international conference on system sciences (HICSS) 2022, Virtual, pp 4796–4805. <https://doi.org/10.24251/HICSS.2022.585>
21. Chaudhary S, Schafeitel-Tähtinen T, Helenius M, Berki E (2019) Usability, security and trust in password managers: a quest for user-centric properties and features. *Comput Sci Rev* 33:69–90. <https://doi.org/10.1016/j.cosrev.2019.03.002>
22. Reese K, Smith T, Dutson J, Armknecht J, Cameron J, Seamons K (2019) A usability study of five two-factor authentication methods. In: Fifteenth symposium on usable privacy and security (SOUPS 2019), pp 357–370
23. “Internet Crime Complaint Center (IC3) | Annual Reports.” Accessed: Sep. 07, 2023. [Online]. Available: <https://www.ic3.gov/Home/AnnualReports>
24. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A (2020) Cybersecurity data science: an overview from machine learning perspective. *J Big Data* 7(1):41. <https://doi.org/10.1186/s40537-020-00318-5>
25. Steelman ZR, Hammer BI, Limayem M (2014) Data collection in the digital age: innovative alternatives to student samples. *MIS Q* 38(2):355–378
26. Dupuis M, Renaud K, Searle R (2022) Crowdsourcing quality concerns: an examination of amazon’s mechanical Turk. In The 23rd annual conference on information technology education, Chicago IL USA: ACM, pp 127–129. <https://doi.org/10.1145/3537674.3555783>
27. Dupuis M, Endicott-Popovsky B, Crossler R (2013) An analysis of the use of Amazon’s mechanical turk for survey research in the cloud. In: International conference on cloud security management, seattle, Washington
28. John OP, Naumann LP, Soto CJ (2008) Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory Res* 3:114–158
29. John OP, Donahue EM, Kentle RL (1991) The big five inventory—versions 4a and 54, Berkeley: University of California. Institute of Personality and Social Research, Berkeley
30. Benet-Martínez V, John OP (1998) Los Cinco Grandes across cultures and ethnic groups: multitrait-multimethod analyses of the Big Five in Spanish and English. *J Personal Soc Psychol* 75(3):729
31. Dupuis M, Geiger T, Slayton M, Dewing F (2019) The use and non-use of cybersecurity tools among consumers: do they want help? In: Proceedings of the 20th annual SIG conference on information technology education, Tacoma WA USA: ACM, pp 81–86. <https://doi.org/10.1145/3349266.3351419>
32. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies, *J Appl Psychol* 88(5):879
33. MacKenzie SB, Podsakoff PM (2012) Common method bias in marketing: causes, mechanisms, and procedural remedies. *J Retail* 88(4):542–555
34. Krathwohl D (2004) Methods of educational and social science research: an integrated approach, 2nd edn. Waveland Press, Long Grove Ill.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Binary Segmentation of Malaria Parasites Using U-Net Segmentation Approach: A Case of Rwanda



Eugenia M. Akpo, Carine P. Mukamakuza, and Emmanuel Tuyishimire

Abstract Malaria is a significant health issue in Rwanda. Its accurate identification is essential for effective treatment. Traditional methods, such as microscopy, often face limitations in these contexts. This paper investigates how advanced machine learning techniques can address diagnostic challenges commonly encountered in resource-limited settings like Rwanda. A powerful deep learning framework known as U-Net was utilized in this study to identify different types of malaria. This method demonstrated the ability to accurately identify the disease at a highly detailed level, yielding promising results. The findings from this study could contribute to the development of computer-aided diagnostic tools specifically designed for regions with limited resources. These tools could assist healthcare professionals in decision-making processes and enhance patient outcomes.

Keywords Malaria diagnosis · U-Net architecture · Malaria parasite segmentation

1 Introduction

Malaria is a deadly disease that is typically present in tropical regions and spread by infected mosquito bites. Fever, headache, chills, nausea, vomiting, diarrhea, anemia, and respiratory distress are among the symptoms [1]. It can cause consequences like cerebral malaria, breathing issues, organ failure, anemia, and low blood sugar if left untreated. By avoiding mosquito bites, using pesticides, sleeping beneath nets, donning long pants and shirts, and using creams or sprays to repel mosquitoes, malaria can be avoided. Physical examination, symptoms, and blood tests are used to make the diagnosis. Depending on the patient's age and health status, several

E. M. Akpo (✉) · C. P. Mukamakuza
Carnegie Mellon University Africa, Kigali, Rwanda
e-mail: eakpo@andrew.cmu.edu

C. P. Mukamakuza
e-mail: cmukamak@andrew.cmu.edu

E. Tuyishimire
College of Science and Technology, University of Rwanda, Kigali, Rwanda

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_12

treatments may be available [1]. Plasmodium parasites can be categorized into five distinct species, namely *P. falciparum*, *P. vivax*, *P. ovale*, *P. knowlesi*, and *P. malariae*. Among these, *P. falciparum* and *P. vivax* are the most virulent and account for most malaria cases [2].

Malaria is a severe disease, with 241 million infections and 627,000 deaths in 2020, with Africa having the highest number of cases and fatalities [3]. Commonwealth governments committed to eliminating malaria by 2030 [1]. There have been various studies on automating the process of malaria detection in various parts of the world. The common consensus is that using the Giemsa-stained specimen/blood sample observed under microscopic slides, known as light microscopy, is the gold standard for malaria detection [4–6]. This process can be time-consuming and is only accurate based on the skill of the operator or technician. It can also be challenging in resource-limited settings, thereby having rapid diagnostic tests as the next alternative [7].

This paper is a part of a broader study in which the authors explore and suggest a digital system for monitoring malaria [8]. The proposed system aims to streamline the process of data collection, delivery, aggregation, classification, treatment reporting, repository updates, and public information services. This could potentially improve the forecasting, management, and treatment of malaria [8]. As a step in the right direction and in an effort to enhance malaria control, [9] created a data management model. The authors examine existing data management models and propose a novel one that is specifically designed for Rwanda. The specific objective of the study discussed in this paper is to enhance malaria prediction and automate the process of diagnosing malaria.

For malaria parasite detection and segmentation, numerous studies have explored both options and have found various results. Previous research in the field has shown that machine learning and artificial intelligence methods have the potential to aid in malaria detection significantly. In a foundational paper [10], advanced image analysis software and machine learning approaches were leveraged to identify malaria, showcasing the pivotal role of modern information technologies in effective disease mitigation. These methods encompassed image processing, cell segmentation, parasite identification, and feature calculation, underscoring the diversity of approaches within this domain [10].

Furthermore, recent research efforts have explored alternative methods for malaria parasite detection. For instance, the work in [5] employed scaled YOLOv4 and YOLOv5 object identification models to classify malaria parasites, achieving notable accuracy rates of 83 and 78.5%. This result suggests the potential for these algorithms to assist medical practitioners in accurately identifying and predicting the malaria stage, a crucial aspect of disease management. While much research has primarily focused on detecting *Plasmodium falciparum*, given its prominence and lethality in sub-Saharan Africa as recognized by the WHO [1], promising strides have been made in diversifying the approaches.

Another promising study is [11], where the authors suggest an automated analytic method for employing quantitative phase pictures to identify *Plasmodium falciparum*-infected red blood cells at the trophozoite or schizont stage. Linear discriminant classification (LDC), logistic regression (LR), and K-nearest neighbor

classification (NNC) are some of the techniques employed. Regarding schizont-stage detection, LDC has the best accuracy (up to 99.7%), whereas NNC has marginally greater accuracy (99.5%). These are all various studies contributing to automating malaria diagnosis.

This work explores Semantic segmentation, specifically the U-Net architecture, to segment malaria plasmodium species in digital image data collected in African settings, specifically Rwanda. Semantic segmentation groups image regions or pixels belonging to the same object class. It provides a helpful solution for various applications, including colon crypt segmentation, tumor identification, localization of surgical instruments, road sign detection, and land use classification. Semantic segmentation offers a well-defined method based on object class semantics, unlike non-semantic segmentation, which clusters pixels based on general object properties [12]. This differentiation highlights its adaptability and usefulness in various contexts, particularly medicine, which helps with disease diagnostics and organ segmentation [12]. U-Net is a famous semantic segmentation architecture proposed by Priyanshu et al. [13]. It is a customized convolutional neural network made specially for the segmentation of biomedical images. With skip connections, it has a recognizable “U” shape with an encoding path for feature extraction and a decoding path for up-sampling and accurate localization. U-Net has gained significant traction in biomedical image analysis, providing precise segmentation findings and inspiring the creation of comparable image segmentation designs across various applications [13].

The choice of employing the U-Net architecture is well justified. The U-Net architecture’s proven success in biomedical image analysis, particularly in segmenting Plasmodium, makes it a suitable and promising approach to address the challenges in this domain. The U-Net architecture has become more well-known in the biomedical industry because of the requirement for precise segmentation. U-Net’s capacity to concurrently integrate low-level and high-level information makes it suitable for medical image segmentation. High-level information makes it possible to extract intricate patterns, whereas low-level information enhances accuracy [14]. Consider the work by Abraham [15] which applies U-Net to segment Plasmodium within thin blood smear images and showcases U-Net’s remarkable accuracy in this task. This study examines three loss functions—mean-squared error, binary cross-entropy, and Huber loss—revealing that the Huber loss function outperforms the others. Testing metrics for F1 score, positive predictive value (PPV), sensitivity (SE), and relative segmentation accuracy (RSA) are notably higher with the Huber loss function, measuring 0.9297, 0.9715, 0.8957, and 0.9096, respectively. This underscores the effectiveness of U-Net coupled with the Huber loss function in achieving precise Plasmodium segmentation in thin blood smear images.

Furthermore, U-Net’s adaptability to different color spaces and its consistently high accuracy rates across RGB, HSV, and GGB color spaces make it a compelling choice for malaria parasite segmentation as demonstrated in [16]. In the RGB, HSV, and GGB color spaces, respectively, the findings demonstrate astounding accuracy rates of 99.40%, 99.36%, and 99.47%, highlighting the suggested technique’s durability [16].

Despite the effectiveness of U-Net, a predominant gap lies in the dataset's composition, consisting primarily of thin images of malaria parasites. This limitation could hinder the generalizability of developed models to more diverse and complex scenarios using thicker blood smears or different image types. Also, there is a dearth of studies addressing the diverse species of malaria parasites. The prevailing focus on a single species or the lack of species-specific identification in the literature could undermine the models' effectiveness in handling multiple species scenarios. Addressing these gaps will not only enhance the comprehensiveness of the research but also lead to more robust and accurate solutions in the field of malaria parasite detection and segmentation.

The rest of this paper is structured as follows: The detection includes the description of the extensive datasets acquisition and the U-Net architecture covered in Sect. 2; the results and discussion are covered in Sect. 3 followed by the limitations and recommendations in Sect. 4, and the conclusions are covered in Sect. 5.

2 The Detection Method

2.1 Datasets, Annotation, and Preprocessing

The dataset consists of a comprehensive collection of microscopic images that capture four types of malaria parasites: *Plasmodium falciparum*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium vivax*. These images were meticulously collected at the Rwanda Biomedical Centre (RBC) [17] using a specialized microscope setup. The setup involved Giemsa-stained slides examined under a microscope, equipped with a camera attached to the eyepiece and connected to a laptop as shown in Fig. 1 [9].

As the microscope was adjusted, snapshots (fields) of the slides were taken and stored for further analysis. The images captured thick and thin film smears, providing a diverse range of samples for study. Each image was carefully annotated using the VGG Image Annotator 2.0.12 to ensure accuracy in identifying infected areas. This tool allowed for precise delineation of infected areas using its polygon feature.

The dataset was strategically divided to facilitate practical training, validation, and testing of the proposed U-Net-based segmentation model on various *Plasmodium* infections. An 80% allocation was made for training, with the remaining 20% equally divided between validation and testing for each parasite species. This division ensures a comprehensive representation of each parasite species across different subsets. This dataset provides a robust foundation for studying malaria parasites and developing effective machine learning models for their identification and classification. The detailed split and total number of images is found in Table 1.

The annotated masks were resized to a uniform dimension of 256×256 pixels and fed into the U-Net architecture. This standardization of image size not only ensures uniformity in training and results but boosts computational efficiency by minimizing



Fig. 1 Camera mounted on microscope at RBC to collect digital images of microscopic slides

Table 1 Summary of data split for all parasites

Parasite	Total	Train	Validation	Test
PF	58	46	6	6
PM	139	111	14	14
PO	191	152	19	20
PV	114	91	11	12

Note Abbreviations used: *PF* Plasmodium falciparum, *PO* Plasmodium ovale, *PV* Plasmodium vivax, *PM* Plasmodium malariae

the time and memory required. This preprocessing step guarantees that the mask sizes are consistent for the segmentation model, enabling smooth integration of the annotated data into the training, validation, and testing stages.

The samples used in this study were collected from patients with high fever who sought consultation at healthcare facilities. Positive slides were subsequently collected by the Rwandan Biomedical Centre for further analysis, including our research. Sample collection is primarily conducted at healthcare facilities across Rwanda on a quarterly basis for quality control purposes. RBC holds the legal authority for quality control, research, and training within Rwanda. This justifies the ethics of our study.

2.2 *U-Net Architecture*

The U-Net is a convolutional neural network architecture designed specifically for biomedical image segmentation tasks. It was introduced in a paper titled “U-Net: Convolutional Networks for Biomedical Image Segmentation,” authored by Ronneberger et al. [13]. The architecture is notable for its distinctive U-shaped design, which includes a contracting path (down-sampling) and an expansive path (up-sampling).

- **Contracting Path (Down-sampling):** The contracting path is responsible for reducing the spatial dimensions of the input image while capturing hierarchical features. It achieves this through a series of convolutional layers designed to detect patterns and features at different scales. The contracting path involves a series of convolutional and max-pooling layers that gradually reduce the spatial dimensions of the input image while extracting hierarchical features.
- **Expansive path (Up-sampling):** On the other hand, the expansive path is responsible for recovering the spatial resolution of the segmented regions. It uses transposed convolutions (deconvolutions) to up-sample the feature maps obtained from the contracting path. This helps the network recreate the finer details of the segmented objects. Additionally, during the up-sampling process, the expansive path incorporates information from the contracting path. This is achieved through concatenation operations, where feature maps from the contracting path are combined with those from the expansive path. This information fusion ensures that the network has context and spatial information for accurate segmentation.

The beauty of the U-Net architecture lies in the synergy between these two paths. The contracting path learns to extract meaningful features and patterns from the input image, while the expansive path uses this information to generate precise segmentation masks. This U-shaped design allows the U-Net to excel at biomedical image segmentation tasks where accuracy and detail preservation are crucial, making it a popular choice for tasks like cell nucleus segmentation or detecting intricate structures within medical images. Inspired by [18, 19], the following U-Net approach in Fig. 2 is used for this study.

2.3 *Training*

This work conducts individual binary segmentation on all four parasite types, with the architectural focus on precisely segmenting objects of interest from images while adeptly addressing the unique challenges inherent in binary image segmentation tasks. The model’s training leverages the Adam optimizer with a learning rate set to $1e-4$ and employs binary cross-entropy loss. Alongside accuracy, the model’s efficacy is gauged through the mean Intersection over Union (IoU) metric, quantifying the overlap between the predicted and actual masks. The training phase encompasses model fitting to the training dataset using a batch size of 16 for 300 epochs.

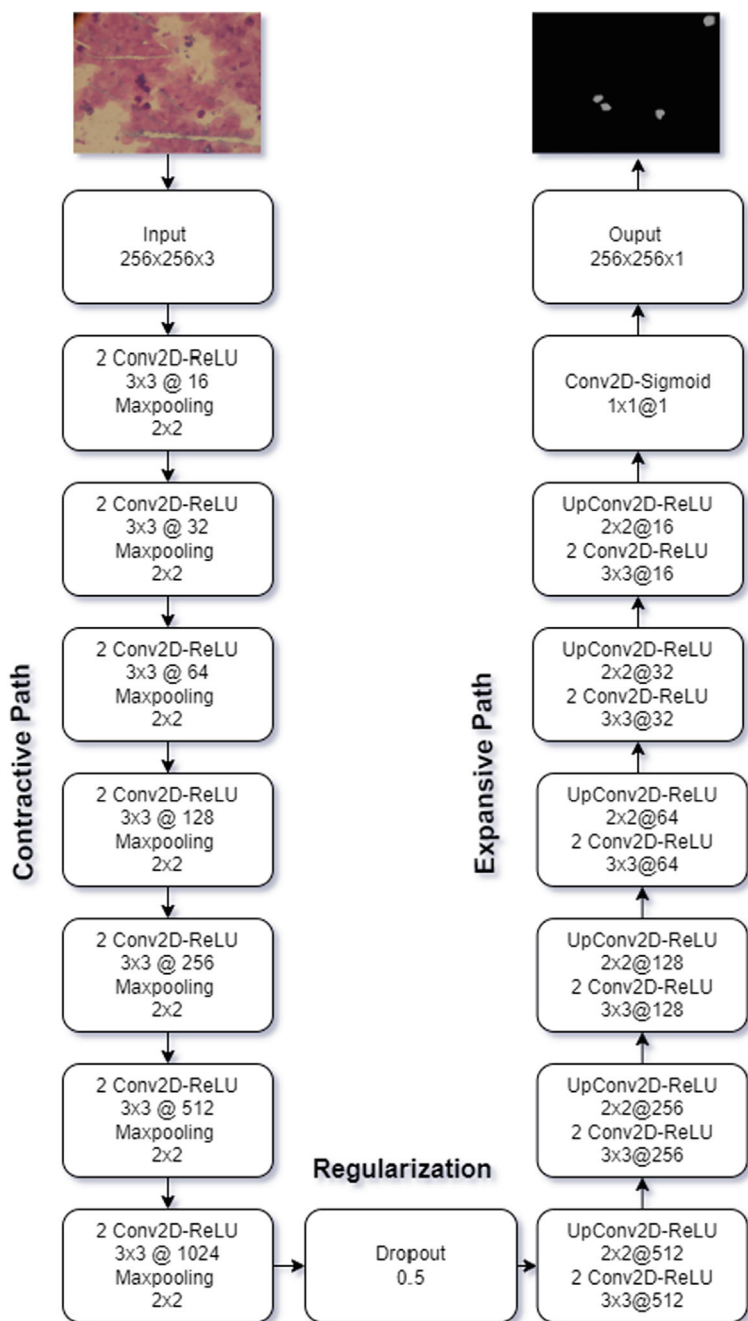


Fig. 2 U-Net architecture

- **Learning Rate (1e-2):** The learning rate controls how much to update the weight in the optimization algorithm. A smaller learning rate could lead to converging to the global minimum but might take more time to train, while a larger learning rate could speed up the training but risks overshooting the global minimum. In this case, a learning rate 1e-2 is chosen to balance convergence speed and assurance of not missing the global minimum [20].
- **Batch Size (16):** The batch size determines the number of samples propagated through the network simultaneously. A smaller batch size is chosen because it requires less memory to process, allows the model to start learning from data earlier, and provides a regular weight update, which can result in a robust model. The data size also accounts for why this figure was chosen. However, it might make the training process noisier and longer [20].
- **Number of Epochs (300):** The number of epochs is the number of times the entire dataset is passed forward and backward through the neural network. A higher number of epochs could lead to better performance until a certain point, after which the model might start overfitting. Therefore, 300 epochs are chosen to allow the model to learn complex patterns in the data without overfitting [21].

The model uses binary cross-entropy loss and Adam optimizer, which combines the advantages of two other extensions of stochastic gradient descent: AdaGrad and RMSProp. The mean Intersection over Union (IoU) metric is used for evaluating segmentation models by quantifying the overlap between the predicted and actual masks. Finally, precision, F_1 score, and recall metrics are explored via hyperparameter tuning to identify an optimal threshold yielding the highest F_1 score. The F_1 score is a performance metric commonly used in binary classification problems. The accuracy and comprehensiveness of the model are assessed using the harmonic mean of recall and precision. The F_1 score runs from 0 to 1, with 1 being the highest number that can be achieved [22, 23]. Unlike accuracy, it offers reliable findings for both balanced and unbalanced datasets.

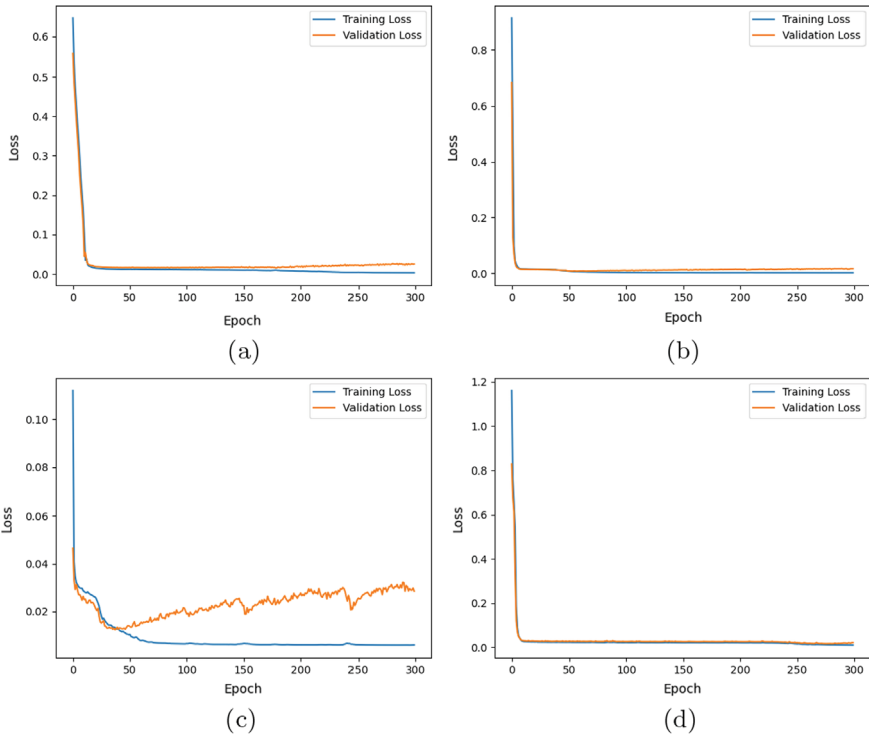
3 Results and Discussion

Table 2 presents a concise overview of the outcomes obtained during the testing phases for all four parasite types. Figure 3 shows the training and validation loss of the model on the different *Plasmodium* species. Additionally, the ensuing figures in Fig. 4 offer visual representations of the model results on the four parasites, providing valuable insights into the project's performance.

The model achieved remarkable accuracy and low losses, yet there is potential for improvement when considering other evaluation metrics. Accuracy is a valuable metric when the classes are approximately equally distributed. It measures the proportion of correct predictions (true positives and negatives) among the total number of cases examined [24]. The F_1 score, combining precision and recall, provides a holistic view of the model's performance. *Plasmodium ovale* displays the highest

Table 2 Results of metrics on four parasites

Parasite	Loss	Accuracy	F_1 score	Precision	Recall
PF	0.0186	0.9955	0.0868	0.1175	0.0688
PM	0.0152	0.9969	0.5420	0.5282	0.5565
PO	0.0298	0.9882	0.6020	0.5444	0.6731
PV	0.0278	0.9840	0.5936	0.5365	0.6642

**Fig. 3** Training result: **a** model on Plasmodium falciparum **b** model on Plasmodium malariae **c** model on Plasmodium ovale **d** model on Plasmodium vivax

F_1 score and recall among parasites, closely trailed by Plasmodium vivax and then Plasmodium malariae. In contrast, Plasmodium falciparum exhibits poorer performance, attributed to limited training data. Data augmentation was omitted to assess the model's inherent learning. F_1 score, precision, and recall collectively showcase the segmentation's efficacy. Plasmodium malariae exhibits peak accuracy, followed closely by falciparum, ovale, and vivax. F_1 scores align with training data, yielding results under 0.7. Visuals highlight accurate data detection, yet recall and precision at the optimal threshold are lacking.

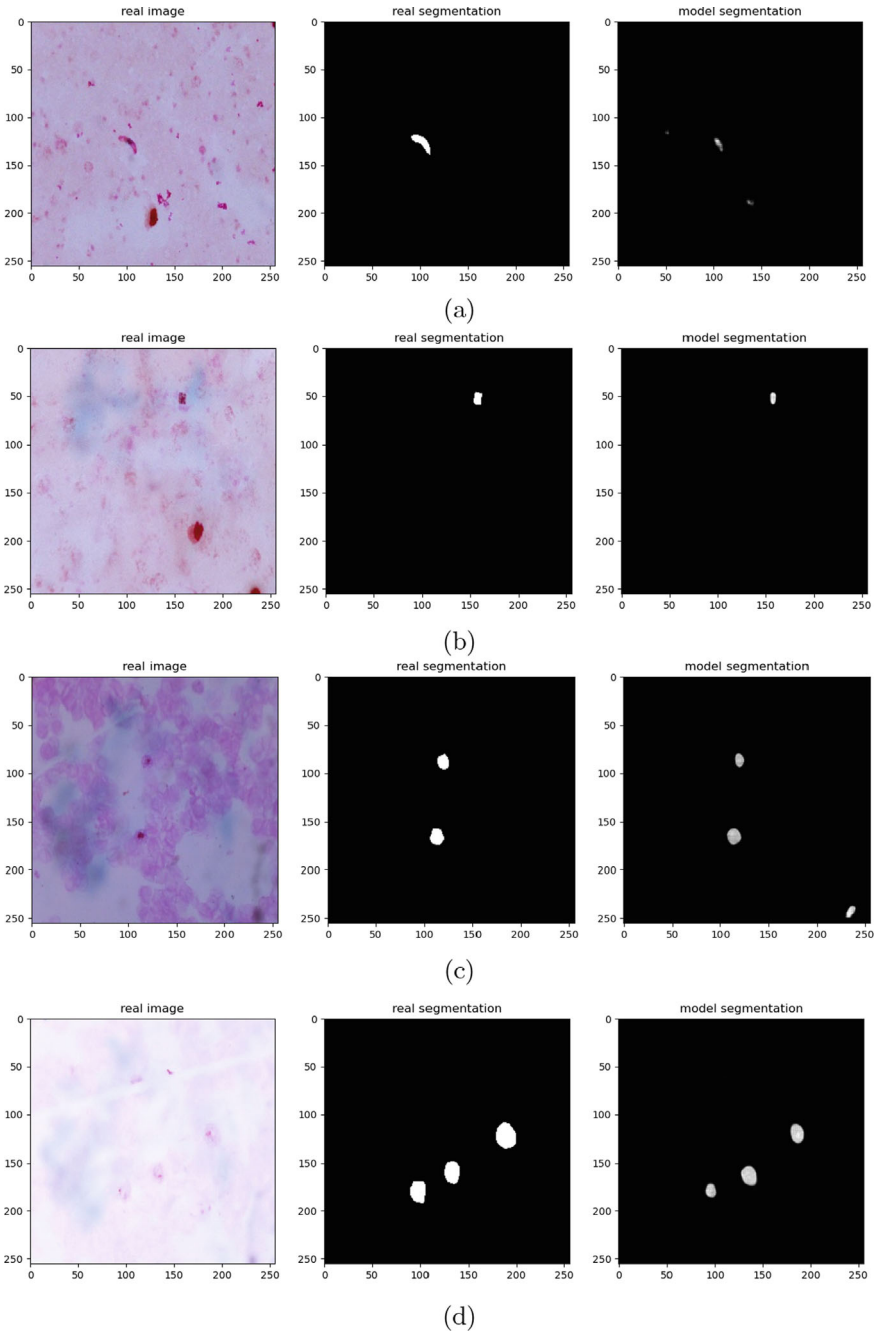


Fig. 4 Real image, ground truth, and model segmentation on the four parasites: **a** *Plasmodium falciparum* **b** *Plasmodium malariae* **c** *Plasmodium ovale* **d** *Plasmodium vivax*

Generally, the model accurately detects 99.6% of *Plasmodium falciparum*, 99.7% of *Plasmodium malariae*, 98.8% of *Plasmodium ovale*, and 98.4% of *Plasmodium vivax*. Despite strong accuracies, they might not fully depict effectiveness. The discrepancies between accuracies and other metrics reveal nuanced evaluation. In the study, while *Plasmodium malariae* exhibits peak accuracy, the F_1 -scores for all parasites are under 0.7. This discrepancy between accuracy and F_1 -score suggests that while the model is generally good at identifying the presence or absence of parasites (high accuracy), it may be less effective at correctly identifying positive cases (lower F_1 -score), particularly for classes with fewer instances in the training data [24].

4 Limitations and Recommendations

Effective diagnosis and treatment of malaria need efficient and precise segmentation of the parasites. As we explore the complexities of automatic segmentation using the U-Net architecture, a few crucial factors come into focus that may significantly impact the model's effectiveness and dependability. Three key limitations stand out and present opportunities for improvement.

One of the primary limitations of the current approach is the lack of a comprehensive dataset that encompasses the wide range of parasite variants, particularly for the *P. falciparum* parasite [25]. The model's effectiveness hinges on its ability to generalize across subtle variations. A larger dataset encompassing diverse parasite stages, morphologies, and image qualities can facilitate more effective learning. This enables the model to better address the complex challenges posed by various parasite species, fostering a comprehensive and adaptable solution. Data augmentation techniques can be employed to artificially expand the size and diversity of the training data, particularly when acquiring new data is challenging or expensive. This may be especially advantageous for the model's detection of *P. falciparum* parasites, which currently perform poorly due to a lack of training data [26].

In addition, the success of any segmentation model is inherently linked to the quality of its ground truth masks [27]. Investigating cutting-edge computer vision methods for producing these masks is essential to improving the precision of the F_1 -score and overall performance. Techniques such as instance segmentation, active contour modeling, and combining annotations from multiple perspectives should be utilized to achieve this [16]. Employing image processing techniques to eliminate or mitigate the impact of artifacts and white blood cells can further enhance the quality of ground truth annotations. Ground truth masks can be meticulously crafted by leveraging the strengths of these methods, effectively bridging the gap between manual annotation and automated detection. The outcome is a more streamlined and reliable training procedure that bolsters segmentation accuracy.

Furthermore, segmentation accuracy extends beyond data collection and model design. The choice of metrics and loss functions employed for evaluating model performance naturally influences the learning process and assessment accuracy. Exploring alternative solutions holds promise [28]. Investigating loss functions tailored to

the specifics of binary segmentation, such as Huber loss or dice loss, may lead to improved convergence and segmentation quality [15]. Additionally, incorporating metrics like the Matthews correlation coefficient (MCC) or the Jaccard index into the evaluation repertoire enables a more comprehensive assessment that considers various aspects of model performance.

Finally, one of the challenges faced is the inadequate training of laboratory technicians, resulting in poorly prepared malaria blood films. These substandard samples are often rejected, leading to delays in acquiring suitable specimens. Additionally, the inconsistent quality of stainers (reagents) can produce images with varying colorations. Experts must regularly check and validate reagents at endemic sites, further hindering the sample collection process.

By addressing the limitations identified and implementing the recommended improvements, the effectiveness and reliability of automatic segmentation of malaria parasites using the U-Net architecture can be significantly enhanced. This can lead to more accurate diagnoses and timely treatment decisions, improving patient outcomes and contributing to the fight against malaria.

5 Conclusion

The study reported accuracy levels of 99.6% for *Plasmodium falciparum*, 99.7% for *Plasmodium malariae*, 98.8% for *Plasmodium ovale*, and 98.4% for *Plasmodium vivax*, indicating a significant improvement in the accuracy of malaria diagnosis. Despite these impressive figures, there is potential for further improvement when considering other evaluation metrics such as the F_1 score, precision, and recall. The performance for *Plasmodium falciparum* was lower due to insufficient training data, highlighting the need for a comprehensive dataset for effective learning. The quality of ground truth masks and the choice of metrics and loss functions also significantly influence the model's success.

To extend this work, more data could be collected, advanced computer vision methods could be explored for each parasite, and other methods of segmentation such as MaskRCNN could be significantly explored. By implementing these steps, it is hoped that the model's performance can be enhanced, leading to more efficient and precise diagnosis and treatment of malaria. This study substantially contributes to malaria parasite detection and segmentation and identifies critical areas for future improvement. There are plans to implement automated malaria diagnosis in collaboration with the Rwandan Biomedical Centre (RBC), our most important stakeholder. This constitutes a part of the investigative phase, with the hope of transitioning to the deployment and testing phase with enhanced outcomes.

References

1. Commonwealth leaders take action in response to the Kigali Summit's call for bold commitments towards ending Malaria and Neglected Tropical Diseases (NTDs) | RBM Partnership to End Malaria. <https://endmalaria.org/news/commonwealth-leaders-take-action-response-kigali-summit%E2%80%99s-call-bold-commitments-towards-ending>
2. Shewajo FA, Fante KA (2023) Tile-based microscopic image processing for malaria screening using a deep learning approach. *BMC Med Imaging* 23(1):39
3. Fact sheet about malaria. <https://www.who.int/news-room/fact-sheets/detail/malaria>
4. Iqbal J, Hira P, Al-Ali F, Khalid N, Sher A (2003) Modified Giemsa staining for rapid diagnosis of Malaria infection. *Med Principles Pract* 12(3):156–159
5. Krishnadas P, Chadaga K, Sampathila N, Rao S, Prabhu S (2022) Classification of Malaria using object detection models. *Informatics* 9(4):76. <https://doi.org/10.1155/2022/3626726>, <https://www.mdpi.com/2227-9709/9/4/76>. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute
6. Shambhu S, Koundal D, Das P, Hoang VT, Tran-Trung K, Turabieh H (2022) Computational methods for automated analysis of Malaria parasite using blood smear images: recent advances. *Comput Intell Neurosci* 2022:3626,726. <https://doi.org/10.1155/2022/3626726>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9017520/>
7. Kigozi RN, Bwanika J, Goodwin E, Thomas P, Bukoma P, Nabyonga P, Isabirye F, Oboth P, Kyoziira C, Niang M, Belay K, Sebikaara G, Tibenderana JK, Gudozi SS (2021) Determinants of malaria testing at health facilities: the case of Uganda. *Malaria J* 20(1):456
8. Mukamakuza CP, Tuyishimire E, Mbituyumuremyi A, Brown TX, Iradukunda D, Phuti O, Happiness RM (2022) A dependable digital system model for Malaria monitoring. preprint, *Mathematics and Computer Science*. <https://doi.org/10.20944/preprints202207.0461.v1>
9. Mary HR, Mukamakuza CP, Tuyishimire E (2023) A data management model for Malaria control: a case of Rwanda. In: 2023 IEEE AFRICON, pp 1–6. <https://doi.org/10.1109/AFRICON55910.2023.10293671>. ISSN: 2153-0033
10. Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G (2018) Image analysis and machine learning for detecting malaria. *Transl Res* 194:36–55. <https://doi.org/10.1016/j.trsl.2017.12.004>. <https://www.sciencedirect.com/science/article/pii/S193152441730333X>
11. Park HS, Rinehart MT, Walzer KA, Chi JTA, Wax A (2016) Automated detection of *P. falciparum* using machine learning algorithms with quantitative phase images of unstained cells. *PLoS ONE* 11(9):e0163,045. <https://doi.org/10.1371/journal.pone.0163045>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5026369/>
12. Thoma M (2016) A survey of semantic segmentation. <https://arxiv.org/abs/1602.06541>. Publisher: arXiv Version Number: 2
13. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. <https://doi.org/10.48550/ARXIV.1505.04597>. Publisher: arXiv Version Number: 1
14. Liu X, Song L, Liu S, Zhang Y (2021) A review of deep-learning-based medical image segmentation methods. *Sustainability* 13(3):1224
15. Abraham JB (2019) Malaria parasite segmentation using U-Net: Comparative study of loss functions. *Commun Sci Technol* 4(2):57–62. <https://doi.org/10.21924/cst.4.2.2019.128>. <https://cst.kipmi.or.id/journal/article/view/128>
16. Nautre A, Nugroho HA, Frannita EL, Nurfauzi R (2020) Detection of Malaria Parasites in thin red blood smear using a segmentation approach with U-Net. In: 2020 3rd International conference on biomedical engineering (IBIOMED), pp 55–59. <https://doi.org/10.1109/IBIOMED50285.2020.9487603>
17. Welcome to RBC. <https://www.rbc.gov.rw/index.php?id=188>
18. Finalproject. <https://kaggle.com/code/alirezatohidi226/finalproject>
19. View of Malaria parasite segmentation using U-Net: comparative study of loss functions. <https://cst.kipmi.or.id/journal/article/view/128/60>

20. Priyanshu A, Naidu R, Miresghallah F, Malekzadeh M (2021) Efficient hyperparameter optimization for differentially private deep learning. <http://arxiv.org/abs/2108.03888>, [ArXiv:2108.03888](https://arxiv.org/abs/2108.03888) [cs]
21. Prechelt L (2012) Early stopping—but when? In: Montavon G, Orr GB, Müller KR (eds) Neural networks: tricks of the trade: second edition, lecture notes in computer science. Springer, Berlin, Heidelberg, pp. 53–67. https://doi.org/10.1007/978-3-642-35289-8_5
22. Hand DJ, Christen P, Kirielle NF (2020) An interpretable transformation of the F-measure. <https://doi.org/10.48550/ARXIV.2008.00103>. Publisher: arXiv Version Number: 3
23. Lipton ZC, Elkan C, Narayanaswamy B (2014) Thresholding classifiers to maximize F1 score. <https://doi.org/10.48550/ARXIV.1402.1892>. Publisher: arXiv Version Number: 2
24. Grandini M, Bagli E, Visani G (2020) Metrics for multi-class classification: an overview. <http://arxiv.org/abs/2008.05756>. [ArXiv:2008.05756](https://arxiv.org/abs/2008.05756) [cs, stat]
25. Su Z, Li W, Ma Z, Gao R (2022) An improved U-Net method for the semantic segmentation of remote sensing images. *Appl Intell* 52(3):3276–3288
26. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for Deep Learning. *J Big Data* 6(1):60
27. Malladi SRSP, Ram S, Rodriguez JJ (2018) A ground-truth fusion method for image segmentation evaluation. In: 2018 IEEE southwest symposium on image analysis and interpretation (SSIAI). IEEE, Las Vegas, NV, pp 137–140. <https://doi.org/10.1109/SSIAI.2018.8470317>
28. Wang Q, Ma Y, Zhao K, Tian Y (2022) A comprehensive survey of loss functions in machine learning. *Annals Data Sci* 9(2):187–212

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



PLC-Based Traffic Light Control for Flexible Testing of Automated Mobility



Tamás Wágner, Tamás Tettamanti, Balázs Varga, and István Varga

Abstract Road infrastructure has evolved to meet diverse needs over time. This makes modernizing traffic control systems also necessary in order to keep pace with technological innovations and changing user habits. This paper presents a novel design and implementation strategy for the next generation of traffic control systems, with particular emphasis on the challenges posed by the integration of automated vehicles. General flexibility in future traffic management, proper use of standardized communication protocols, and safety solutions are essential factors in addressing the traffic challenges of the coming decades. The system has been developed primarily for PLCs (Programmable Logic Controllers) and has been tested and verified, providing the advantages of safety elements and software safety solutions. The proposed system demonstrated efficient and reliable performance while the applied OPC UA (Open Platform Communications Unified Architecture) communication protocol assured robustness and secure client-to-PLC communication.

Keywords PLC · Traffic light controller · CCAM · Automated vehicles · OPC UA

1 Introduction

Essentially, traffic lights are control systems designed to manage road traffic flow and improve safety for all road users. The main progress in traffic light technology has not been in control methods, but in the Traffic Light Controller (TLC) hardware. These devices have become increasingly reliable and safe over time, as microcontroller-based systems have replaced older, relay-based ones.

However, there have been fewer advancements in control software compared to the TLCs. Presently, systems that control traffic lights can be categorized into two

WWW home page: www.traffic.bme.hu.

T. Wágner (✉) · T. Tettamanti · B. Varga · I. Varga
Department of Control for Transportation and Vehicle Systems, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Budapest H-1111, Muegyetem rkp. 3, Hungary
e-mail: wagnertomi991025@edu.bme.hu

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_13

177

main groups: fixed-time (open-loop) systems and adaptive (closed-loop) systems. The former is unable to react to external information such as traffic flow, whereas the latter is equipped with sensors that gather information about its surroundings and adjust the phase plan accordingly. It is worth noting that many adaptive systems can only provide estimates of the number of vehicles and traffic throughput at an intersection for the local area [1, 2].

The emergence of automated cars has made it necessary to create a new type of traffic light control system, primarily for testing purposes. These new automated systems need to be more interconnected than traditional traffic management systems, as there are more requirements for proving grounds. Although current traffic control systems have proven to be safe and robust, they cannot cope with the complexity of modern traffic systems. These traditional systems are primarily used to test driver assistance systems that do not require continuous information exchange with other vehicles and infrastructure [3, 4].

Our proposed system has the capacity to handle traffic scenarios that cannot be implemented in conventional traffic management systems, such as conflicting green signals. The main aim is to present corner cases to the automated vehicles to test decision making layers under unusual circumstances. During the design stage of development, it is important to consider external V2X (Vehicle to Everything) communication system integration, as this will be essential for automated vehicle testing in the future [5]. Additionally, we aim to prioritize the possible implementation of existing traffic measurement devices, such as inductive loops, magnetic sensors, or cameras at a later stage [6–8]. Given the innovative nature of the system, it was deemed essential to integrate proven safety protocols, such as the Intergreen Time Matrix, to ensure that the system guarantees maximum safety for road users [9].

This system has been designed with the testing interests of the automotive industry in mind, making it highly effective when used in automotive proving grounds. ZalaZONE, based in Zalaegerszeg, Hungary, is one such proving ground where it can be used to create a test environment that is ideal for connected and automated vehicle (CAV) testing. However, it is noteworthy that ZalaZONE can offer services not only for commercial use, but can also significantly support various research and development projects. Scientific and academic institutions can utilize the facility for their research projects as well [10].

This paper examines the potential of PLCs as control devices for traffic signal systems, particularly with regard to their feasibility and reliability. A special focus is put on measuring the latency of the network communication between the different elements of the system as well as on how much processor capacity was used by controlling the intersection.

The structure of the paper is as follows: Sect. 2 is a presentation of the system architecture, while Sect. 3 is a description of the test methods and results.

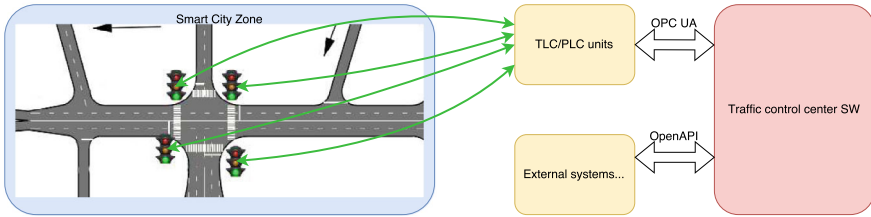


Fig. 1 System elements and their communication protocols

2 System Architecture

The conceptual system is divided into three subsystems: local programs running on Programmable Logic Controllers (PLCs), central server software, and client applications. The PLC software is the only component capable of autonomous operation independently of the rest of the system, but the opposite is not true. This architecture enables safe fallback to a traditional, island mode operation in case of failure of the central system. The central software is responsible for coordinating all of the subcomponents, ensuring continuous communication with both the various PLCs and client applications. The network links between the system components are shown in Fig. 1.

A high level of safety and real-time control is essential for each system component. The selection of implementation tools considered these requirements. The central software’s code is written in Rust, a modern programming language focused on safety. This enabled us to integrate C libraries while maintaining high memory safety and excellent performance [11]. The central subsystem does not control the traffic lights directly but only sends instructions to the PLCs, which process these instructions and control the traffic lights via their own PLC programs. This structural setup ensures that the traffic lights operate reliably and securely, even if the communication between the central system and the PLC is interrupted.

It’s critical to note that reliable communication between the PLCs and the central software is crucial in our system. With this in mind, we’ve chosen to use the Open Platform Communications Unified Architecture (OPC UA) protocol—a network protocol that is specifically designed for industrial environments and ensures robustness, security, and scalability [12]. It’s worth noting that most state-of-the-art PLCs support this protocol by default. Moreover, numerous modern PLCs support this protocol out-of-the-box, which does not overload the PLC’s processors due to manufacturer optimizations.

Three modes of operation are available for PLCs, with the first referred to as “static,” whereby the traffic lights operate according to a predefined phase plan without any modifications. In this mode, the PLC only sends status reports to the server and doesn’t receive any instructions from the central software. In the “custom” mode, the central software uploads a phase plan created by the user to the PLC. Along with this, the user has the option to deactivate the built-in safety functions. The third mode, the “interactive” mode, enables the user to control the traffic lights in real-time

through a client application of their choice. In this scenario, there is no predefined phase plan. PLCs are autonomous entities and are not required to operate in the same mode. Traffic light control systems are crucial because malfunctions can put human lives at risk. To ensure that the necessary safety protocols are in place, the design and implementation of these systems must be approached with great care. The PLC was chosen as the hardware because of its built-in basic safety components, including redundant subsystems, real-time clock, output monitoring functions, and so on. In addition, the PLC includes not only hardware but also software safety features, such as data integrity control mechanisms or watchdog timers.

Despite these built-in safety modules, PLCs cannot provide complete protection, so additional safety software must be developed. These programs operate in parallel with the light control program and stop it immediately if an anomaly is detected. The two most crucial extra programs are the intergreen time matrix checker and the output cable status checker.

The former verifies that there are no conflicting green phases and that there is enough time between phases to clear the intersection. If an error occurs, the PLC responds based on the current operating mode. A warning is sent to the central server in the event of an error. If the PLC is in “static” mode, it will attempt to restart the control when a problem occurs. If a series of restarts fails to solve the problem, the PLC reports the problem to the central server.

The output cable status checker monitors the status of the cables based on feedback from the PLC I/O cards. If the program detects a cable break, it indicates that maintenance is required to the central server and turns off the traffic lights until the necessary repairs are made to ensure the safety of the road.

3 Evaluation and Testing

To ensure the proper and reliable functionality of our systems, comprehensive testing is essential. This chapter presents the testing process and the technologies used.

Throughout the development process, multiple testing levels were utilized, one of the most significant and comprehensive being hardware-in-the-loop (HiL) testing. This form of testing allowed us to test our software directly with real hardware under realistic conditions. During this phase, we thoroughly tested the functionality of both the PLC programs and the entire system. This included the OPC UA communication protocol.

To replicate actual environmental conditions, we developed a customized circuit capable of emulating the behavior of signal lights and capturing logic values at the PLC output through a microcontroller. Additionally, the circuit is capable of simulating cable faults, which was achieved by using the Sziklai pair circuit, see Fig. 2. We used a Raspberry Pi 4 (RPI) and two Arduino Nano microcontrollers to conduct our measurements. The link between the modules is shown in Fig. 3. The RPI was responsible for controlling the testing process, while the Arduino microcontrollers managed the simulation of cable breaks and the reading of PLC output values.

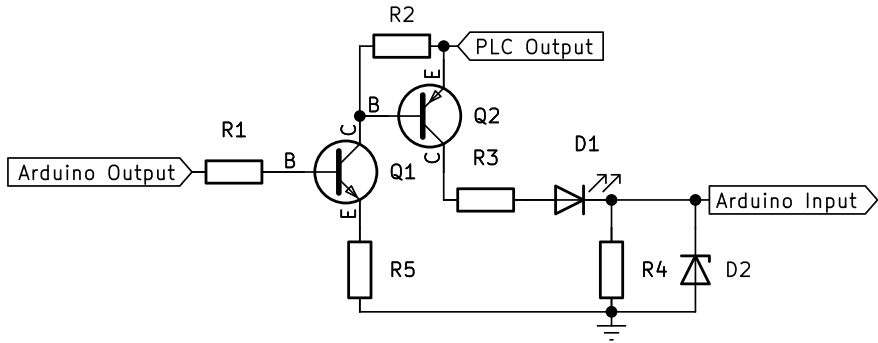


Fig. 2 Measurement circuit for one PLC output

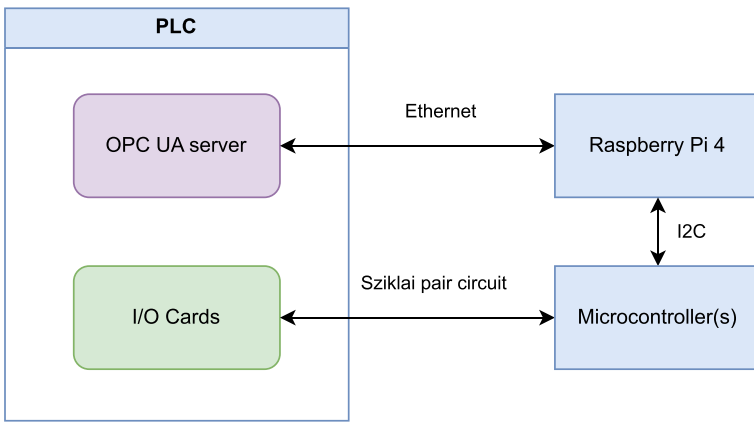


Fig. 3 Modules in the HiL test

Two versions of the HiL test control program are executed by RPi4. One version is written in C because the open62541 [13] library utilized to implement OPC UA communication is also written in C. Additionally, the RPi library primarily supporting I2C communication was also written in C.

The other version, a Rust version of the HiL test control program, was also created later in the development process. This version also uses the open62541 library, but its usage is primarily handled by the Rust compiler and linker. The advantage of this version is that the central system is completely written in Rust. This allows to test the suitability of the open62541 library for our requirements in the traffic control center software in advance and thus to determine whether a more advanced library should be used at a later stage.

Our primary testing objectives focused on validating the performance of the PLC processor, including its capacity for handling the concurrent execution of internal programs and OPC UA communication. We also evaluated the system’s ability to operate in real-time and quantified the network traffic it generated.

Table 1 OPC UA communication measurements

	C	Rust
Average network latency	111.52 ms	193.14 ms
Maximum network latency	187 ms	301 ms
Average network traffic	650 byte/s	768 byte/s

3.1 Network Wide Testing

The primary goal of this testing phase was to measure the end-to-end latency of the communication between the PLC and the client. This latency is described as the duration between the change in the PLC output signal until the client detects the change. Accuracy in recording these times is achieved by synchronizing all devices' clocks in the system to the nearest thousandth of a second utilizing NTP (Network Time Protocol). The PLC's built-in logging function was used to record all output signal changes. Each change was time-stamped. Similarly, on the client side, all incoming messages were logged, including their time stamp and content upon arrival.

Wireshark was used for detailed monitoring of the network traffic. Wireshark is software that captures and organizes network traffic for analysis [14]. This data enabled us to identify the number and type of data packets transmitted between the client and the PLC.

Additionally, a significant objective of the test was to determine the amount of data transferred between network nodes and whether the system could operate smoothly under this load. In the test configuration, a single PLC communicated with a sole client, rendering a simulation of the actual operating conditions accurate as only one PLC would be present per intersection.

The results of the test for C and Rust versions are shown in Table 1. The average network latency for C measures 111.52 ms and for Rust measures 193.14 ms. It is worth noting that this value has the potential to decrease further by using specialized industrial equipment instead of generic devices and optical solutions instead of Ethernet-based network infrastructure. As our current implementation is mainly intended for data monitoring purposes, the latency results for the C version are entirely satisfactory, however, for the Rust version it is unacceptable. The slight increase in latency in this case is expected, as while the Rust programming language's performance is nearly on par with C, it still falls slightly short [15].

3.2 Processor Load Testing

Overloading the processor of the PLC can lead to critical system failures. When system processes like OPC UA communications overload the processor, there is a possibility that the task manager responsible for the PLC programs will not be able

Table 2 CPU load measurements

Average CPU load	26 %
Maximum CPU load	34 %
Minimum CPU load	23 %

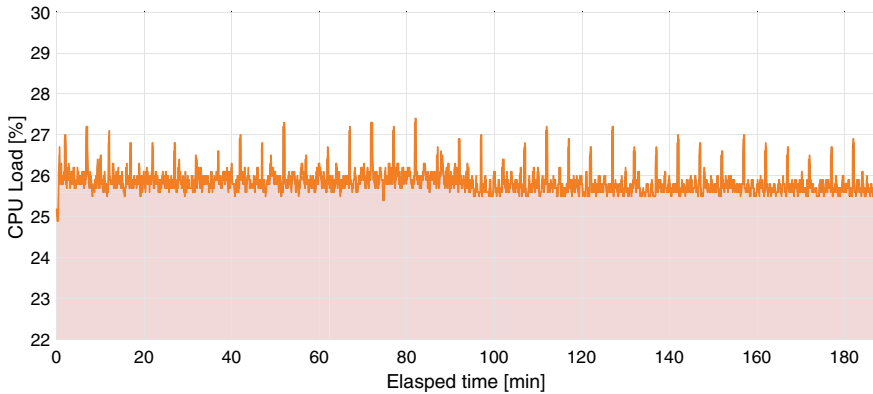


Fig. 4 CPU load over time with moving average (10 sample window)

to maintain the predetermined cycle times. In such overloads, various hazardous situations can occur. Therefore, it is essential that the processor load is kept low and that the system has sufficient computing capacity in reserve. To ensure adequate reserve capacity, we will analyze the minimum, maximum, and average CPU workload during the execution of control programs and OPC UA communication in HiL testing.

The tested PLCs, specifically the WAGO 750-8211/040 models, possess a high-performance processor capable of handling resource-intensive functions. Nevertheless, it is crucial to continuously monitor the processor load to prevent potential system risks and enable future integration of additional functions. Table 2 and Fig. 4 present the results of the CPU test. The data indicate that the system is not utilizing the entire CPU capacity available, and the CPU load varies in a consistent interval.

4 Conclusion

The paper presented the feasibility evaluation of a PLC-based traffic controller for a novel, dynamic traffic control system that can address current and future traffic challenges, especially those posed by the growing number of automated vehicles. The system architecture can seamlessly integrate with future V2X technologies to enable effective testing of automated vehicles.

Safety was our top priority in developing this system. The use of PLCs was advantageous due to their fundamental safety elements and the implemented additional software safety solutions. Throughout the testing phase, the system demonstrated efficient and reliable performance. The implementation of OPC UA communication protocol assured system stability, robustness, and secure client-to-PLC communication.

The performed network tests showed that the system works with acceptable latency in its current state and that this value could be further reduced with additional optimizations in the future. The tests on CPU load also indicated that the processor of the PLC is working under a constant and acceptable load.

Acknowledgements The research was supported by the European Union within the framework of the National Laboratory for Autonomous Systems (RRF-2.3.1-21-2022-00002). Supported by the ÚNKP-23-2-1-BME-250 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

References

1. Shankaran S, Rajendran L (2021) Real-time adaptive traffic control system for smart cities. In: 2021 International conference on computer communication and informatics (ICCCI). IEEE
2. Trivedi JD, Devi MS, Dave DH (2021) A vision-based real-time adaptive traffic light control system using vehicular density value and statistical block matching approach. *Transport Telecommun J* 22(1):87–97
3. Horváth MT, Lu Q, Tettamanti T, Török A, Szalay Z (2019) Vehicle-in-the-loop (VIL) and scenario-in-the-loop (SCIL) automotive simulation concepts from the perspectives of traffic simulation and traffic control. *Transport Telecommun J* 2(20):153–161
4. Jayan K, Muruganatham B (2020) Advanced driver assistance system technologies and its challenges toward the development of autonomous vehicle. In: *Intelligent computing and applications*. Springer, Singapore, pp 55–72
5. Wágner Tamás, Ormándi Tamás, Tettamanti Tamás, Varga István (2023) SPaT/MAP V2X communication between traffic light and vehicles and a realization with digital twin. *Comput Electr Eng* 106:108560 Mar
6. Anderson RL (1970) Electromagnetic loop vehicle detectors. *IEEE Trans Veh Technol* 19(1):23–30
7. Febrian Rachmadi M, Afif FA, Jatmiko W, Mursanto P, Manggala EA, Anwar Ma'sum M, Wibowo A (2011) Adaptive traffic signal control system using camera sensor and embedded system. In: *TENCON 2011 IEEE region 10 conference*, pp 1261–1265
8. Zarnescu A, Ungurelu R, Iordache AG, Secere M, Spoiala M (2017) Crossroad traffic monitoring using magnetic sensors. In: *2017 IEEE 23rd international symposium for design and technology in electronic packaging (SIITME)*, pp 413–418
9. HG Retko and Manfred Boltze (1987) Timing of intergreen periods at signalised intersection; The German method. *J Inst Transp Eng* 57(2):45–50
10. Szalay Zsolt (2021) Next generation X-in-the-loop validation methodology for automated vehicle systems. *IEEE Access* 9:35616–35632
11. Klabnik S, Nichols C (2023) *The Rust programming language*. No Starch Press
12. Mahnke W, Leitner SH, Damm M (2009) *OPC unified architecture*. Springer, Berlin. OCLC: ocn268784080

13. Julius Pfrommer I (2023) Open62541: open source implementation of OPC UA (IEC 62541). <https://github.com/open62541/open62541>
14. Beale J, Orebaugh A, Ramirez G (2006) Wireshark and ethereal network protocol analyzer toolkit. Elsevier
15. Bugden W, Alahmar A (2022) Rust: the programming language for safety and performance

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



User Perceptions of Progressive Web App Features: An Analytical Approach and a Systematic Literature Review



Tulio Marchetto and Marcelo Morandini

Abstract This article presents an in-depth analysis of the usability of progressive web apps (PWAs) across different operating systems, based on a survey that had 226 responses. The study covers five different apps and two operating systems, providing a comprehensive understanding of user experiences and preferences. Key findings include the relative ease of installation across all apps and operating systems, with iOS users reporting lower levels of difficulty. The visibility of app icons on home screens was generally high, underscoring the importance of app accessibility and user engagement. However, there is room for improvement in certain areas. The dataset obtained from the responses also reveals a high level of user engagement and willingness to provide feedback, which is crucial for the continual improvement of apps. The insights gleaned from this survey are invaluable for developers seeking to optimise their PWAs, ensuring they meet user expectations and deliver a superior app experience.

Keywords Progressive web apps · Usability · User experience · Installation difficulty · App visibility · User engagement · Feedback · App optimisation

1 Introduction

In the rapidly evolving digital landscape, progressive web apps (PWAs) have emerged as a game-changer, offering a seamless user experience across various operating systems. This report delves into a comprehensive analysis of a survey conducted to evaluate the usability of PWAs, providing valuable insights into user experiences and preferences. The survey encompasses a wide range of aspects, including installation

WWW home page: <https://www.usp.br>.

T. Marchetto (✉) · M. Morandini
Universidade de São Paulo, São Paulo, SP 03828-000, Brazil
e-mail: tuliomarchetto@usp.br

M. Morandini
e-mail: m.morandini@usp.br

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_14

difficulty, app visibility, functionality, and overall user experience, offering a holistic view of the current state of PWA usability.

The data collected from the survey paints a vivid picture of user experiences with PWAs. It reveals a spectrum of experiences, with mean scores indicating general trends in usability aspects, while standard deviations highlight areas of significant user opinion divergence. From installation experiences to app visibility and functionality, the survey uncovers critical facets of user interaction with PWAs, particularly on iOS and Android platforms.

In essence, this report serves as a mirror reflecting the current state of PWA usability, highlighting both its strengths and areas for improvement. It underscores the importance of user-friendly installation, effective app visibility, and the role of features like initialisation screens and text selection functionality in enhancing user experience. The insights gleaned from this survey are invaluable for developers seeking to optimise their PWAs, ensuring they meet user expectations and deliver a superior app experience.

In the realm of digital technology, progressive web apps (PWAs) have emerged as a significant player, offering a seamless user experience across various operating systems. This article delves into a comprehensive dataset derived from a survey conducted to evaluate the usability of PWAs. The dataset, comprising 226 responses and 46 columns, encapsulates a wide range of questions and response types, providing a holistic view of user experiences and preferences.

The survey covers responses about five different apps: Calculator, Instagram, Periodex, The Cube, and X (former Twitter), across two operating systems: Android and iOS. The responses shed light on various aspects of PWA usability, including installation difficulty, app visibility, and functionality. Notably, the dataset reveals that users generally found the installation process relatively easy across all apps and operating systems, with iOS users reporting lower levels of difficulty compared to Android users for the same apps.

The dataset also underscores the importance of app visibility and identification in enhancing user experience. Most users reported seeing the app icon and name on their home screens, which is crucial for app accessibility and user engagement. However, there are instances where this aspect could be improved. The responses also indicate a considerable level of user engagement and willingness to provide feedback, which is essential for the continual improvement of apps. This dataset serves as a valuable resource for developers seeking to optimise their PWAs, ensuring they meet user expectations and deliver a superior app experience.

2 Systematic Literature Review

We found some usability assessment method studies, but none on progressive web apps (PWAs) except this preliminary scoping review study. The study by Insfran and Fernandez [1] examined web usability evaluation methods. From 410 articles, 51

Table 1 Search strings for primary articles

Source	Search strings
ISIS web of science	TI = (usability AND evaluat* OR assess* OR measu* OR method* OR techni* OR approah*) AND AB = (pwa OR “progressive web app*” OR “web app*” OR “mobile app*”) AND LA = (English)
SCOPUS	TITLE (usability AND evaluat* OR assess* OR measu* OR method* OR techni* OR approah*) AND ABS (pwa OR “progressive web app*” OR “web app*” OR “mobile app*”)) AND (LIMIT-TO (LANGUAGE , “english”))

were reviewed. Web application usability evaluation methods were briefly discussed in this study.

Hornbaek [2] used a thematic review system-inspired research method to review usability measures’ current state. Usability study measures were examined to determine if they accurately measure and cover broad issues. This review found several problems with usability research. These include finding and comparing objective usability measures, making measures for learning and remembering, and looking at the relationships between usability measures to make sure they are correct.

In a systematic review by Mendes [3], only 5 web engineering research claims were found to be rigorous. This review also found incorrect terminology in several papers. They used “experiment” instead of “experience report” and “case study” instead of “proof of concept”. Improvements to Web Engineering practices were suggested.

2.1 Conducting the SRL

Kitchenham guidelines were used for this project’s scoping review. The scoping review involves three phases: planning, conducting, and reporting results, followed in order [4].

The first phase identified the need for the review and established a review protocol to reduce research bias. After the review protocol was approved, the review began. The review summarises impartial primary software usability studies. To reduce research bias, a review protocol is created through meetings among project-related professors Kitchenham [4].

The search strings were formed by alternative terms and synonyms using the boolean expression OR combining the main search terms using AND. Table 1 presents the general search terms used for identifying primary articles related to software usability with real-time data. Once the search terms were identified, significant source portals were selected. To search the primary studies, the search engines explored were:

- ISIS Web of Science
<https://www.webofscience.com/wos/woscc/advanced-search>
- SCOPUS
<https://www.scopus.com/search/form.uri?display=advanced>

2.2 *Study Selection Criteria*

In 1977, the first model of software usability was introduced. However, this work focused on studies from 2006 to 2021, as the rapid updates of GUIs in systems and the emergence of GUIs for portable touch screen-controlled devices make earlier studies very outdated. The relevant research areas of software usability studies were classified into five main categories. In addition, relevant primary studies are identified through obtaining full-text documents based on the inclusion and exclusion criteria mentioned below. This review considers empirical and analytical primary studies for inclusion in the list of primary studies references.

Thus, for further processing, a total of 245 unique studies were identified. Some studies are excluded because they do not provide answers to any of the five research questions and to the quality assessment questions in step 2d. Subsequently, 106 articles were selected based on the following inclusion-exclusion criteria:

Inclusion Criteria

1. Empirical and analytical studies for models and standards of software usability with real-time data.
2. Studies using usability evaluation methods.
3. Empirical studies contributing to usability metrics.
4. Studies contributing to the integration of usability engineering with software engineering.
5. Empirical studies for usability applications in other domains.

Exclusion Criteria

1. Studies in the specific context of a single software.
2. Studies of systems for a specific sector.
3. Studies without empirical analysis of usability evaluation methods.
4. Similar studies with the same results by the same authors but in different conferences and journals.
5. Studies written in a language other than English.

2.3 *Final Considerations of the Review*

This literature scoping review followed guidelines and steps from [4]. From 2003 to 2021, 245 primary articles on software usability with real-time data were found.

The study begins with common software usability model and standard attributes. The document recognises common system usability evaluation methods. This review lists several usability metrics. It also finds the phases where software usability issues are most addressed and explores different domains for their application.

From the selected primary studies, the main conclusions are obtained as follows:

1. Learnability, efficiency, satisfaction, and effectiveness are attributes commonly addressed in different existing software usability models and standards.
2. Usability testing, heuristic evaluations, and questionnaires are the most frequently used methods for usability assessment.
3. The metrics of ISO 9241-11 [5] and other measurement approaches [6] are used to estimate software usability.
4. In the design phase of software development, software usability problems are more addressed compared to other phases like the requirements and implementation phases.
5. Domains such as aviation, finance, industry, medicine, mobile, navigation, and news were assessed for PWA usability, but only for the overall application in each specific case.
6. No studies were found that collected important PWA development features.

Researchers disagree on PWA-developed app usability models. Thus, a consistent, non-redundant usability model is needed. Developers struggle to implement usability evaluation methods for these software products. Developers without usability engineering experience can use this software usability model. Thus, a single framework that is easy to develop and use in the application is needed.

Future research guidelines for progressive web application usability engineering:

Many current usability measurement methods do not encompass all aspects of ISO 9241-11 [5] in a single score or metric. Furthermore, some measures are challenging to calculate. Thus, a usability metric that covers PWA development and user experience is needed.

Using practical implementation, researchers should address usability issues at various software development stages. Thus, a single framework for usability engineering–software engineering integration is needed. Usability evaluation can improve online services.

Numerous studies on the usability evaluation of systems exist, but we have found a lack of research on PWA technologies like web and mobile in a single application. A scoping review is needed to create a characteristic study that explains software usability evaluation models, standards, evaluation methods, metrics, and approaches.

3 PWA Development Features

This project aims to present a quantitative study that assists in comparing PWA development solutions. It intends to analyse the attributes of APIs, selected based on their

relationship with usability and user experience, as well as their evaluation criteria. The solutions of these attributes served to guide aspects and improve the separation of concerns related to patterns for developers, although only four implementations oriented to these aspects have shown significant reusability.

A number of recent studies have compared usability assessment methods. One of the first analyses of these methods was by Gray et al. [7]. Five experiments were conducted to compare these methods. They wanted to prove that such experiments require scientific rigour. The authors say most experiments without method comparisons don't identify the comparisons. They also found that same-condition results can be misleading when comparing methods' efficacy.

Although studies [7] may be relevant to HCI, we will propose standards based on usability evaluations of PWA features. According to Hartson et al. [8], the lack of comparison criteria hinders competent assessment and comparison of usability evaluation methods. We examined several studies to determine usability assessment methods and evaluation measures. Most studies measured clear depth (the ratio between the number of actual usability problems found and the total number of actual usability problems). This research indicated that most usability evaluation and effectiveness comparison studies lacked the descriptive statistics needed for a meta-analysis.

A heuristics-based evaluation method was also presented. This study proposes that evaluators be given real problems to better determine the applicability of a set of heuristics to these problems. New procedures were created to select these issues [9]. These procedures can improve the comparison base for usability evaluation methods, but they only cover a small subset of inspection methods and aim to solve more immediate problems in human-computer interfaces. The studies' criticisms apply to web interfaces.

In [10], Alva et al. evaluated seven usability evaluation methods and tools for software products and web artefacts. This study assessed method similarity using ISO Standard 9241-11 principles [5]. This informal research has no structured questions or method identification.

In [11], Ivory and Hearst examined usability evaluation methods, reviewed automation capabilities, and proposed a classification taxonomy. Usability evaluation includes testing, inspection, evaluation, analytical modelling, and simulation. After applying the taxonomy to 128 usability evaluation methods, 58 were web-friendly. Usability tests and investigations are the only way to discover subjective user preferences and misconceptions, according to this research. Analytical modelling and simulation can help designers choose designs before costly development. The study suggests promising automated usability evaluation methods.

A rough web development model was created to link usability testing methods from the literature with the web development process [12]. The authors acknowledge that the research was incomplete but suggest it could guide web designers and developers. The research suggests competitive analysis, scenarios, inspection methods, log analysis, and questionnaires as evaluation methods. Many of the suggested methods are informal guidelines or ways to gather user interaction data.

Batra and Bishu [13] presented results from two web application usability studies. The first study compared user tests and heuristic evaluations for efficiency and

effectiveness. Both methods are efficient and effective for web usability evaluation and address different usability issues. The second study compared remote and traditional usability tests. The results show no significant difference between methods.

Most of the studies that are relevant to this project come from unstructured literature, surveys, or comparisons that do not have clear research questions, research methods, data extraction, or analysis processes. The authors chose the usability evaluation methods that were looked at based on their criteria. Moreover, most of these studies look at how to test usability in general user interfaces across all types of systems. However, there are not many that focus on testing methods related to the web, and even fewer that look at testing methods related to progressive web applications.

As a starting point for judging the APIs' metrics, 14 features were chosen that were most relevant to PWAs and had an effect on usability or user experience. Certainly, not all features have been explored, as the APIs are in constant process of review and improvement. However, one of the results already obtained was the elaboration of PWA metrics with the following relationships:

- A. One of the main foundations of a PWA is that it can be installed on the device, potentially differentiating it from a standard web application, as without this feature, the user would need to open the browser, type the application link, and then start using it. In a PWA, the user can access the application with a simple tap on an icon directly from the device's launch screen.
- B. After the PWA is installed on the user's device, it can have an icon along with its name to represent it on the home screen, just like apps from stores.
- C. During the first milliseconds of application startup, a splash screen, described in this project as the launch screen, is displayed. This screen occupies the entire available screen area of the device, indicating that the application is loading.
- D. Applications developed with native technologies do not allow users to select text in elements apart from editing fields. This happens because, in these technologies, text selection is restricted only to specific elements, unlike what occurs with web technologies, which were initially designed for reading texts. Allowing users to select texts from buttons, menus, and other non-paragraph text or editing field elements can provide a poor experience for the user, in addition to making it difficult to select texts that really need to be selected.
- E. Web pages optimised for mobile devices, especially PWAs, can block zooming across the entire application or specific elements. This is not a standard, but it is the default behaviour of native apps and should be considered during development to provide a closer experience to native applications.
- F. Mobile devices may have rounded corners, screen cutouts to accommodate cameras or sensors, or even folds in the screen. These screen spaces need to be respected, and no interactive elements should be used within these areas, as the user will not be able to interact with them.
- G. Web applications can use the offline loading feature, which is particularly recommended for PWAs, as the mobile device may be without a connection at any time.

- H. Like in the entire application, the OS status bar text must have sufficient contrast for the user to read the information. There is a particularity in developing this feature, as attention must be paid to the colours chosen for the background and text of the status bar.
- I. Some applications forget to provide specific navigation features for mobile devices, such as a side menu and especially a button to move back or forward between screens. This can produce an unwanted block in the user's navigation.
- J. Applications can be developed with light themes, recommended for the day or well-lit environments, as well as dark themes where there is little lighting.
- K. Each button in the application should have at least a visual or tactile response when it is pressed. This is particularly necessary for touch screens, where the precision of touch is less than that of a mouse pointer and it is not possible to hear and feel the click.
- L. Mobile devices allow easy manipulation of their orientation, which can be used both vertically and horizontally. In PWAs, it is important to check if the content adapts well to both cases, as the web technology standard is that it operates in two ways. Often, the developer might want the application to rotate to one of the sides.
- M. Various applications need to operate in different countries, consequently with different languages. In native applications, it is common to detect the language set in the operating system (OS) and apply it to the app.
- N. One of the main forms of engagement for native apps is the possibility for the user to receive notifications while the application is not in use. This feature can also be implemented with web technologies.

3.1 Usability Evaluation

A survey was made to find out how to make progressive web applications (PWAs) more usable. It will use a questionnaire to look at 14 features that are typical of PWAs. The results will show design guidelines that can be used to compare how satisfied users are with different applications already on the market.

As it is the responsibility of developers to decide whether to use API features, this decision may depend on the type of application. Therefore, the questionnaire seeks to validate the usability of these features in different applications, as well as evaluate the importance of implementing or not implementing each feature.

Usability Evaluation Questionnaire for PWA Features

- **A1** How difficult would you rate this installation?
- **A2** Were you able to open this application from the launch screen?
- **A3** Is it possible to observe the "App Name" in the list of open applications?
- **B1** Is there an icon with the application name "App Name" on your device's start screen?
- **B2** Do you think this icon helped you identify the application?

- **B3** Why do you agree or disagree?
- **C1** Was a launch screen with a brand or logo of the application displayed?
- **C2** Does the application's splash screen helped me identify it and provide a better experience than displaying a blank screen?
- **D1** Is it possible to select texts in the application?
- **D2** Was the text selection feature useful in this application?
- **E1** Is it possible to zoom in on the screen with this app?
- **E2** Does the zoom help you or create more complications in this app?
- **F1** Are there buttons or other interaction elements too close to the rounded corners and other cutouts of your device's screen?
- **F2** Do you find it more difficult to interact with these elements?
- **G1** Is it possible to use this application with the Internet connection turned off?
- **G2** I find it useful to be able to use this application without an Internet connection.
- **H1** The texts in the status bar (clock, signal, battery level) are legible.
- **I1** I can navigate within the app easily.
- **J1** Does this application change its colours when you switch the dark/light theme of the OS?
- **J2** The application implemented the light and dark theme feature as I expected.
- **K1** When pressing buttons or menu items, does the application show visual feedback of the element being pressed?
- **K2** I consider the interaction feedbacks shown by the application helpful in identifying which elements I was interacting with.
- **L1** When using the device in landscape mode, did the application content adapt to the new screen format?
- **L2** The application managed to use the screen space well, with the device in landscape or portrait.
- **M1** When changing the language of your device, did the application also change languages?
- **M2** I prefer the app to start in the same language as my device.
- **N1** With the app closed and notifications enabled, was it possible to receive any notifications?
- **N2** I believe that notifications are useful in reminding me to use the application.

4 Results Obtained

Preliminary Data Structure Overview The dataset comprises responses from a survey about the usability of various progressive web applications (PWAs) across different operating systems. Here's a brief overview of its structure:

- Total Entries: 226 responses.
- Columns: 46 columns, encompassing various questions and response types.
- Data Types: Percentages, booleans, scores ranging from 1 to 5.

Key Features

A. The survey covers responses about five different apps:

- Calculator: 37 responses
 - <https://bhar.app/calculator>
- Instagram: 48 responses
 - <https://instagram.com>
- Periodex: 54 responses
 - <https://periodex.co>
- The Cube: 47 responses
 - <https://bsehovac.github.io/the-cube>
- X (former Twitter): 40 responses
 - <https://x.com>

B. Two operating systems are represented:

- Android: 154 responses
- iOS: 72 responses

C. Questions: The survey includes a mix of questions with numerical scores and categorical responses (e.g. Yes/No, multiple-choice).

The reason for the lower response rate in the app “Calculator” is that the PWA in question is no longer available and it was successfully accessed on September 27th, 2023. Despite this, the other apps did not report any problems, and the most recent time they were accessed was on November 19th, 2023.

4.1 Questions A: PWA Installation and Initialisation

In assessing the difficulty of installation (Question A1), it was observed that users generally found the process relatively easy across all apps and operating systems. Notably, “The Cube” on iOS had the lowest average difficulty rating at 1.33, indicating a particularly user-friendly installation experience. In contrast, the highest average difficulty was reported for “Instagram” on Android, with an average rating of 2.48. Across the board, iOS users reported lower levels of difficulty compared to

Android users for the same apps, suggesting a smoother installation process on iOS devices.

Regarding the ability to open the app from the home screen (Question A2), the majority of users successfully accessed the apps, indicating good integration of PWAs with device interfaces. This success rate was especially prominent in “Instagram” and “X (Twitter)” on iOS, and “X (Twitter)” on Android, where all users could open the app from the home screen. However, “The Cube” on Android showed a slight deviation, with 6 out of 32 users unable to open the app from the home screen, signalling room for improvement in the app’s integration with Android devices.

4.2 Questions B: User Experience, App Visibility and Identification

Most users reported seeing the app icon and name on their home screens, which is crucial for app accessibility and user engagement. For instance, in “Periodex” on Android, about 91.4% (32 out of 35) of users saw the icon, while 8.6% did not. “The Cube” on Android had 100% of respondents (26 out of 26) confirming the icon’s visibility. However, “Calculator” showed a lower visibility rate with 96.4% on Android and 88.9% on iOS, indicating some users did not see the app icon. These figures suggest that while most PWAs are successful in making their icons visible, there are instances where this aspect could be improved.

The app icons were generally perceived as helpful in identifying the apps. In the case of “Instagram” on Android, for example, 96.3% (26 out of 27) of users found the icon helpful. This trend was consistent across most apps and operating systems, highlighting the importance of a well-designed and recognisable app icon in enhancing the user experience.

The number of responses to this question indicates a considerable level of user engagement and willingness to provide feedback, essential for the continual improvement of apps. While the specific content of the feedback is not detailed in the dataset, the response counts reflect active user interaction. This engagement is vital for developers to understand user experiences and make informed enhancements to the apps.

4.3 Questions C: Splash Screen Experience

Looking at how people answered questions C1 and C2 shows how the initialisation screen affects the user experience in different PWAs on different operating systems.

The majority of users across most apps and operating systems reported seeing an initialisation screen with the app’s branding. For example, “Instagram” on Android saw 96.3% of users confirming the presence of such a screen, and “The Cube” on both Android and iOS platforms reported 100% visibility. However, there were

notable exceptions, such as “Calculator” on iOS and “The Cube” on iOS, where 88.9 and 46.7% of users, respectively, did not see the initialisation screen. This variation suggests differences in the implementation of the initialisation screen across different apps and platforms.

The initialisation screen was generally found to be effective in aiding app identification and providing a better experience than a blank screen. For instance, “Instagram” on Android had 85.2% of users affirming the effectiveness of the initialisation screen, and “X (Twitter)” on iOS reported 92.3% positive responses. However, the effectiveness varied, with “Periodex” on iOS showing a lower effectiveness rate at 43.75%.

4.4 Questions D: Text Selection and Its Utility

Answers to survey questions D1 and D2, which were about how text selection works and how useful people think it is, show how people interact with content in different PWAs on different operating systems.

The capability to select text within apps varied significantly across different apps and operating systems. For example, in “Periodex” on Android, approximately 94.3% of users were able to select text, indicating a high level of functionality. In contrast, “The Cube” on both Android and iOS platforms showed a significantly lower rate, with only 7.7% and 6.7% of users, respectively, reporting the ability to select text. This disparity suggests that text selection functionality is not uniformly implemented across all PWAs, impacting the user’s ability to interact with content.

Regarding the utility of the text selection feature, responses varied based on user perception and app context. For “X (Twitter)” on Android, a significant 80.8% of users found the text selection feature useful, indicating its relevance in the app’s context. Conversely, “The Cube” on both Android and iOS exhibited lower utility rates, with only 7.7% and 6.7% of users, respectively, finding the feature useful. These differences highlight that the utility of text selection is highly dependent on the specific use case and design of each PWA.

4.5 Questions E: Screen Zoom User Experience

The ability to enlarge the screen varies considerably among apps and operating systems. In “Periodex” on Android, 100% of users reported the ability to enlarge the screen, showcasing high functionality. However, “The Cube” on Android and iOS showed a significant limitation in this aspect, with only 11.5 and 0% of users, respectively, being able to enlarge the screen. “Instagram” also exhibited a lower rate of screen enlargement capability, with only 33.3% on Android and 35.7% on iOS. These figures indicate that the screen enlargement feature is not uniformly available

or functional across all PWAs, affecting the adaptability of the app interface to user preferences.

The perceived utility of the screen enlargement feature demonstrates varied user experiences. For instance, “Periodex” on Android had 48.6% of users finding this feature useful, while “The Cube” on Android had a much lower utility rate at 7.7%. This variation suggests that the usefulness of screen enlargement depends heavily on the app’s content and design. In some cases, it enhances the user experience by providing better visibility and readability, while in others, it may not add significant value or could even complicate the interface.

The implementation and perceived utility of the screen enlargement feature in PWAs are diverse. While some apps excel at providing this functionality, thereby enhancing the user experience, others lack either the availability or utility of this feature. These insights suggest a need for developers to consider the specific requirements and design elements of their apps when incorporating such features, ensuring they align with user needs and improve the overall app experience.

4.6 Questions F: Screen Edge and Corners Experience

When users were asked if there were buttons or other interaction elements too close to the rounded corners or other cuts on the screen, the majority of respondents, approximately 61.65%, answered no, indicating no issues with element placement. Around 19.42% of respondents answered yes, suggesting some difficulty with interface elements due to their positioning. Additionally, 14.56% reported that their device’s screen did not have rounded corners, special cutouts for the camera, sensors, or folds, while 4.37% indicated their device did not have rounded corners or folds at all.

About the difficulty level in interacting with these elements, the responses were numerical. The average score was approximately 2.25 on a scale, with a standard deviation of about 1.35. This suggests a mild level of difficulty overall, with some variation among users. The 50th percentile (median) score was 2, indicating that the central tendency of the responses leans towards a lower difficulty level.

4.7 Questions G: Offline Usability

In the examination of the offline usability of various apps across different operating systems, as queried in question G1, a nuanced pattern emerges. The “Calculator” app exhibits universal offline functionality among Android users and most iOS users. In stark contrast, Instagram, predominantly an Internet-dependent application, is reported to be usable offline by a mere fraction of its user base, regardless of the operating system. “Periodex” and The Cube demonstrate robust offline capabilities across both Android and iOS platforms. Conversely, X (Twitter) shows limited offline

usability, with only a small segment of users across both platforms affirming this feature.

Regarding the perceived utility of offline functionality, as explored in question G2, divergent views are apparent across apps and operating systems. Users of the Calculator app, particularly on iOS, highly value its offline utility, while Instagram users exhibit a notably lower appreciation for offline usage, aligning with its Internet-centric nature. “Periodex” and “The Cube” are commended for their offline utility on both platforms, underscoring a positive user perception. “X (Twitter)”, however, garners moderate scores, reflecting a less favourable view of its offline capabilities.

4.8 *Question H: Status Bar Readability*

The assessment of the legibility of status bar information, as probed in question H1, reveals generally favourable responses across all apps and operating systems. The Calculator app, while scoring lower on Android, achieves perfect scores for iOS users. Instagram and The Cube maintain high legibility scores, indicative of user satisfaction with the visibility of crucial status bar elements. Similarly, Periodex and X (Twitter) also receive high scores, suggesting that these apps effectively present essential status information across both platforms.

4.9 *Question I: Navigation*

Regarding question I1, which aims to evaluate the ease of navigation within the app, the data reveals a strong positive response. A significant 81.07% of participants rated their experience at the highest level, suggesting that the app is exceptionally user-friendly and intuitive in terms of navigation. This high satisfaction rate is indicative of a well-designed user interface, facilitating a seamless and efficient user experience. On the other hand, the lower ratings, collectively constituting 18.93%, point towards areas where the app’s navigation could be enhanced.

4.10 *Questions J: OS Light and Dark Mode*

In question J1, the focus is on assessing the app’s ability to adapt its colour scheme in response to the OS’s dark or light mode. Here, 68.93% of respondents affirm the app’s adaptability to these theme changes, indicating a good level of integration with iOS’s features. This adaptability is crucial for providing a consistent user experience that aligns with the user’s system preferences. However, the 31.07% of respondents who did not observe this adaptability highlight a potential area for improvement, either in the app’s functionality or in communicating this feature to the users.

Question J2 deals with user satisfaction regarding the implementation of light and dark themes within the app. In this regard, 61.17% of users expressed high levels of satisfaction, reflecting that the app meets or even surpasses their expectations in implementing these themes. The variety in satisfaction levels, including 17.96% of users rating their satisfaction as the lowest, underscores the subjective nature of user experience and highlights the challenge of catering to a diverse user base with varying expectations.

4.11 Questions K: Design Elements Touch Feedback

For question K1, which assesses whether the app provides visual feedback when buttons or menu items are pressed, the results vary significantly across different apps and OS. For instance, in the Android environment, the highest affirmation for visual feedback is observed in the app “Periodex” (97.14%), while the lowest is in “The Cube” (53.85%). In contrast, on iOS, “X (Twitter)” and “Instagram” show high levels of positive feedback at 92.31% and 92.86%, respectively, but “The Cube” again has a lower rate, with only 26.67% affirming visual feedback. These discrepancies suggest that the user experience regarding interaction feedback is highly dependent on both the specific app and the OS.

Question K2 delves into how helpful the users find these interaction feedbacks in identifying the elements they are interacting with. The responses here also show significant variation across different apps and OS. On Android, the app “X (Twitter)” has a high level of satisfaction (5 out of 5) at 58.82%, while “Calculator” has a lower satisfaction level at 34.78%. In the iOS environment, “X (Twitter)” again stands out with high satisfaction at 84.62%, compared to “The Cube” which has a lower satisfaction rate of 40%. This variability indicates that the effectiveness and perception of interaction feedback are not only by app but also across different OSs.

4.12 Questions L: Horizontal Mode

Focusing on the usability of a change in the device orientation into landscape mode, notable insights were gathered from user responses. Question L1 delved into the flexibility of application usage in different orientations. A significant majority, 66.99% of the respondents, confirmed their ability to use the application in both horizontal and vertical modes. This adaptability highlights the versatile technical nature of PWAs.

Additionally, question L2 provided further interesting data. The average response score was 3.47, with a median of 4. Notably, 57.28% of participants either agreed or strongly agreed on the usefulness of this feature. These findings were particularly pronounced in the context of the “Periodex” application, where 84.31% of users acknowledged its utility. These statistics not only underscore the effectiveness of

the device's landscape mode but also emphasise the specific success of the certain applications like "Periodex" application in enhancing user experience.

4.13 Questions M: OS Default Language

For question M1, which assesses whether the app changes its language in response to the language settings of the user's device, the results exhibit considerable variation across different apps and OS. On Android, the percentage of users confirming this functionality is highest in the app "X (Twitter)" (84.62%), indicating strong language adaptability, while it is notably lower in "Calculator" (32.00%) and "The Cube" (38.46%). In the iOS environment, "Instagram" and "X (Twitter)" demonstrate high adaptability, with 92.86 and 92.31% affirming the change, respectively. However, "Calculator" and "The Cube" show a significantly lower percentage, with 11.11% and 13.33% respectively, suggesting limited language adaptability in these apps on iOS.

Question M2 explores user preferences regarding the app starting in the same language as their device. The responses indicate that on Android, a high preference for this feature is seen in "Periodex" (77.14%) and "Instagram" (81.48%). On iOS, the preference is overwhelmingly high in "Instagram" "Periodex" and "X (Twitter)", all scoring 100%. This suggests a strong user preference for apps to align with the device's language setting, enhancing user comfort and app accessibility.

4.14 Questions N: App Notification

Question N1 evaluates whether users can receive notifications from the app. The data shows a high affirmation of notification receipt across both Android and iOS for "Instagram" (92.59% on Android and 92.86% on iOS) and "X (Twitter)" (84.62% on both Android and iOS). This high percentage suggests that both apps are effectively utilising notifications to keep users informed and engaged. The lower rate of negative responses (ranging from 7.14 to 15.38% across apps and platforms) indicates that a small subset of users either do not receive notifications or are unaware of this feature.

Question N2 explores user beliefs about whether receiving notifications increases their engagement with the app. On Android, a substantial majority believe in the positive impact of notifications on engagement, with 77.78% for "Instagram" and 80.77% for "X (Twitter)" rating their agreement as 5 out of 5. Similarly, on iOS, the belief in the positive impact of notifications is prevalent, albeit slightly less pronounced, with 64.29% for "Instagram" and 69.23% for "X (Twitter)" rating their agreement as 5 out of 5. This reflects a strong user perception that notifications enhance engagement with these apps.

5 Conclusions

This report provides a detailed analysis of a survey that was carried out to evaluate the usability of various progressive web applications (PWAs) on various operating systems. The responses from users who evaluated the app on various criteria, including how difficult it was to install, how visible it was, and how well it functioned. The overall user experience is included in the dataset. The data showed that users had a wide variety of experiences, with the mean scores revealing broad patterns in terms of usability considerations. The areas in which user opinions varied significantly were highlighted by the standard deviations.

The findings point to a user-friendly installation experience for both iOS and Android users, with a more favourable tendency towards iOS users than Android users. PWAs have been effectively integrated with the user interface of the operating system, as evidenced by the high success rate of opening apps from the home screen, particularly on iOS; however, there are still some challenges to be overcome to perfect this experience on Android platforms.

In conclusion, the responses to the survey shed light on the successful visibility of app icons in the majority of PWAs, their role in assisting app identification, as well as the active participation of users in providing feedback. Even though the effectiveness of icon visibility and identification is high, there is room for improvement, particularly in ensuring that the visibility is consistent across all devices and operating systems.

The responses highlight the significance of an initialisation screen in improving the user experience and assisting in app recognition. Although this function has been successfully incorporated into the vast majority of apps, the usefulness of such screens and whether they are present varies widely across both app types and operating systems. This suggests that certain applications have room for improvement about their initialisation screens, which will help to ensure that users have positive and consistent first impressions.

Text selection functionality is present in many PWAs; however, its implementation and the extent to which it is perceived to be useful varies greatly between apps and operating systems. This suggests that developers need to consider the specific context and user needs of their apps when implementing and designing features like text selection, making sure that they add value to the user experience and meet user expectations in the process.

The screen zoom feature in PWAs can be implemented in various ways, and users' perceptions of its usefulness can vary. Some applications excel at providing this functionality, which in turn improves the user experience. Other applications, on the other hand, either do not have this feature available or do not make effective use of it. When incorporating such features into their apps, developers should consider the specific requirements and design elements of their apps to ensure that the features are in line with user needs and contribute to an overall improvement in the experience provided by the app.

The research reveals that there is a complex landscape when it comes to the usability of mobile applications, particularly regarding the functionality of apps that can be used offline and the clarity of information displayed in status bars. The nature of both the operating system and the application has a significant impact on user experiences and perceptions, which highlights the complex challenges involved in optimising app design for the wide range of user requirements and technological environments. The findings highlight the necessity for app developers to consider these nuanced user preferences and technical constraints in their design and development processes, with the goal of increasing overall user satisfaction and engagement.

The findings point to a generally favourable experience with PWAs, although they highlight the necessity for careful consideration of zoom and text selection functionalities, as well as the need to optimise the use of screen space. These findings provide PWA developers with valuable insights that they can use to improve the overall usability of their applications as well as the user experience.

The features of PWA development that make use of Large Language Models (LLMs), smart devices with natural language communication, wearable devices, or any embedded systems were not evaluated in this study (IoT). The usability of the designs for desktop computers or laptop computers was not considered, even though these types of computers require more specialised research and evaluation methods for usability.

The usability features of the app were well received by users, particularly in regard to how easily it could be navigated and how responsive the themes were. The fact that the app is perceived by the vast majority of users to be both user-friendly and responsive to adjustments made to the operating system is evidence of successful user interface and user experience design. Despite this, the areas in which a sizeable proportion of users have voiced concerns or expressed dissatisfaction, most notably in terms of the adaptability of the theme, point to areas in which further refinements could be made. It is essential to address these concerns to enhance the appeal and functionality of the app, which allows it to cater to a wider variety of user preferences and expectations.

The examination of questions K1 and K2 across various apps and operating systems reveals that the ways in which users interact with visual feedback and the extent to which they find it helpful are highly dependent on the context in which they find themselves. A significant part in the formation of these experiences is played by elements such as the characteristics of the operating system and the design of the application. Apps such as “X (Twitter)” and “Instagram” tend to show higher levels of user satisfaction in terms of the visual feedback they provide and how helpful it is across both Android and iOS, whereas apps such as “The Cube” tend to show lower levels of user satisfaction, especially on iOS. This highlights the importance of individualised interaction design strategies for various platforms, as well as the requirement for continuous evaluation and improvement, to cater to a wide variety of user expectations and technological contexts.

The responses to questions M1 and M2 across various apps and operating systems reveal that users place a high value on language adaptability in apps, and they prefer apps that are compatible with the language settings of their respective devices. While

some applications, such as “X (Twitter)” and “Instagram” are thought to perform exceptionally well in this regard on both Android and iOS platforms, others, such as “Calculator” and “The Cube” may need to improve the language adaptability features that they offer, particularly on iOS. These findings highlight the importance of linguistic adaptability in app design, which is crucial for providing a personalised and user-friendly experience across diverse user bases. This is crucial for providing a linguistic adaptability in app design.

According to the findings of the analysis of N1 and N2, “Instagram” and “X (Twitter)” are both able to successfully deliver notifications to the vast majority of their users on both the Android and iOS platforms. In addition, a sizeable percentage of users on both platforms believe that these notifications have a positive impact on the degree to which they engage with the apps. These findings highlight the importance of well-integrated notification systems in mobile apps as a means of maintaining user engagement and interest as a means of keeping users engaged and interested. Having said that, the fact that there is a subset of users who either do not receive notifications or do not consider them to be beneficial suggests that there is room for improvement in the delivery of notifications as well as in the education of users regarding this feature.

In conclusion, this comprehensive study on progressive web applications (PWAs) usability across operating systems illuminates the complexity of user experience and app functionality. We found that PWAs generally provide a good user experience, but feature implementation, particularly text selection, screen zoom, and theme adaptability, needs improvement. As user satisfaction varies across apps and operating systems, the study emphasises the importance of platform characteristics and user preferences in app design. Additionally, linguistic adaptability and integrated notification systems improve user engagement and satisfaction. These insights reflect the current state of PWA usability and provide a roadmap for developers optimising their apps for diverse user groups and technological environments. As the digital landscape changes, PWAs must continuously evaluate and adapt to meet user needs and expectations, resulting in more intuitive, efficient, and satisfying user experiences.

- We would like to extend our gratitude to Fapesp (*Fundação de Apoio à Pesquisa no Estado de São Paulo*) for their assistance with this research.

References

1. Insfran E, Fernandez A (2008) A systematic review of usability evaluation in web development. In: vol 5176. https://doi.org/10.1007/978-3-540-85200-1_10
2. Hornbæk K (2006) Current practice in measuring usability: challenges to usability studies and research. *Int J Hum Comput Stud* 64(2):79–102. Academic Press, Inc., USA. <https://doi.org/10.1016/j.ijhcs.2005.06.002>
3. Mendes E (2005) A systematic review of web engineering research. In: 2005 international symposium on empirical software engineering. <https://doi.org/10.1109/ISESE.2005.1541857>

4. Kitchenham B (2004) Procedures for performing systematic reviews. Keele University. Technical Report TR/SE-0401, Department of Computer Science, Keele University, UK
5. ISO 9241-11 (2018) Ergonomics of human-system interaction—part 11: usability: definitions and concepts. Standard, International Organization for Standardization, Geneva/Switzerland
6. Schusteritsch R, Wei CY, LaRosa M (2007) Towards the perfect infrastructure for usability testing on mobile devices. In: CHI '07 extended abstracts on human factors in computing systems, pp 1839–1844. Association for Computing Machinery, New York, NY, USA. ISBN 9781595936424
7. Gray WD, Salzman MC (1998) Damaged merchandise? A review of experiments that compare usability evaluation methods. *Hum Comput Interact* 13(3):203–261. L. Erlbaum Associates Inc., USA. https://doi.org/10.1207/s15327051hci1303_2
8. Hartson H, Andre T, Williges R (2001) Criteria for evaluating usability evaluation methods. *Int J Hum Comput Interaction* 13:373–410. https://doi.org/10.1207/S15327590IJHC1304_03
9. Somervell J, Mccrickard D (2004) Comparing generic versus specific heuristics: illustrating a new UEM comparison technique. In: Proceedings of the human factors and ergonomics society annual meeting, vol 48. <https://doi.org/10.1177/154193120404802108>
10. Alva MEO, Martinez P, AB, Cueva L, JM, Sagastegui Ch, TH, Lopez P, B (2003) Comparison of methods and existing tools for the measurement of usability in the web. In: *Web engineering*, pp 386–389. Springer, Berlin
11. Ivory MY, Hearst MA (2001) The state of the art in automating usability evaluation of user interfaces. *ACM Comput Surv* 33(4):470–516. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/503112.503114>
12. Cunliffe D (2000) Developing usable Web sites—a review and model. In: *Internet research*, vol 10, pp 295–308. <https://doi.org/10.1108/10662240010342577>
13. Batra S, Bishu RR (2007) Web usability and evaluation: issues and concerns. In: *Usability and internationalization. HCI and culture*. Springer, Berlin, pp 243–249

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



An Automated and Goal-Oriented Clustering Procedure



Oded Koren , Michal Koren , and Or Peretz 

Abstract Clustering techniques are convenient tools for preparing and organizing unstructured and unclassified data. Depending on the data, they can be used to prepare for an analysis or to gain insight. However, choosing a clustering technique can be challenging when dealing with high-dimensional datasets. Most often, application requirements and data distribution need to be considered. Since clustering is defined as a complex problem to calculate, different algorithms may produce different results that meet the application's needs. This study presents an automated threshold-based and goal-oriented clustering procedure. It is based on the AutoML mechanism to estimate the most suitable hyperparameters according to predefined needs and can learn four clustering performance metrics thresholds for a given dataset. The significant advantages of this method are the automatic selection of clustering technique (i.e., partitional, hierarchical, density-based, or graph-based) and the ability to determine the output dynamically, according to predefined goals. We tested our method over four datasets and analyzed the results according to different goals. The results show that our method improved the silhouette score by 549.5% (from 0.105 to 0.682) compared to popular and commonly used K-means. Furthermore, clustering based on multiple metrics yielded more information than clustering by a single metric.

Keywords Clustering · AutoML · Threshold learning · Decision-making

O. Koren · M. Koren · O. Peretz (✉)

School of Industrial Engineering and Management, Shenkar—Engineering, Design, Art, Anne Frank 12, Ramat-Gan, Israel

e-mail: or.perets@shenkar.ac.il

O. Koren

e-mail: odedkoren@shenkar.ac.il

M. Koren

e-mail: michal.koren@shenkar.ac.il

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_15

207

1 Introduction

Unsupervised learning algorithms are trained without predefined knowledge [1] (i.e., unlabeled data are used as input) to search for hidden patterns or perform data groupings. Data similarity (or dissimilarity) measures are a central tool in this type of learning [2]. These measures are used to identify behavior patterns or data groups as there is no prior information on the data. The division of the data into groups is called clustering [3], and many types of clustering techniques exist, such as partitioning [4], hierarchical clustering [5], density-based clustering (also known as kernel density estimation) [6, 7], and graph-based clustering [8–12]. However, there are many challenges in the field of clustering:

- (a) It is impossible to determine the optimal number of clusters, which is known as an NP-hard problem.
- (b) Choosing the suitable type or technique for high-dimensional datasets is difficult.
- (c) It may be challenging to interpret the meaning of each cluster among high-dimensional data.

To address the above issues, studies have suggested using the field of ensemble learning, which involves learning several algorithms simultaneously with similar goals [13]. This can be accomplished by executing the same algorithm with different hyperparameters or running different algorithms and analyzing the results in several ways (mainly by voting). Thus, it is possible to approximate an almost optimal solution by weighting the algorithms using the ensemble manner [14, 15]. The main advantage of this learning method is the improvement of predictions and achievement of a better performance than any single contributing model.

Further attempts have been made to optimize and solve the clustering issues for high-dimensional data clustering [16]. New clustering algorithms have been developed that focus on both (a) clustering by subspace (some features are used, and the cluster models include the relevant features for each cluster) and (b) clustering by correlation, which seeks out rotated (i.e., correlated) subspace clusters that can be modeled by correlating their features. Examples of such algorithms are CLIQUE-clustering [17] and SUBCLU [18], among others.

Measuring the performance of an unsupervised algorithm is different from supervised learning as there is no dependent variable and no predictions to compare. There are many valuable metrics to measure the performance of a clustering algorithm, such as the silhouette score [19–21], Calinski–Harabasz (CH) score for indicating the variance ratio [22], Dunn index (DUNN) [23] for measuring the distance ratio between the clusters, and Davies Bouldin index (DBI) for describing the level of similarity between clusters [24].

Automated machine learning (AutoML) is emerging along with ML techniques to allow the automation of processes and tasks required to solve ML problems [25, 26]. Several researchers in the field of AutoML have addressed this issue by automatically determining appropriate algorithms and their hyperparameters [27].

However, these techniques mainly apply to supervised learning tasks [28]. Even with the help of automatic methods, the main challenge is to explore and find the most suitable hyperparameters (clustering technique, number of clusters, etc.) for a given dataset. AutoML procedures for clustering have been proposed using threshold learning [29–31], analysis of autonomous vehicles to estimate the risk in decision-making [32], and automated clustering for the event logs of a system [33].

This study presents an automated and goal-oriented clustering procedure. It includes the selection of a clustering technique and estimation of the optimal hyperparameters according to predefined goals. It is based on threshold learning, ensuring that all performance measures are as close to optimal as possible. There are two main advantages of this method: first, the automatic selection of clustering technique (i.e., partitional, hierarchical, density-based, or graph-based), and second, the ability to determine the output dynamically according to the goal. Section 2 presents the Automated Clustering procedure, implementation, and performance metrics used in this study. Section 3 describes the empirical study and a detailed scenario, including different goal demonstrations. The results of our method were compared over four datasets, and the results are presented in Sect. 4. Last, Sect. 5 discusses the main conclusions and suggestions for future directions.

2 Automated Clustering Procedure

This section presents and describes the Automated Clustering procedure. It uses eight clustering techniques, divided into four main categories: partitional, hierarchical, density-based, and graph-based. First, we will detail the methods and grid of parameters defined for exploration. Second, we will describe the four clustering performance metrics used for threshold learning. Last, we will present the algorithm and its implementation.

2.1 Clustering Methods and Hyperparameters

This study used eight clustering algorithms. The algorithms were chosen to cover the four existing main categories. The following are the algorithms and their categories:

1. (Partition) **K-Means** [34]—Partitions the records into k clusters such that the intracluster similarity is high and the intercluster similarity is low. The sum of squared error (SSE) is mainly used as the cost function.
2. (Hierarchical) **Hierarchical Clustering** [35]—Builds a hierarchy of clusters by partitioning the dataset sequentially. Here, the agglomerative version was used, which describes the bottom-up approach: observations are placed in clusters, and pairs of clusters are merged while moving up the hierarchy.

3. (Hierarchical) **BIRCH** [36]—Balanced Iterative Reducing and Clustering using Hierarchies is a clustering algorithm that is highly useful among large datasets. It produces small and compact summaries of large datasets that retain as much information as possible.
4. (Graph-based) **Spectral Clustering** [37]—This approach is based on graph theory, which uses the eigenvalues of special matrices to identify communities of nodes in a graph (subgraphs) representing clusters.
5. (Density-based) **Mean Shift** [38, 39]—Iteratively shifts records toward the mode to assign them to clusters. At each iteration, a record will move closer to where the most points are.
6. (Density-based) **DBSCAN** [40]—Density-Based Spatial Clustering of Applications with Noise (DBSCAN) inputs two parameters: epsilon, which specifies how close points should be to each other to be considered a part of a cluster, and min samples, which are the minimum number of points necessary to form a cluster. It starts with an arbitrary record and creates clusters such that the records in each cluster satisfy the epsilon condition.
7. (Density-based) **OPTICS** [41]—Ordering Points To Identify Cluster Structure (OPTICS) is a generalization of the DBSCAN algorithm with two additional parameters: core distance, meaning the minimum radius (i.e., epsilon) that is required to make a record a core point, and reachability distance, which is the smallest distance from the record to its core object.
8. (Graph-based) **Affinity Propagation** [42, 43]—Density records (also known as the exemplars) are identified by the concept of message passing and form clusters around the density areas. To avoid numerical oscillations, damping reduces the responsibility and availability of messages to avoid using many exemplars.

To understand the motivation for this study, Fig. 1 presents a comparison of clustering techniques and their performances over different data distributions. For the same distribution, different clustering algorithms return different results. The figure illustrates the difficulty in choosing an appropriate clustering algorithm for a given dataset as the number of features increases. It is important to note that the source code for the comparison is based on the Scikit-learn tool for Python [45].

2.2 Clustering Performance Metrics

For threshold learning, we used the following clustering performance measures:

1. **Silhouette Coefficient (SIL)**—ranges from -1 to 1 , where higher scores indicate better cluster definition. It is possible for clusters to overlap and yield coefficients around zero [19–21].
2. **Calinski–Harabasz (CH) Score**—represents the variance ratio [22]. Higher scores indicate that clusters are dense and well separated [46, 47].
3. **The Dunn Index (DUNN)** [23]—measures the distance between clusters as the ratio between the smallest intercluster distance and the largest intracluster

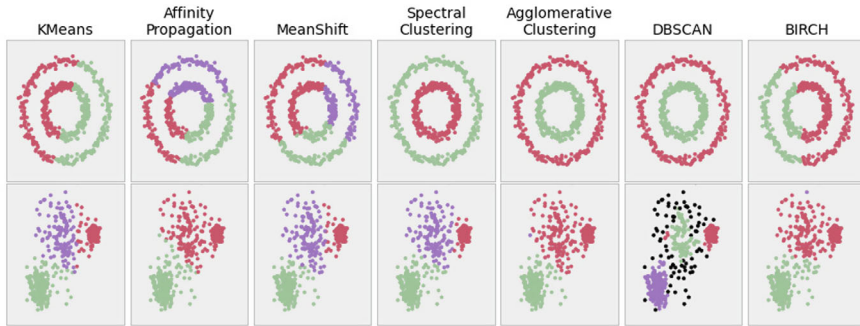


Fig. 1 A comparison of clustering algorithm performances

distance. High DUNN values improve clustering as the clusters are closer together than individual observations, despite the clusters themselves being farther apart.

4. **Davies Bouldin Index (DBI)**—indicates the level of similarity between clusters [24]. In many fields, clustering techniques are optimized using this method [47–49]. However, this technique can only be used to measure clustering performance due to a drawback—an optimal DBI value does not necessarily imply the best information retrieval results. In contrast to the other measures, lower values indicate a higher performance rate. To standardize all scales, we transformed those values as follows:

Let $A = a_1, a_2, \dots, a_n$ be a set of DBI values. For each $a_i \in A$ perform:

$$\max(A) - a_i + \min(A)$$

The new value is bounded between the original values and in reverse of the original order. For any $a_i, a_j \in A$, such that $a_i \geq a_j$, it holds that:

$$\begin{aligned} a_i &\geq a_j \\ -a_i &\leq -a_j \\ \max(A) - a_i &\leq \max(A) - a_j \\ \max(A) - a_i + \min(A) &\leq \max(A) - a_j + \min(A) \end{aligned}$$

2.3 Automated Clustering Algorithm

The algorithm inputs a dataset (D), consisting of m features denoted as $F = \{F_1, \dots, F_m\}$, a threshold (t) between zero and one, a set of clustering algorithms (A) and performance metrics (M), and the user goal (G). The goals are divided into two main groups: supervised and unsupervised analysis. There can be maximization

or minimization aims, such as maximizing the accuracy, minimizing the rate of false positive predictions, maximizing the silhouette score, a combination of goals, and more. The algorithm consists of a main procedure, Auto-Clustering, and a helper procedure, search-hyperparameters.

To begin, the method normalizes the dataset by Z-normalization and uses the search-hyperparameters procedure to evaluate all clustering algorithms, as described in Sect. 2.1. Each algorithm is then applied in an ensemble manner—multiple and independent algorithms running with the parameters presented in Table 1. Let M be a set of performance metrics, as described in Sect. 2.2. The results are assigned to matrix R with $|M| + 2$ columns, in which the two additional columns represent the algorithm name and tested hyperparameter. Each row has the form:

(algorithm – name, hyper – parameter, SIL, CH, DBI, DUNN)

Next, the method examines the percentage of improvement between two consecutive iterations of the same technique. Consider the i^{th} iteration, with the improvement for each $m_j \in M$ calculated by $\frac{R[i+1,m_j]-R[i,m_j]}{R[i,m_j]}$. In cases where the percentage is higher than the predefined threshold (t), the method marks m_j and the iteration parameters. At the end of the search-hyperparameters procedure, the marked clustering algorithms and their hyperparameters, denoted as P , are returned.

In the second phase of the algorithm, the Auto-Clustering creates the labeled dataset using the output of each clustering algorithm existing in P and in consideration of two options. First, in the case where G is the supervised analysis, it evaluates a supervised learning algorithm and calculates the performance metrics for analysis. Otherwise, if G is the unsupervised analysis, it uses the clustering performance

Table 1 The grid of hyperparameters used for each clustering algorithm

Category	Algorithm	Hyperparameters	Range
Partition	K-means	Number of clusters	2, 3, ..., \sqrt{n}
Hierarchical	Agglomerative		
	BIRCH		
Graph-based	Spectral	Damping	0.5, 0.55, ..., 1.0
	Affinity Propagation		
Kernel Density Estimation	DBSCAN	Min samples to perform a cluster	2, 3, ..., $\frac{n}{2}$
	OPTICS		
	Mean shift		

Notes

1. For the DBSCAN algorithm, the epsilon (i.e., the maximum distance between two points) was calculated by the distance to the nearest records, sorting, and choosing the epsilon that maximizes the differences [44].
2. The value n is the number of records in the dataset.
3. A variety of parameters and influences can affect the measures and values presented here, such as the dataset, use cases, business requirements, targets, and market needs.

metrics, as described in Sect. 2.2, and returns the dataset that satisfies G (i.e., the user goal) with an additional feature that represents the desired cluster.

2.4 Implementation

Search-Hyperparameters(\mathbf{D} , \mathbf{A} , \mathbf{M} , \mathbf{t})

```

 $F \leftarrow \{F_1, \dots, F_m\}$  set of features in  $D$ 
Normalize each  $F_i \in F$  by Z-score
 $R \leftarrow \emptyset$ 
For each  $ALG \in \mathbf{A}$ :
    Run ALG with the hyperparameters range
    For  $m_j \in \mathbf{M}$ : calculate  $m_j$  for ALG and store it in
         $R[ALG, m_j]$ 
for  $i \leftarrow 1$  to  $|R|$ :
    If  $|\frac{R[i+1, m_j] - R[i, m_j]}{R[i, m_j]}| \geq t$ :
        Mark TRUE for technique  $m_j$  in iteration  $i + 1$ 
    Otherwise, mark FALSE
Return the rows in  $R$  that satisfy TRUE in all measures

```

AutoClustering(\mathbf{D} , \mathbf{A} , \mathbf{M} , \mathbf{t} , \mathbf{G})

```

 $P \leftarrow$  Search-Hyperparameters( $\mathbf{D}$ ,  $\mathbf{A}$ ,  $\mathbf{M}$ ,  $\mathbf{t}$ )
For each technique ( $t_i$ )  $\in P$ :
    Create (sub)dataset labeled with the clusters yielded from  $t_i$ 
    If  $G$  is the supervised analysis:
        Evaluate the supervised learning model
    Otherwise:
        Evaluate the unsupervised learning performance metrics
Return the dataset that satisfies  $G$ 

```

Notes

1. The ALG parameter is one of the clustering algorithms presented in Sect. 2.1.
2. Note that, in cases of a supervised analysis, it is possible to calculate all the unsupervised performance metrics.

Table 2 provides the legend and description of the flow presented in Fig. 2.

Table 2 Legend and description for Fig. 2

Step	Input/output	Description
1	Dataset (D), a set of goals, and a threshold	An example of input: (<i>dataset = D, goals = [maximize homogeneity, silhouette > 0.5], t = 0.2</i>)
2	Inputs step (1) and produces clustering performance metric values for each technique and hyperparameter	The procedure begins by exploring the hyperparameters' values over different clustering algorithms, based on the clustering performance metrics described in Sect. 2.2
3	Inputs step (2) and yields a set of valid clustering techniques for dataset (D)	After completing step two, it examines the percentage of improvement between two consecutive iterations of the same technique. If the improvement is less than the predefined threshold, it eliminates the technique and the hyperparameter
4	Inputs step (3) and updates the dataset (D) by adding clusters' feature(s)	For example: <i>BIRCH with 7 clusters, Mean Shift with a bandwidth of 4, K-means with 5 clusters</i>
5	Inputs step (4) and examines which clusters meet the goal(s)	The procedure updates the dataset (D) and adds clusters' feature(s), one for each valid clustering technique achieved in step four. It examines which clustering technique(s) meet the defined goal(s)
6	Inputs step (5) and returns the clusters that satisfy the predefined goal(s)	Returns the dataset with the relevant clusters feature(s)

3 Empirical Study

3.1 Data Sources

Four datasets were compared to examine the automated clustering method:

1. **Students' Academic Success** [50]—A dataset containing information about undergraduate students from multiple higher education institutions. It provides information about 4424 students' enrollment and academic performance across 36 features. The target variable has three options: graduate, dropout, or enrolled.
2. **Fetal Health** [51]—A fetal health classification dataset to prevent maternal and infant mortality. There are 2126 observations over 21 features, as well as one target variable with three possible values: normal, suspect, or pathological.
3. **Heart Failure** [52]—A total of 299 patients who experienced heart failure. The dataset contains 12 clinical features and a Boolean target variable that represents if the patient had heart failure.
4. **Hepatitis C** [53]—Laboratory values and demographic information about 611 blood donors and Hepatitis C patients over 13 features.

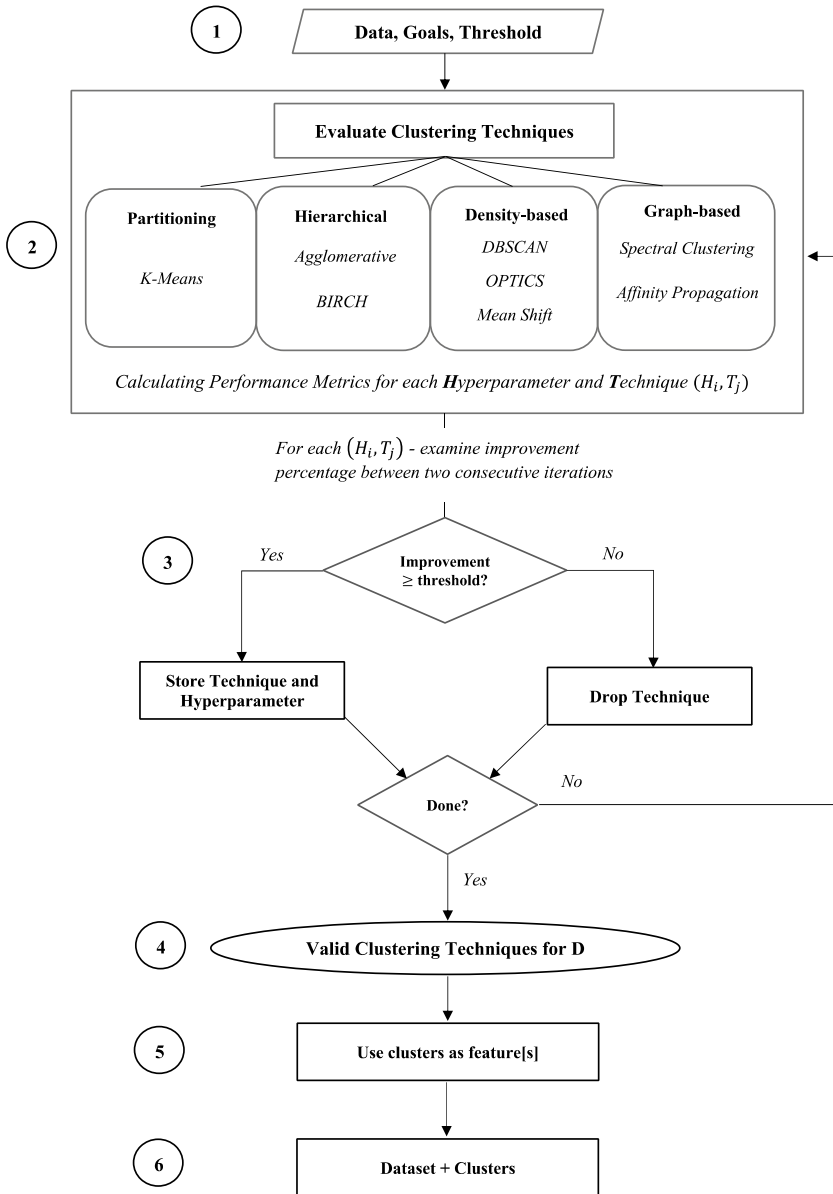


Fig. 2 The automated clustering procedure flow

3.2 *Experimental Procedure*

In the experiments, the predefined parameters were set as equal for all datasets. The following describes the setup for each experiment:

- **Datasets**—Each dataset was analyzed using both supervised and unsupervised methods. To illustrate the results for both learning types, the target variable in the unsupervised analysis was ignored. A labeled dataset was used, and the target variable was removed to compare our results to the original value.
- **Preprocessing**—Datasets were used consisting of numerical features and normalized using Z-score.
- **Parameters**—For all datasets, a threshold of 10% ($t = 0.1$) was used, meaning that an improvement of at least 10% between the iterations was required.
- **Figures**—For the presentation, the principal components analysis (PCA) was used to perform dimensionality reduction to 2D. Notably, this is not possible for every dataset, although here, all the features were numerical and PCA could be performed.

3.3 *Use Case: The Hepatitis C Dataset*

To simplify the demonstration, the Hepatitis C dataset was chosen, consisting of 13 features over 611 observations. We denoted the feature names as F_i for all $1 \leq i \leq 13$. The target variable was removed for the scenario demonstration and stored for future comparisons.

To begin, the algorithm evaluated all clustering techniques and hyperparameters (presented in Table 3) and calculated the four clustering performance metrics: Silhouette score, CH-score, DB index, and Dunn index. For each algorithm, the method examined the improvement (if it existed) in the percentage between two consecutive iterations. In this case, only four techniques passed the threshold condition out of a total of eight clustering techniques. Table 3 presents the four selected techniques and the differences between all of the measures given a hyperparameter.

For comparison, the classical K-means algorithm failed to optimize all four measures. Table 4 presents the K-means algorithm's learning process over different values of the number of clusters. As none of the rows satisfied TRUE in all measures, the K-means may not be the appropriate clustering technique for this case.

At this phase, the algorithm created four sub-datasets (one for each clustering technique) and labeled the records using the output clusters. Although the algorithm found four suitable clustering techniques for this dataset, not all of them necessarily satisfy the predefined goal. The following sections will describe the two main options for the goal.

Supervised Analysis. This section presents the case of the dataset that included a target variable and the ground truth labels for comparison. Here, given the original target and the output clusters, the algorithm calculated the mutual information (MI),

Table 3 Clustering techniques and their performance metrics after the threshold learning

Technique	Value	SIL	CH	DBI	DUNN
Affinity clustering (damping)	0.5	0	0	0	0
	0.55	0	0	0	0
	0.6	0	0	0	0
	0.65	0	0	0	0
	0.7	0	0	0	0
	0.75	0	0	0	0
	0.8	0	0	0	0
	0.85	0	0	1	0
	0.9	1	0	1	1
	0.95	0	0	0	0
	0.99	1	1	1	1
Mean shift (bandwidth)	10	0	0	0	0
	9	1	0	1	0
	8	0	1	1	1
	7	0	0	1	1
	6	1	1	1	1
	5	0	1	1	0
	4	1	1	1	1
	3	1	0	1	1
	2	0	1	0	1
OPTICS (min samples)	10	0	1	1	0
	9	1	1	1	0
	8	1	1	1	0
	7	1	1	1	1
	6	0	1	1	1
	5	1	1	1	0
	4	1	1	1	1
	3	1	1	1	0
	2	0	1	1	1
Spectral clustering (number of clusters)	2	1	1	1	0
	3	0	1	1	1
	4	0	0	1	0
	5	0	0	1	1
	6	1	1	1	1
	7	1	1	1	1
	8	0	1	1	0
	9	1	1	1	0

(continued)

Table 3 (continued)

Technique	Value	SIL	CH	DBI	DUNN
	10	1	1	1	0
	11	1	0	0	1
	12	1	1	1	0

Note This is a Boolean table in which 1 signifies that the threshold has been passed; otherwise, it is assigned as 0. The bolded cells are the selected hyperparameters by our method
Bold shows the selected hyperparameters by author’s method

Table 4 K-means performance metrics after the threshold learning

	SIL	CH	DBI	DUNN
<i>k</i> = 2	0	0	0	0
<i>k</i> = 3	0	0	0	0
<i>k</i> = 4	0	0	0	0
<i>k</i> = 5	0	0	1	1
<i>k</i> = 6	1	0	0	0
<i>k</i> = 7	0	0	0	0
<i>k</i> = 8	0	0	0	0
<i>k</i> = 9	0	0	0	1
<i>k</i> = 10	0	0	1	1
<i>k</i> = 11	0	0	0	0

homogeneity, V-measure (weighted score for homogeneity and completeness), and Fowlkes–Mallows index (FM index; describes the ratio between the precision and recall) for each clustering technique. Table 5 compares those measures between the four selected clustering techniques.

The final output of our method can be changed according to the goal. For example, if seeking to maximize the clusters’ homogeneity, the returned dataset is created by Affinity Clustering. The homogeneity measure ranges between zero to one and describes the percentage of each cluster only containing members of a single target [54]. When all samples in some clusters have the same label (target), the homogeneity

Table 5 Clustering performance metrics with the original target value

	Affinity clustering	Mean shift	OPTICS	Spectral clustering
MI	0.197	0.427	0.483	0.477
Homogeneity	0.451	0.345	0.388	0.390
V-measure	0.222	0.445	0.488	0.491
FM index	0.553	0.930	0.944	0.937

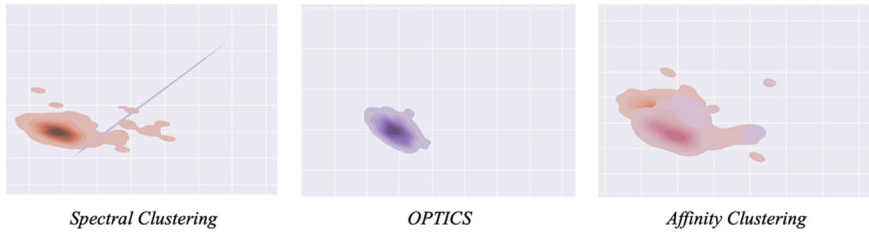


Fig. 3 The kernel density estimation (KDE) of the output clusters

is equal to one. Note that maximizing/minimizing a specific performance metric may not yield optimal results, although it would satisfy the predefined goal.

In another case, when the goal is to maximize the ratio between precision and recall, the algorithm returns using the OPTICS technique due to the maximality of the FM index, which is the geometric mean of the precision and recall measures:

$$FMI = \frac{TP}{\sqrt{(TP + FP) \cdot (TP + FN)}}$$

Figure 3 presents three sub-figures of the kernel density estimation (KDE) of the output clusters achieved by the different clustering techniques. It describes the challenge in estimating the appropriate clustering procedure. Indeed, the FM index has a maximum value. However, according to Fig. 3, this is due to the density of the data and the lack of difference between the resulting clusters. Thus, it is an example of uncertainty in unsupervised learning as there is no prior information about the data.

Unsupervised Analysis. A clustering algorithm’s performance cannot be measured as easily as supervised classification algorithms’ precision or recall. The clustering algorithm uses the performance metrics described in Sect. 2.2 and examines the values to satisfy the goal. This section presents the case of the dataset that does not include a target variable. Table 6 compares the clustering performance metrics that do not require an original target.

As mentioned earlier, optimizing a single metric can be helpful for a purposed goal but not necessarily for all. For example, if the goal is to maximize the CH-score, the method returns the dataset created by Mean Shift with a silhouette score of 0.591. If the silhouette score and the DB index were prioritized for optimization,

Table 6 A comparison of clustering performance metrics

	Affinity clustering	Mean shift	OPTICS	Spectral clustering
Silhouette score	0.139	0.591	0.530	0.461
CH-score	43.351	42.308	71.974	42.573
DB index	0.925	0.598	2.245	0.987

the algorithm returns the labeled dataset achieved by Mean Shift, which satisfies the maximal silhouette score and minimal DB index. Although, if the goal was to maximize both the DB index and the silhouette score, OPTICS presents a near optimal silhouette score with a maximal CH-score (almost double) compared to the other options. This case illustrates the uncertainty in clustering in the absence of the original target (labeling) and the different types of clustering that can be used to achieve different results.

4 Results

Four datasets (Sect. 3.1) were compared using the identified parameters to illustrate our method. For each dataset, we evaluated the automated clustering procedure with both supervised and unsupervised analyses. Furthermore, two additional goals were added to emphasize the diversity of our method: maximizing the Rand Index (RI) and maximizing the accuracy score of a supervised learning algorithm. The RI represents the ratio between the correct decisions and all decisions made by the algorithm. Table 7 presents the datasets, the predefined goal, and the selected clustering technique.

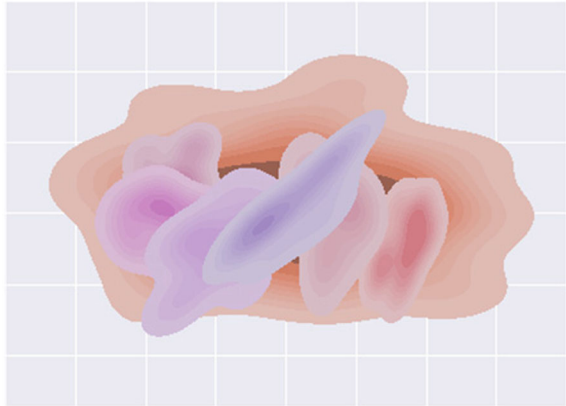
As presented in Table 7, for the heart failure dataset, both goals (maximizing the random index and minimizing the DB index) returned the OPTICS clustering algorithm. Figure 4 presents the KDE of the principal components of the heart failure dataset. The explained variance ratio was 81.19%, indicating that the PCA successfully preserved 81.19% of the original information; thus, the total information loss was 18.81%.

First, we concluded that the results were reliable and represented the original data distribution. Second, it can be seen (Fig. 4) that the clusters had a high and good separation rate. The range of values in the 2D figure was different due to the PCA's normalization and compression. However, the performance measures may be high due to the separation existing in higher dimensions.

Table 7 A comparison of the four datasets with different goals

Dataset	Goal	Clustering technique
Students	Maximize accuracy	Mean shift
	Maximize completeness	Spectral clustering
Fetal health	Maximize homogeneity	Affinity clustering
	Maximize CH-score	OPTICS
Heart failure	Maximize RI	OPTICS
	Minimize DB index	OPTICS
Hepatitis C	Maximize Homogeneity	Affinity Clustering
	Maximize Silhouette	Mean Shift

Fig. 4 The heart failure dataset clusters achieved by OPTICS



Analyzing the fetal health dataset yielded results that can address different issues. Since it is a medical dataset that aims to prevent maternal and infant mortality, it is essential to understand the meaning of the extreme values. Such an understanding could reveal a new insight or divide the data into groups to find features and relationships that could help in future predictions. Figure 5 compares the clusters achieved by Affinity Clustering (satisfying the supervised goal) and OPTICS.

When the goal was to create clusters such that the homogeneity was maximized, the algorithm chose Affinity Clustering with a homogeneity of 0.618. On the other hand, when the goal was to find a set of noisy observations, the algorithm chose OPTICS. It can be seen in Fig. 5 that additional clusters were obtained whose distribution differed from the gravity center distribution. In this way, it is possible to understand other patterns in the data that would not necessarily be obtained in traditional clustering.

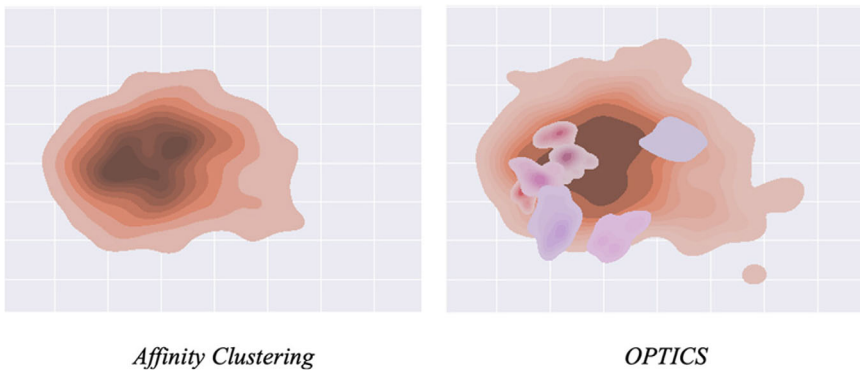


Fig. 5 The kernel density estimation (KDE) comparison between affinity clustering and OPTICS applied in the fetal health dataset

Table 8 A comparison of silhouette scores between k-means and our method

	K-means		Auto-clustering		
	Clusters	Silhouette score	Algorithm	Clusters	Silhouette score
Students	8	0.105	Mean shift	4	0.682
Fetal Health	7	0.152	Mean shift	4	0.412
Heart Failure	11	0.112	K-means	11	0.112
Hepatitis C	14	0.159	Spectral clustering	6	0.591

4.1 The Comparison to K-Means

K-means is the popular choice when a clustering algorithm is required. To estimate the number of clusters, it is required to examine the SSE for different values and choose the optimal number of clusters according to the elbow method (the point on the plot where distortion slows down). We repeated this process for each dataset (using the preprocessing described in Sect. 3.2) and calculated the silhouette score. Table 8 compares the number of clusters and the silhouette score between the traditional K-means and our method.

Table 8 highlights the importance of testing additional clustering algorithms besides K-means. Except for the heart failure dataset, all silhouette scores presented a significant increase and a decrease in the number of clusters using our method. However, in the heart failure dataset, our method estimated that the optimal clustering technique for this dataset was K-means. Thus, the measures did not change.

An extremely significant increase in the silhouette score was seen in the student’s dataset. For further examination, Fig. 6 presents both cases—the left figure shows the clusters found by K-means, and the right presents the Mean Shift clusters. First, in K-means, there are additional small clusters (with less than 3% of the data) that are further from the general distribution of the data, perhaps even defined as outliers. As a result, a silhouette score of 0.105 was yielded, indicating that the cluster division was controversial. On the other hand, the Mean Shift presented four well separated clusters with a silhouette score of 0.682 (549.5% improvement). Additionally, no clusters existed that were considered outliers or small clusters. This example reinforces the claim that K-means is not the ultimate choice for all cases and that the clustering algorithm must be adapted according to the distribution and shape of the data.

5 Conclusions and Discussion

This study proposed a goal-oriented clustering procedure based on the AutoML mechanism for selecting a clustering technique and its hyperparameters. It is based on threshold learning, ensuring that all performance measures are as close to optimal as possible. The main challenge in clustering appears in high-dimensional data (i.e.,

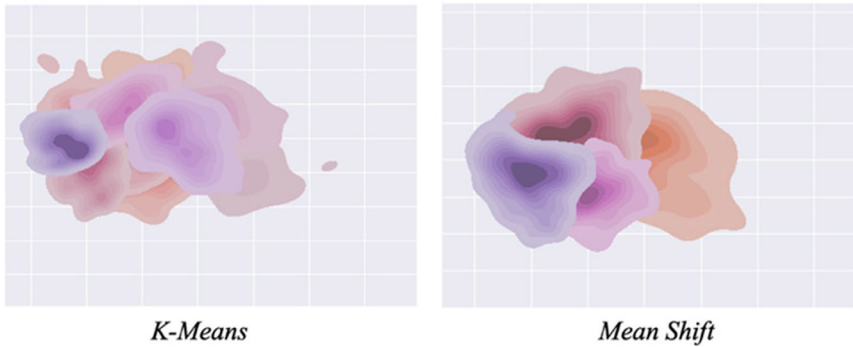


Fig. 6 Clusters comparison between two cases of the student's dataset

a large number of features and records) since the data distribution is hard to approximate. Due to the exponential growth of the number of possible values in this type of data, it is unclear what type of clustering technique is suitable for a given dataset. Moreover, a clustering algorithm's performance metrics are not as easily measurable as metrics for a supervised learning algorithm (classification or regression). Our method presents two main innovative aspects: automatic selection of clustering type (i.e., partition, hierarchical, density-based, or graph-based) without human intervention and the ability to determine the output dynamically according to a predefined goal.

Here are the main results:

1. Our method presented an extremely high increase in the silhouette score compared to the popular K-Means. The student's dataset achieved a silhouette score of 0.105 in K-Means and 0.682 in our method (which selected the Mean Shift technique). The fetal health dataset also presented a high increase in the silhouette score, from 0.152 to 0.412. Since it is a medical dataset, the new clusters may reveal a new insight or find relationships between features, which could help in future predictions. There was a single case (i.e., the heart failure dataset) in which our method chose K-means as the best fit clustering technique; thus, our results are the same.
2. Using different clustering techniques to analyze and explore meaningful conclusions is necessary. For example, in the case of the fetal health dataset, when the goal was to achieve as homogeneous clusters as possible, the Affinity algorithm produced an optimal result (Fig. 5). On the other hand, the OPTICS algorithm was used when the goal was to find noisy records in the data, thereby identifying patients with abnormal values. This conclusion reinforces the difficulty in performing clustering without defining a goal in advance.
3. Performing clustering based on multiple performance metrics produces more information than a single metric. For example, in the use-case demonstration (Sect. 3.3), the traditional K-means failed to optimize all four measures, although

the other four techniques succeeded. By comparing the results (Fig. 3), we concluded that clusters exist that can separate the dataset (the spectral clustering).

The quality of the data has a direct influence on the results, especially regarding unsupervised learning. An incorrect adjustment of these parameters or lack of data may lead to failure. This study presented two main issues that need to be addressed in future studies. First, we used a constant improvement threshold of 10%. Future studies ought to examine the optimal value of this threshold, which should be determined by preprocessing calculation to enhance its meaning and performance. Second, significant feature selection preprocessing can reduce the total complexity/running time. Furthermore, if the distribution of a given dataset is easy to estimate, clustering techniques that do not fit can be removed in advance. Approaches to these topics must be explored in future studies to further improve the presented method.

References

1. Barlow HB (1989) Unsupervised learning. *Neural Comput* 1(3):295–311. <https://doi.org/10.1162/neco.1989.1.3.295>
2. Sindhu Meena K, Suriya S (2019) A survey on supervised and unsupervised learning techniques. In: *International conference on artificial intelligence, smart grid and smart city applications*. Springer, Cham, pp 627–644. https://doi.org/10.1007/978-3-030-24051-6_58
3. Berry MW, Mohamed A, Yap BW (eds) (2019) *Supervised and unsupervised learning for data science*. Springer Nature, Cham
4. Elavarasi SA, Akilandeswari J, Sathiyabhama B (2011) A survey on partition clustering algorithms. *Int J Enterp Comput Bus Syst* 1(1):1–14
5. Patel S, Sihmar S, Jatain A (2015) A study of hierarchical clustering algorithms. In: *2015 2nd International conference on computing for sustainable global development (INDIACom)*. IEEE, New Delhi, pp 537–541
6. Ige AO, Noor MHM (2002) A survey on unsupervised learning for wearable sensor-based activity recognition. *Appl Soft Comput* 127:109363. <https://doi.org/10.1016/j.asoc.2022.109363>
7. Singh HV, Girdhar A, Dahiya S (2022) A literature survey based on DBSCAN algorithms. In: *2022 6th international conference on intelligent computing and control systems (ICICCS)*. IEEE, Madurai, pp 751–758. <https://doi.org/10.1109/ICICCS53718.2022.9788440>
8. Hazan H, Saunders D, Sanghavi DT, Siegelmann H, Kozma R (2018) Unsupervised learning with self-organizing spiking neural networks. In: *2018 international joint conference on neural networks*. IEEE, Rio de Janeiro, pp 1–6. <https://doi.org/10.1109/IJCNN.2018.8489673>
9. Liu Q, Mukhopadhyay S (2018) Unsupervised learning using pretrained CNN and associative memory bank. In: *2018 international joint conference on neural networks (IJCNN)*. IEEE, Rio de Janeiro, pp 1–8. <https://doi.org/10.1109/IJCNN.2018.8489408>
10. Nikbakht R, Jonsson A, Lozano A (2020) Unsupervised learning for parametric optimization. *IEEE Commun Lett* 25(3):678–681. <https://doi.org/10.1109/LCOMM.2020.3027981>
11. Serb A, Bill J, Khat A, Berdan R, Legenstein R, Prodromakis T (2016) Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses. *Nat Commun* 7(1):1–9. <https://doi.org/10.1038/ncomms12611>
12. Xie T, France-Lanord A, Wang Y, Shao-Horn Y, Grossman JC (2019) Graph dynamical networks for unsupervised learning of atomic scale dynamics in materials. *Nat Commun* 10(1):1–9. <https://doi.org/10.1038/s41467-019-10663-6>

13. Sagi O, Rokach L (2018) Ensemble learning: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 8(4):e1249. <https://doi.org/10.1002/widm.1249>
14. Dong X, Yu Z, Cao W, Shi Y, Ma Q (2020) A survey on ensemble learning. *Front Comput Sci* 14(2):241–258. <https://doi.org/10.1007/s11704-019-8208-z>
15. Rincy TN, Gupta R (2020) Ensemble learning techniques and its efficiency in machine learning: A survey. In: 2nd international conference on data, engineering and applications (IDEA). IEEE, Bhopal, pp 1–6. <https://doi.org/10.1109/IDEA49133.2020.9170675>
16. Kriegel HP, Kröger P, Zimek A (2012) Subspace clustering. *Wiley Interdiscip Rev Data Min Knowl Discov* 2(4):351–364. <https://doi.org/10.1002/widm.1057>
17. Agrawal R, Gehrke J, Gunopulos D, Raghavan P (2005) Automatic subspace clustering of high dimensional data. *Data Min Knowl Discov* 11(1):5–33. <https://doi.org/10.1007/s10618-005-1396-1>
18. Kailing K, Kriegel HP, Kröger P (2004) Density-connected subspace clustering for high-dimensional data. In: Proceedings of the 2004 SIAM international conference on data mining. SIAM, Lake Buena Vista, pp 246–256. <https://doi.org/10.1137/1.9781611972740.23>
19. Kingrani SK, Levene M, Zhang D (2018) Estimating the number of clusters using diversity. *Artif Intell Res* 7:15–22. <https://doi.org/10.5430/air.v7n1p15>
20. Shi C, Wei B, Wei S, Wang W, Liu H, Liu J (2021) A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *J Wireless Com Network* 31:1–16. <https://doi.org/10.1186/s13638-021-01910-w>
21. Ünlü R, Xanthopoulos P (2019) Estimating the number of clusters in a dataset via consensus clustering. *Expert Syst Appl* 125:33–39. <https://doi.org/10.1016/j.eswa.2019.01.074>
22. Wang X, Xu Y (2019) An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conf Ser: Mater Sci Eng* 569:052024. <https://doi.org/10.1088/1757-899X/569/5/052024>
23. Dunn JC (1974) A graph theoretic analysis of pattern classification via Tamura’s fuzzy relation. *IEEE Trans Syst Man Cybern SMC*-4(3):310–313. <https://doi.org/10.1109/TSMC.1974.5409141>
24. Davies DL, Bouldin DW (1979) A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell* 1:224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
25. Wu Y, Xi X, He J (2022) AFGSL: automatic feature generation based on graph structure learning. *Knowl Based Syst* 238:107835. <https://doi.org/10.1016/j.knsys.2021.107835>
26. Yao Q, Wang M, Chen Y, Dai W, Li YF, Tu WW, Yang Q, Yu Y (2018) Taking human out of learning applications: A survey on automated machine learning. *arXiv:1810.13306*. <https://doi.org/10.48550/arXiv.1810.13306>
27. He X, Zhao K, Chu X (2021) AutoML: A survey of the state-of-the-art. *Knowl Based Syst* 212:106622. <https://doi.org/10.1016/j.knsys.2020.106622>
28. Tschechlov D (2019) Analysis and transfer of AutoML concepts for clustering algorithms [Master’s thesis]. University of Stuttgart. <https://doi.org/10.18419/opus-10755>
29. Koren O, Hallin CA, Koren M, Issa AA (2022) AutoML classifier clustering procedure. *Int J Intell Syst* 37:4214–4232. <https://doi.org/10.1002/int.22718>
30. Koren O, Koren M, Peretz O (2022) Automated feature selection threshold-based learning for unsupervised learning [Manuscript submitted for publication]. *Int J Intell Manuf*
31. Koren O, Koren M, Peretz O (2022) AutoML Threshold learning for feature selection optimization [Manuscript submitted for publication].
32. Shi X, Wong YD, Chai C, Li MZF (2020) An automated machine learning (AutoML) method of risk prediction for decision-making of autonomous vehicles. *IEEE Trans Intell Transp Syst* 22(11):7145–7154. <https://doi.org/10.1109/TITS.2020.3002419>
33. Barbon Jr S, Ceravolo P, Damiani E, Tavares GM (2021) Selecting optimal trace clustering pipelines with AutoML. *arXiv:2109.00635*. <https://doi.org/10.48550/arXiv.2109.00635>
34. Lloyd SP (1982) Least squares quantization in PCM. *IEEE Trans Inf Theory* 28(2):129–137. <https://doi.org/10.1109/TIT.1982.1056489>
35. Nielsen F (2016) Hierarchical clustering. In: Nielsen F (ed) Introduction to HPC with MPI for data science. Springer, Cham, pp 195–211

36. Zhang T, Ramakrishnan R, Livny M (1997) BIRCH: A new data clustering algorithm and its applications. *Data Min Knowl Discov* 1(2):141–182. <https://doi.org/10.1023/A:1009783824328>
37. Ng A, Jordan M, Weiss Y (2001) On spectral clustering: Analysis and an algorithm. *Adv Neural Inf Process Syst* 14:1–8
38. Cheng Y (1995) Mean shift, mode seeking, and clustering. *IEEE Trans Pattern Anal Mach Intell* 17(8):790–799. <https://doi.org/10.1109/34.400568>
39. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24(5):603–619. <https://doi.org/10.1109/34.1000236>
40. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proceedings* 96(34):226–231
41. Kriegel HP, Kröger P, Sander J, Zimek A (2011) Density-based clustering. *Wiley Interdiscip Rev Data Min Knowl Discov* 1(3):231–240. <https://doi.org/10.1002/widm.30>
42. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. *Science* 315(5814):972–976. <https://doi.org/10.1126/science.1136800>
43. Wang K, Zhang J, Li D, Zhang X, Guo T (2008) Adaptive affinity propagation clustering. [arXiv:0805.1096](https://arxiv.org/abs/0805.1096). <https://doi.org/10.48550/arXiv.0805.1096>
44. Rahmah N, Sitanggang IS (2016) Determination of optimal epsilon (eps) value on DBSCAN algorithm to clustering data on peatland hotspots in Sumatra. *IOP Conf Ser: Earth Environ Sci* 31(1):012012. <https://doi.org/10.1088/1755-1315/31/1/012012>
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
46. Wang Y, Xu Y, Gao T (2021) Evaluation method of wind turbine group classification based on Calinski Harabasz. In: 2021 IEEE 5th conference on energy internet and energy system integration. IEEE, Taiyuan, pp 2630–2635. <https://doi.org/10.1109/EI252483.2021.9713300>
47. Morales F, García-Torres M, Velázquez G, Daumas-Ladouce F, Gardel-Sotomayor PE, Gómez-Vela F, Divina F, Vázquez Noguera JL, Sauer Ayala C, Pinto-Roa DP, Mello-Román JC, Becerra-Alonso D (2022) Analysis of electric energy consumption profiles using a machine learning approach: a Paraguayan case study. *Electronics* 11(2):267. <https://doi.org/10.3390/electronics11020267>
48. Sitompul BJD, Sitompul OS, Sihombing P (2019) Enhancement clustering evaluation result of Davies-Bouldin index with determining initial centroid of k-means algorithm. *J Phys Conf Ser* 1235:012015. <https://doi.org/10.1088/1742-6596/1235/1/012015>
49. Wijaya YA, Kurniady DA, Setyanto E, Tarihoran WS, Rusmana D, Rahim R (2021) Davies Bouldin index algorithm for optimizing clustering case studies mapping school facilities. *TEM J.* 10:1099–1103
50. Realinho V, Martins MV, Machado J, Baptista L (2021) Predict students' dropout and academic success. UCI Machine Learning Repository. <https://doi.org/10.24432/C5MC89>
51. Ayres-de-Campos D, Bernardes J, Garrido A, Marques-de-Sa J, Pereira-Leite L (2000) SisPorto 2.0: A program for automated analysis of cardiocograms. *J Matern Fetal Med* 9(5):311–318. <https://doi.org/10.3109/14767050009053454>
52. Dua D, Graff C (2019) UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>
53. Hepatitis. (1988) UCI Machine Learning Repository. <https://doi.org/10.24432/C5Q59J>
54. Rosenberg A, Hirschberg J (2007) V-measure: A conditional entropy-based external cluster evaluation measure. In: *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. pp 410–420.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Proposal for a New Separation Method for Reproducing Images with Properties in the Visible and Near-Infrared Spectrum



Jana Žiljak Gršić, Silvio Plehati, Tomislav Bogović, and Roko Vujić

Abstract Painting dyes and artworks exhibit duality with different states in the visible (V) and near-infrared (Z) spectrum. “V” denotes the solar spectrum ranging from 400 to 750 nm, while “Z” indicates the solar spectrum up to 1000 nm. The infrared state of an artwork can be an independent, “hidden” image. The paper demonstrates a new procedure of graphic preparation for the separation of process inks for reproducing artworks in monographs, with a focus on integration in the visible and near-infrared spectrum. The practical part of the paper uses the Projectina PAG B50 forensic scanner, which is equipped with 24 filters for blocking light from 220 to 1000 nm. Image reproduction involves photography of artwork in RGB, then conversion into CMY, with the addition of K from the Z image. “VZ” separation is introduced into the reproduction of the artwork, simulating the dual states of the original artwork. The GCR procedure is applied by subtracting CMY but based on information for the K component that is equal in the Z image. This process provides a deeper understanding of artworks and offers extended information about their state in the visible and near-infrared part of the spectrum.

Keywords VZ color separation · Reproduction of NIR images · NIR spectra of dyes · Hidden image · VIS/NIR-Z · GCR

J. Ž. Gršić (✉) · S. Plehati · T. Bogović
Zagreb University of Applied Sciences, Vrbik 8, 10000 Zagreb, Croatia
e-mail: jana@tvz.hr

S. Plehati
e-mail: splehati@tvz.hr

T. Bogović
e-mail: tbogovic@tvz.hr

R. Vujić
Faculty of Graphic Arts, University of Zagreb, Getaldićeva 2, 10000 Zagreb, Croatia
e-mail: rvujic@grf.hr

1 Introduction

Artworks have a visual (visible) state visible to human eyes and a near-infrared state visible with near-infrared (NIR—Near InfraRed) cameras. Each dye, and each image, has this “duality”. We introduce the concept of duality to expand the vocabulary in the works of dual photography, restoration projects, new interpretations of artworks, involving the coexistence of two spectral regions. The artist creates artwork with secrets embedded in a space invisible to the naked eye, and a hidden image becomes an integral part of the artwork. Therefore, an artwork today is viewed in both its visible and near-infrared states [1–3].

This paper demonstrates the importance of the near-infrared state of the image and an innovative method of reproducing artworks in monographs or web editions with properties in both the visible and near-infrared spectra. The near-infrared spectrum is crucial because it displays potential authorial imprints, the sequence of layers, and the connection of colors. The infrared and near-infrared spectrum is particularly important for art historians, forensic experts, and others interested in proving artworks’ authenticity [4–6].

Current practices in publishing art monographs reproduce artworks only for viewing in the visible part of the spectrum. We have developed a method that can reproduce an artwork using conventional printing techniques such as offset or digital printing, without significantly altering the printing process, as all works are done in software preparation [7–9].

We can present an artwork with properties in two spectral regions as an animation that transitions from view in the visible spectrum to the final view in the near-infrared spectrum.

During past decades, our method, Infraredesign, has been used to produce works intentionally expressing the artist’s vision in two spectral regions. The method of capturing and reproducing images for the visible and near-infrared spectrum is crucial for these artworks as it presents them in their entirety [10, 11].

According to other published articles, we show the original and reproduction here, its simulation with about ten light filters. We reveal, discover, and execute reproduction and equality in two spectral regions [12, 13].

The brushstroke has always been studied as a personalized signature and a means of confirming the author of the artwork. Marking an artwork with the NIR (Z) process has initiated new learning about dyes and new ways of blending dyes for the purpose of individualizing a painting [14].

The original work needs to be viewed instrumentally in the “V” and “Z” spectra. However, in case when we only have access to the reproduction or are prohibited from approaching the painting instrumentally in the museum, such dual viewing is impossible. This paper resolves this problem by developing a novel model of reproducing the image in its duality in monographic editions.

2 Artistic Dyes Under Light Blockades

Some dyes absorb NIR light. Regarding dye materials, there are no published data on how their components behave in the near-infrared spectrum. NIR (Z) cameras are recommended for the creation of hidden, dual images. Using an NIR camera, the painter can monitor their colors and brushstrokes, as well as the overall process of creating the image in the NIR spectrum. The painter may have made several corrections or started a multilayer execution of the image. It is possible that adjustments were made to the image, repairs carried out on certain parts of the picture, or that the artist abandoned the original idea and repainted the image. There is no recipe for the composition of store-bought dyes. We have selected eight artistic dyes and recorded them with the forensic device Projectina Documenter 4500 [15] (Fig. 1).

Yellow and red dyes do not absorb light above 645 nm. Other dyes, as well as cyan and blue, have the property of absorbing light in the visible part of the spectrum up to 715 nm (with reduced intensity beyond that value).

These properties are for store-bought set of dyes that are used in this work. Although these colors can be mixed, for example, with C, M, Y, dyes with Z values equal to zero value (Table 1).

Values given in Table 1 are information for painters on how to use dyes in the NIR-Z spectrum.

3 Artwork Duality in VIS/NIR-Z Spectra

Paintings exhibit duality as dyes manifest dual characteristics during the absorption of sunlight. Painters have researched the properties of dyes in the V and the Z spectra, because, depending on wavelength, surveillance cameras provide different information about the same object. Photographing the NIR state of the artwork has initiated work in IR restoration. The goal is to faithfully restore each artwork, ensuring that the colors used in restoration have identical or (very) similar properties to the original colors.

Photographing artworks with a PAG B50 camera allows for multiple states of the image at different wavelengths and enables a detailed approach to the analysis of the image.

In this study, a series of photographs were taken using four blockades (visible (V), 550, 665, and 850 nm). These photographs provide a multidimensional (multispectral) view of artwork, particularly emphasizing changes and characteristics resulting from the combination of the visible and near-infrared spectra. This deepens the understanding of artistic processes and opens up space for further analysis and interpretation.

We selected an artwork by academic painter Nada Žiljak, who started painting, drawing, using the “VZ” process about ten years ago (Fig. 2).

Fig. 1 **a** Eight acrylic dyes in the visible (V) part of the spectrum. **b** Eight acrylic dyes with light blockade at 645 nm. **c** Eight acrylic dyes with light blockade at 715 nm. **d** Eight acrylic dyes with light blockade at 850 nm (Z)

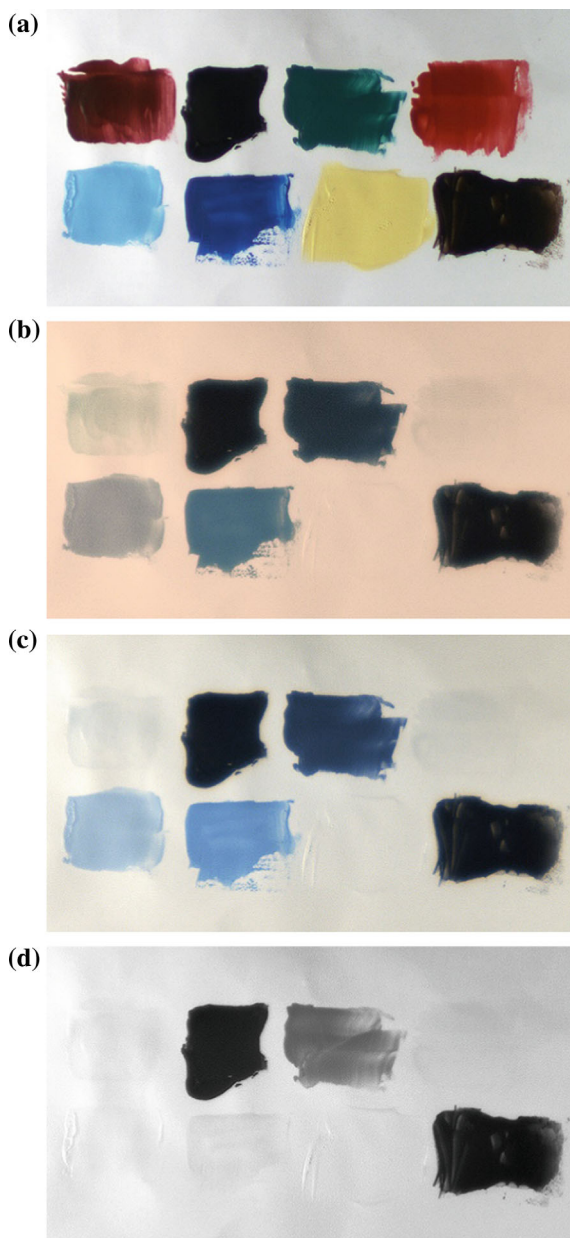


Table 1 Acrylic dyes (1–8), RGB values, L*a*b* visible, Z at 850 nm

Dye	RGB	L*a*b* visible	Z at 850 nm (%)
1. Permanent red violet, 576	51, 24, 20	12, 13, 9	3
2. Oxide black, 735	16, 17, 18	5, 0, - 1	90
3. Emerald green, 615	20, 54, 54	20, - 13, -4	64
4. Primary magenta, 369	97, 24, 16	22, 33, 25	0
5. Sky blue light, 551	114, 180, 218	70, -14, -26	0
6. Primary cyan, 572	21, 46, 95	19, 5, - 33	4
7. Nickel titan yellow, 274	206, 192, 108	77, - 3, 44	0
8. Vandyke brown, 403	16, 15, 14	4, 0, 1	90

Author's signature

In painting, there are no standards of the signing and labeling of images. Author's signatures, as well as other markings, are freely executed. Many signs appear in the paintings, from letters to small graphically shaped strokes. Today, the author is aware that they determine the recognition of dyes with two different instruments: for the visible part of the spectrum (V) and the near-infrared part of the spectrum (Z—NIR).

In the selected artwork, the signature is executed with black ink of type "S" (a mixture of equal parts magenta, yellow, and cyan). When signature is observed with the "NIR-Z camera", the camera does not register any of its components. In monographs of artworks, we find different solutions from page to page. Some signatures are executed in blue tones, some in black, some in green. Why is that? Only the authors of the images know. In our example, the author has developed a series of secrets about mixing dyes concerning their appearance in the V and Z spectra.

Animation of Nada Žiljak artwork with properties in two spectral regions is shown at: <https://nada.ziljak.hr/245.mp4>.

4 CMYK Reproduction and Reproduction in Art Monographs

Painters use a wide range of different colors and dyes. By mixing C, M, Y, K dyes, a simulation of colors in all these tones is achieved. This is sufficient for the reproduction of an artistic work in the visible and near-infrared spectrum. Colors and CMYK dyes are called "process colors" and are most commonly used for reproduction through conventional printing processes.

By mixing C, M, Y dyes, all color tones detected and distinguished by our eyes (ranging from 700 to 750 nm) are simulated. By mixing C, M, Y (in equal quantities), a black dye called "S" is achieved. However, this dye does not absorb light in the Z spectrum (range from 750 to 1000 nm) because it consists of individual dyes that do not absorb Z light on their own.

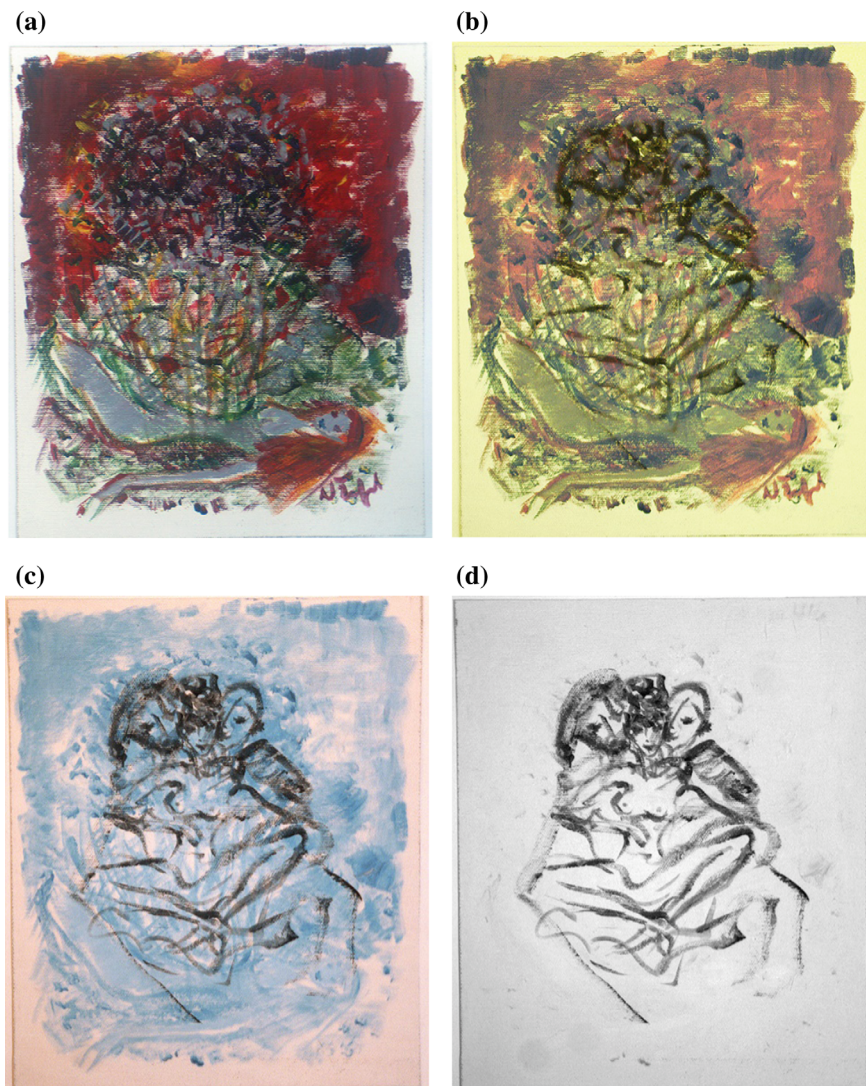


Fig. 2 a Artwork shown in the visible part of the spectrum. b Artwork shown with light blockade at 555 nm. c Artwork shown with light blockade at 665 nm. d Artwork shown with light blockade at 850 nm

Another black dye, as an integral part of the “process colors” set, is commonly called K or “Carbon black dye”. K is a dye that absorbs infrared light in the range from 750 to 1000 nm. It is precisely because of the K dye that the subject of this work is the printing technique Infrared Reproduction (IRR) or, briefly, “VZ”.

The channels C, M, Y, K are shown for reproduction in a monochromatic display Fig. 3a–d by subtracting the values of C, M, Y from the Z image for the K channel.

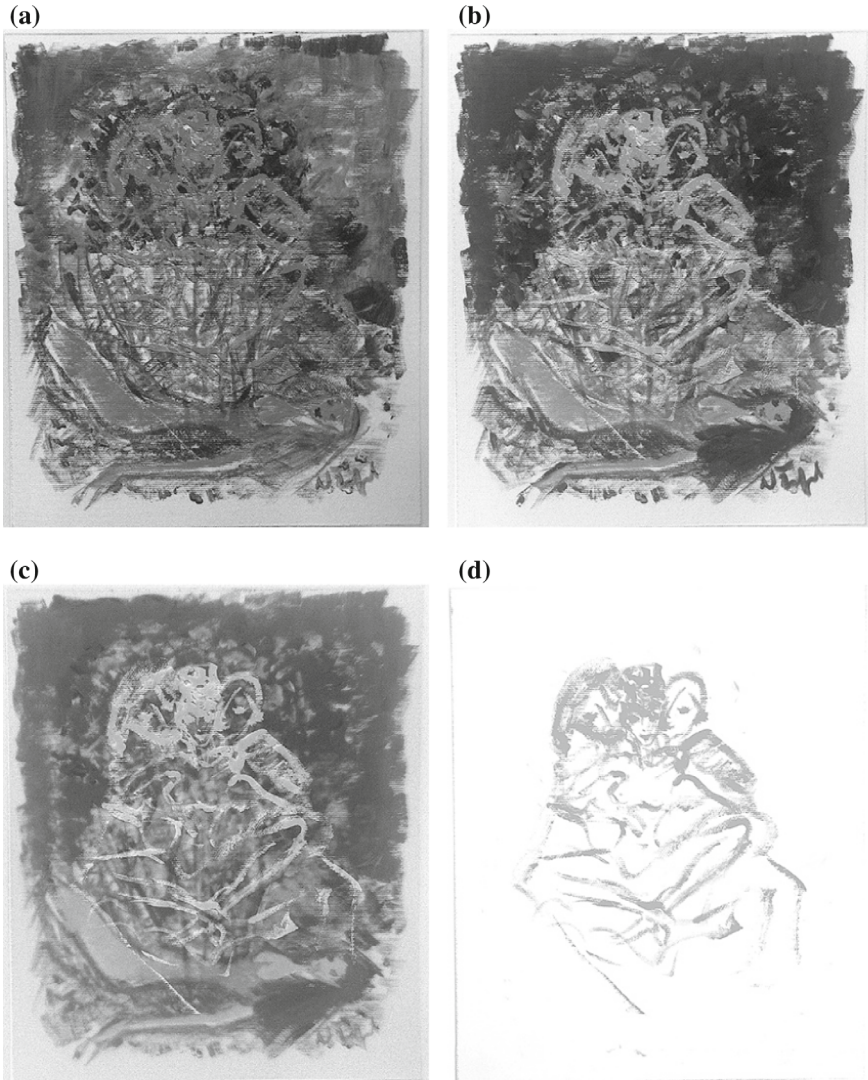


Fig. 3 a Cyan separation for reproduction. b Magenta separation for reproduction. c Yellow separation for reproduction. d K (carbon black) separation for reproduction

The condition of the images is determined in two spectra: visible and near-infrared. This expands the theme of the originality of the artwork. We demonstrate and propose how such image can be reproduced in two spectral areas in monographs, books, and textbooks on visual arts.

The printed image will show only the visual spectrum (visible part of the spectrum), and when viewed with an infrared camera, the image in the near-infrared part of the spectrum will become visible. Seeking and discovering are new activities

for both experts and ordinary observers, all aimed at uncovering new information, whether hidden or intentional.

In conventional printing, printing inks called “process inks” include “carbon black ink” marked with K. In conventional separation procedure gradations of K participation are offered with the names: none, light, medium, heavy, maximum according to the GCR procedure, replacing C, M, Y with K. Fortunately, K ink absorbs NIR light. We have developed the “VZ” method so that adding K (and replacing C, M, Y) does not result merely in ink savings through the GCR process. More importantly, this ink will carry information from the Z image. The CMYK image with Z information is employed for conventional printing to reproduce artwork in the monograph (Fig. 4).

Surrounded by surveillance cameras that register the Z spectrum, we suggest using the properties of printing inks for dual interpretation of the artistic image in both the visible and NIR spectra.

VZ separation performs the K value, obtained from the Z image. Our GCR works (subtracts) based on the Z value of the image, which does not exist as a programming tool in today’s separation software. Our image reproduction process is oriented toward the state of duality, just like the original. The K channel takes on a new significance. It represents the Z value. The remaining CMY channels are adjusted according to the Z value of the image using conventional GCR procedures. Nevertheless, the objective is to ensure that the reproduction in the visible spectrum closely resembles the visual perception of the original artwork. In cases where the Z value

Fig. 4 CMYK image with Z information for reproduction in art monograph



from the original image cannot be fully realized, a compromise is made by adjusting the interpretation of the Z image.

5 Conclusion

The article introduces reproducing artworks with an extension to the near-infrared (NIR) light spectrum. Museums are obliged to photograph protected artworks in both the visible and near-infrared parts of the spectrum. This is because each set of dyes has colors with more or less properties of NIR light absorption. Human eyes do not perceive this, but when viewed instrumentally, for example, with surveillance cameras, we gain the ability to separate and isolate only the specified wavelength.

Based on previous research on the authenticity of artworks, it is concluded that the visible (V) and near-infrared (Z) spectral states of the image represent key points in the analysis of historical and contemporary unintentional/intentional IR painting. Special attention is given to series of photographs through light blockades, emphasizing intentional manipulation of the visible/NIR spectrum. Studies on NIR components of artworks have been a privilege of a narrow number of researchers because reports and printed monographs presented only the visual part of the spectrum.

We are opening a new chapter in the reproduction of artworks. Based on the knowledge that process inks from the field of printing also have light absorption properties in the NIR spectrum, the idea emerged to include the state of its hidden NIR component in the reproduction of the image. Reproducing artworks through process inks from the printing field implies high circulation. The widespread availability of monographic editions also opens a new chapter in general “VZ” analyses of both old and new paintings.

By photographing V and Z in today’s procedures for archiving images of artworks, information is obtained that is not available by observing the image with the naked eye alone. New knowledge about the content of images observed in the V and NIR spectra has initiated new restoration work, which would include the state of the image in the visible and near-infrared spectra. The goal is to faithfully restore each artwork and ensure that the colors used have the same or very similar properties to the dyes the author used when creating their work. Observing images in museums with NIR cameras makes us aware that the image has significantly changed after restoration. NIR photography in museums is not welcomed. Galleries will not allow such shootings. A big question remains: How will restorers deal with an intentionally painted dual image, or a hidden image such as the artwork of Nada Žiljak shown in this paper?

References

1. Žiljak Gršič J, Tepeš Golubić L, Stanačev Bajzek S (2022) Hidden fine art with IR colorants In: 9th SWS international scientific conferences on ART and HUMANITIES - ISCAH Proceedings 2022, 9, pp 89–96. SGEM World Science, Vienna. <https://doi.org/10.35603/sws.iscah.2022/s08.10>, <https://nada.ziljak.hr/n196.mp4>
2. Žiljak Gršič J, Plehati S, Žiljak Stanimirović I, Bogović T (2023) Properties of dyes for painting with spectroscopy in the visible and near infrared range. *Appl Sci* 13:2483. <https://doi.org/10.3390/app13042483>
3. Li C, Wang C, Wang SJ (2013) A black generation method for black ink hiding infrared security image. *Appl Mech Mater* 262:9–12. Trans Tech Publications, Baech SZ (2013). <https://doi.org/10.4028/www.scientific.net/AMM.262.9>
4. Žiljak Gršič, J., Jurečić, D., Morić Kolarić, B., Jeličić, T.: The Technique of Security Print on Textiles with a Hidden Sign in the Near-Infrared Spectrum. *Journal Tehnički vjesnik – Technical Gazette*, 27, 2, pp. 633–637. (2020). <https://doi.org/10.17559/TV-20191001173959>, <https://hrcak.srce.hr/file/343936>
5. Žiljak Gršič J, Tepeš Golubić L, Jurečić D, Matas M (2019) Hidden information in paintings that manifests itself in the near infrared spectrum. *Informatologia* 52, 1–2, pp 9–16. <https://doi.org/10.32914/i.52.1-2.2>. <https://hrcak.srce.hr/file/325245>
6. Wang CY, Li C, Huo LJ (2012) A security printing method by black ink hiding infrared Image. *Appl Mech Mater* 200:730–733. <https://doi.org/10.4028/www.scientific.net/amm.200.730>
7. Žiljak Gršič J, Politis A, Jurečić D (2020) Dye spectroscopy in the visual and near infrared spectrum. *Polytechnic Des* 8(1):29–37. <https://doi.org/10.19279/TVZ.PD.2020-8-1-14>. <https://polytechnicanddesign.tvz.hr/index.php/ojs/article/view/328/300>
8. Jurečić D, Žiljak V, Kudumović M, Kelčec Pester B (2018) Packaging with dual information for visual and infrared spectrum. *Acta Graphica* 29(2):41–46. ISSN 0353-4707, e-ISSN 1848-3828, <https://hrcak.srce.hr/file/314843>
9. Shin H, Reyes NH, Barczak AL, Chan CS (2010) Colour object classification using the fusion of visible and near-infrared spectra. In: Zhang BT, Orgun MA (eds) *PRICAI 2010: trends in artificial intelligence. PRICAI 2010. Lecture Notes in Computer Science*, 6230, pp 498–509, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15246-7_46
10. Žiljak V, Tepeš Golubić L, Žiljak Gršič J, Jurečić D (2018) Security methods in the art and science using spectroscopy in the visual and near infrared range. *Int J New Technol Res (IJNTR)*, 4(8):48–52. <https://doi.org/10.31871/IJNTR.4.8.14>, https://www.ijntr.org/download_data/IJNTR04080014.pdf
11. Žiljak Gršič J (2019) Near infrared spectroscopy of colorants in security design of postage stamps. *Polytech Des* 7(3):195–198. <https://doi.org/10.19279/TVZ.PD.2019-7-3-17>. <https://polytechnicanddesign.tvz.hr/index.php/ojs/article/view/272/254>
12. Xiaohe C, Liuping F, Peng C, Jianhua H, Jianle Z (2017) Research on anti-counterfeiting technology of print image based on the metameric properties. In: *Proceedings of the 2017 2nd international conference on communication and information systems*, pp 284–289. Association for Computing Machinery, New York, NY, USA (2017). <https://doi.org/10.1145/3158233.3159358>
13. Nazor Čorda D, Žiljak Gršič J (2021) Extended method of restoration retouching in the near-infrared spectrum. In: 8th SWS international scientific conference on social sciences—ISCSS 2021, section media & communications, pp 667–680. SGEM World Science, Vienna. <https://doi.org/10.35603/sws.iscss.va2021/s10.65>
14. Žiljak V, Žiljak Gršič J, Jurečić D, Jeličić T (2019) Near-infrared spectroscopy and hidden graphics applied in printing security documents in the offset technique. *Tehnički glasnik/ Technical J* 13(4):311–314. <https://doi.org/10.31803/tg-20191004140247>
15. Projectina Docucenter 4500 with SP-2000 color spectroscopy module, PAG B50 custom designed camera with 24 barrier filters, <https://www.projectina.ch>, last accessed 2023/11/19, http://www.telectronics.biz/assets/mainmenu/104/editor/PDF_leaflet_Docucenter_4500.pdf. Last accessed 19 Nov 2023

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Epidemic Information Extraction for Event-Based Surveillance Using Large Language Models



Sergio Consoli, Peter Markov, Nikolaos I. Stilianakis, Lorenzo Bertolini, Antonio Puertas Gallardo, and Mario Ceresa

Abstract This paper presents a novel approach to epidemic surveillance, leveraging the power of artificial intelligence and large language models (LLMs) for effective interpretation of unstructured big data sources like the popular ProMED and WHO Disease Outbreak News. We explore several LLMs, evaluating their capabilities in extracting valuable epidemic information. We further enhance the capabilities of the LLMs using in-context learning and test the performance of an ensemble model incorporating multiple open-source LLMs. The findings indicate that LLMs can significantly enhance the accuracy and timeliness of epidemic modelling and forecasting, offering a promising tool for managing future pandemic events

Keywords Health informatics · Epidemiology · Event-based surveillance · Natural language processing · Large language models

S. Consoli (✉) · P. Markov · N. I. Stilianakis · L. Bertolini · A. P. Gallardo · M. Ceresa
European Commission, Joint Research Centre (JRC), Ispra, Italy
e-mail: Sergio.Consoli@ec.europa.eu

P. Markov
e-mail: Peter.Markov@ec.europa.eu

N. I. Stilianakis
e-mail: Nikolaos.Stilianakis@ec.europa.eu

L. Bertolini
e-mail: Lorenzo.Bertolini@ec.europa.eu

A. P. Gallardo
e-mail: Antonio.Puertas@ec.europa.eu

M. Ceresa
e-mail: Mario.Ceresa@ec.europa.eu

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_17

1 Introduction and Background

Epidemic modelling is a challenging task as emphasized by numerous studies [15]. The use of novel unstructured datasets has been increasingly encouraged, as they can significantly contribute to improving early warning systems [11]. The recent COVID-19 pandemic has underscored the importance of these novel approaches [8]. The pandemic revealed the long delays in the release of official statistics, emphasizing the need to track pandemic information in alternative ways and at a higher frequency.

These opportunities and challenges in infectious disease epidemiology are inspiring some of the research activities at the Digital Health unit of the Joint Research Centre (JRC)¹ of the European Commission (EC). Ongoing work, described in this paper, aims at tracking pandemic outbreaks in the European Union (EU) using datasets which are considered unconventional in classic epidemiological modelling. The project aims at exploring novel (big) data sources to provide a better response to epidemics. It links to other ongoing relevant initiatives on the subject.² An important source of such information is represented by moderated news and reports, since they discuss important events and experts opinions, and can substantively inform policy-making decisions [6]. However, translating this new source of data into valuable information is challenging, given that the data derived from these sources are often unstructured and large. This data also exhibits nonlinear relationships across variables, which adds to the complexity of its interpretation. Despite these challenges, the potential benefits of utilizing these unconventional data sources are considerable. By effectively translating this data into actionable insights, we can significantly improve our understanding of epidemics and, consequently, our response to them.

In this paper, we aim to leverage the vast capabilities of artificial intelligence (AI) [3], specifically the power of generative pre-trained large language models (LLMs) [2, 14], for the exploration and effective interpretation of such innovative big data sources, with the ultimate goal of improving epidemic response. LLMs are a type of generative AI models that utilize deep-learning (DL) algorithms (involving billions of parameters) to calculate the likelihoods of word sequences. These probabilities are determined based on substantial text corpora that the model has previously learned from. Significant advancements to LLMs have been made with the advent of the transformers architecture [14]. Transformers designed to handle sequential data by using a mechanism called attention, which allows the model to weigh and prioritize different parts of the input data [12]. Unlike traditional sequential models such as recurrent neural networks (RNNs), transformers process all input data concurrently, which allows for more efficient computation and the ability to handle longer sequences.

¹ https://ec.europa.eu/info/departments/joint-research-centre_en.

² WHO EIOS: Epidemic Intelligence from Open Sources, <https://www.who.int/initiatives/eios>.

Considering the rapid advancements in LLMs, this study aims to evaluate the most significant and recent open-source and commercial models for the specific task of epidemic information extraction. The use of the information extracted from large-scale, unstructured datasets in epidemic infectious disease, coupled with the integration into surveillance systems, can significantly enhance the accuracy and timeliness of epidemic outbreaks modelling and forecasting.

2 Data

In this section, we describe the unstructured datasets used in the study. Although we have focused on these two datasets, the approach is completely generalizable to any textual source concerning infectious disease epidemics.

2.1 *ProMED*

ProMED³, the Program for Monitoring Emerging Diseases, is an initiative by the International Society for Infectious Diseases (ISID). As the largest publicly accessible system for infectious disease outbreak reporting, it serves as a critical resource for healthcare professionals and the public alike. Its platform offers a wealth of daily posts detailing the latest developments in the field of infectious diseases. Launched in 1994, ProMED has been at the forefront of outbreak reporting, being the first to report on numerous disease outbreaks, including SARS, Chikungunya, Ebola, Zika, MERS and, most recently, COVID-19 [7]. Its users comprise international public health leaders, government officials, physicians, veterinarians, researchers, companies, journalists and the general public worldwide. Reports are produced, and commentary is provided by a multidisciplinary global team of subject matter experts in a variety of fields, including virology, parasitology, epidemiology and entomology. Since its inception, ProMED has generated so far more than 66,000 mail posts, from 1994-08-19 to date.

2.2 *WHO Disease Outbreaks News*

The World Health Organization (WHO)'s Disease Outbreak News (DONs)⁴ serve as a vital source of information on confirmed acute public health events or potential events of concern. The platform has been providing crucial updates since January 1996. The essence of DONs lies in its commitment to sharing news about confirmed

³ <https://promedmail.org/>.

⁴ <https://www.who.int/emergencies/disease-outbreak-news/>.

or potential public health events. This includes incidents of unknown cause that carry significant or potential international health concerns and could affect international travel or trade. It also encompasses diseases of known cause which have shown the capacity to cause a serious public health impact and spread internationally. Since its inception, DONs has published over 3000 pieces of curated epidemic news, from 1996-01-22 to date.

3 Methods

In this section, we describe the epidemic information extraction models considered in our study. These consist of the most popular and recent LLMs, spanning from open-source to commercial ones, along with a significant semantic method from the literature for epidemic information extraction, namely EpiTator. Please note that the LLMs employed in this study have been used through the GPT@JRC initiative of the Joint Research Centre (JRC) of the European Commission, which enables JRC staff to explore the potential uses of AI pre-trained LLMs. The initiative, which is part of a broader study on the new technology's applications within the European Commission, is a central hub offering secure access to various AI models. GPT@JRC is hosted at the JRC datacentre and supports both open-source AI models, deployed on-premises at the JRC Big Data Analytics Platform⁵ and commercial OpenAI's GPT models running in the European Cloud under a Commission contract with an opt-out clause on prompt analysis by third parties.

3.1 *EpiTator*

EpiTator⁶ is a comprehensive open-source epidemiological annotation tool, specifically designed for the extraction of epidemiological information from texts [1]. This tool is essentially a Python script that leverages the powerful natural language processing (NLP) spaCy library⁷ to extract critical named entities that are of particular interest in epidemiology applications. These entities include diseases, locations, dates and counts. One of the remarkable features of EpiTator is the *Resolved Keyword Annotator*. It utilizes an SQLite database of entities, which ensures that multiple synonyms for an entity are resolved to a single id, thereby enhancing the accuracy and consistency of the extracted data. EpiTator also imports information about infectious diseases and animal species from several reliable sources, including Disease Ontology, Wikidata and the Integrated Taxonomic Information System (ITIS). The *Count Annotator* feature of EpiTator extracts and parses count values, identifying

⁵ <https://jeodpp.jrc.ec.europa.eu/bdap/>.

⁶ <https://github.com/ecohealthalliance/EpiTator>.

⁷ <https://spacy.io/>.

attributes such as whether the count refers to cases or deaths, or if the value is approximate. This provides a more nuanced understanding of the data. Instead, the *Date Annotator* feature identifies and parses dates and date ranges. EpiTator employs a key entity filtering process that condenses the output to a single entity per class that best describes the corresponding event. Despite EpiTator's capability to return all entities of an entity class (e.g. disease) found in a text, this filtering process is necessary to distil the most pertinent information. It achieves this using the most-frequent approach filtering, which identifies the key entity from all the entities by selecting the most frequently mentioned one per class.

3.2 *Pythia-12b*

The Pythia-12b model is a state-of-the-art transformer-based LLM that follows the architecture of the highly popular GPT3 model. This open-source model is owned by EleutherAI⁸, a collective of AI specialists aiming at advancing AI in an open and collaborative manner. Pythia-12b has been primarily designed to facilitate research on the behaviour, functionality and limitations of LLMs. The model operates within a context length of 4096 tokens and consists of a whopping 12 billion parameters.

3.3 *Mpt-30b-chat*

The Mpt-30b-chat model is a state-of-the-art, open-source LLM owned by MosaicML.⁹ One of the key features is its specialization in chat and dialogue generation. It has been fine-tuned to handle conversational dynamics, making it an excellent choice for creating interactive AI applications. It was fine-tuned on several datasets including ShareGPT-Vicuna, Camel-AI, GPTeacher, Guanaco, Baize and some more generated datasets. In terms of technical specifics, the Mpt-30b-chat model boasts a context length of 8192 tokens and is built with an impressive 30 billion parameters. This vast number of parameters allows the model to capture and learn from a wide variety of linguistic nuances, thereby greatly enhancing its conversational capabilities and predictive performance.

⁸ <https://huggingface.co/EleutherAI/pythia-12b>.

⁹ <https://huggingface.co/mosaicml/mpt-30b-chat>.

3.4 *Llama-2-70b-chat*

The Llama-2-70b-chat model [13] is an advanced open-source LLM owned by Meta.¹⁰ It is one of the most powerful open LLMs currently available. The base model, Llama 2, was pre-trained on a diverse range of publicly available online data sources. The fine-tuned version of this model, known as Llama-2-70b-chat, further enhances this understanding by leveraging publicly available instruction datasets and over 1 million human annotations. This extensive fine-tuning process ensures a high level of precision and adaptability in its responses. In terms of its technical specifications, the Llama-2-70b-chat model operates with a context length of 4096 tokens. Moreover, the model is built with an impressive 70 billion parameters.

3.5 *Mistral-7b-openorca*

The Mistral-7b-openorca model¹¹ is an open-source model, owned and fine-tuned by OpenOrca using its datasets¹² on top of the Mistral LLM [9]. Despite its smaller size with 7 billion parameters, the model is fast at inference and delivers robust performance across a wide range of tasks. With a context length of 4096 tokens, the Mistral-7b-openorca is one of the best-performing open-source models available for models under 30 billion parameters. This makes it a strong candidate for many API use-cases due to its combination of size, speed and performance.

3.6 *Zephyr-7b-alpha*

The Zephyr-7b-alpha model is an open-source model owned by HuggingFace.¹³ This model has been fine-tuned by HuggingFace using a combination of publicly available and synthetic datasets on top of the Mistral LLM. Despite its small size with only 7 billion parameters, it often outperforms the larger Llama 2 13B model, demonstrating performance comparable to several models in the 20–30B range.

3.6.1 *Gpt-3.5-turbo-16k*

The Gpt-35-turbo-16k model is a robust commercial model owned by OpenAI. This model is an advanced version of OpenAI's GPT-35-turbo, colloquially known as ChatGPT, but with four times the context. The versatility of this model is apparent in

¹⁰ <https://ai.meta.com/llama/>.

¹¹ <https://huggingface.co/Open-Orca/Mistral-7B-OpenOrca>.

¹² <https://huggingface.co/datasets/Open-Orca/OpenOrca>.

¹³ <https://huggingface.co/HuggingFaceH4/zephyr-7b-alpha>.

its ability to perform a wide range of tasks. With its context length of 16,384 tokens, the Gpt-35-turbo-16k model offers an expansive reach for various applications. However, being a commercial model, its extensive usage comes with a considerable cost and some usage constraints.

3.7 *Gpt-4-32k*

The Gpt-4-32k model, another commercial offering from OpenAI, is a powerful tool which sets itself apart in terms of its problem-solving capabilities. It can solve difficult problems with a greater level of accuracy than any of OpenAI's previous models, leading to its reputation as the best LLM available for any task. While it may come at a higher cost, the Gpt-4-32k model's superior performance and its ability to manage approximately 80 pages of text make it one of the best LLM options available. Its context length of 32,768 tokens provides an even greater scope for processing and managing large amounts of data, making it a perfect option to handle large texts.

3.8 *Gpt-4-FewShots*

The Gpt-4-FewShots is a variant to the Gpt-4-32k model obtained by passing three examples (three shots) of epidemic information extraction in the context to the Gpt-4-32k prompt, such that to try to specialize the model to handle the specific epidemic information extraction task. In-context learning (ICL) [2] refers to a learning approach where an AI model learns from its context. LLMs have already shown an ability for ICL as they scale in terms of model and corpus sizes [5]. The fundamental principle of in-context learning is learning through analogy. Initially, ICL needs a handful of examples to create a demonstration context, usually framed in natural language. Following this, ICL combines a query question with a demonstration context to create a prompt. This prompt is then input into the language model for prediction. We have leveraged the ICL approach on the Gpt-4-32k model¹⁴, leading to Gpt-4-FewShots, to allow for a more efficient and accurate way to extract and process information on epidemics.

3.9 *Open-Ensemble*

The Open-Ensemble model is an AI approach that leverages the power of multiple high-performing open-source models to generate outputs. It uses a majority voting system to determine the best result from the contributed models. In contrast to ordi-

¹⁴ Please note that ICL was only applied to Gpt-4-32k due its large context capability, which was not amenable in the other studied LLMs which are characterized by a way more lower context.

nary machine learning approaches which try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use [10].

The models included in this ensemble are Llama-2-70b-chat, Mistral-7b-openorca and Zephyr-7b-alpha, that, as it will be shown next, resulted to be the best performing open-source LLMs for epidemic IE. Each of these models brings unique strengths and capabilities to the ensemble, enhancing its overall performance. This Open-Ensemble model, with its majority voting mechanism, ensures that the most agreed-upon output from these three models is chosen, thereby increasing the likelihood of accuracy and reliability in its predictions. We have used this approach to test whether the obtained ensemble model improves over the single open-source LLMs and is able to compete with the performance of the more advanced commercial GPT models.

4 Computational Experiments

In this section, we delve into the application of LLMs for the purpose of epidemic information extraction. The extracted entities include the name of the disease, the country where the cases were reported, the confirmed case count and the date of the case count. We compare the performance of the different LLMs for infectious disease epidemic information extraction, along with EpiTator. The comparison has been made over a subset of the Incident Database (IDB), a repository developed for the purpose of event-based surveillance.¹⁵ The IDB is structured to house a comprehensive collection of data related to epidemic events, making it an important resource for researchers and healthcare professionals. The database includes a sample of ProMED and WHO DONs that have been meticulously annotated by domain experts. For the purpose of comparing the performance of the tested epidemic information extraction models, a gold standard subset of the IDB comprising 171 carefully selected samples has been selected. This subset provides a benchmark against which the performance of the models have been assessed.

We have evaluated the information extraction task as a binary classification problem where the negative class means that no information (“None”) is associated with the IDB data sample, while the positive class indicates that the IDB is labelled with a specific epidemic information. The models have been tested on their ability to recognize such positive and negative classes by considering the following widely adopted precision, recall and F_1 -score metrics [4]: Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$, $F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$, where TP, FP, FN and TN are, respectively, the number of True Positives, the number of False Positives, the number of False Negatives and the number of True Negatives [4]. In other words, the precision is the accuracy for the positive class predictions. The recall (sensitivity) is the ratio of the positive class instances that are correctly detected as such by the classifier. The F_1 -score is the harmonic mean of precision and recall, whereas the regular mean treats all

¹⁵ <https://github.com/aauss/EventEpi>.

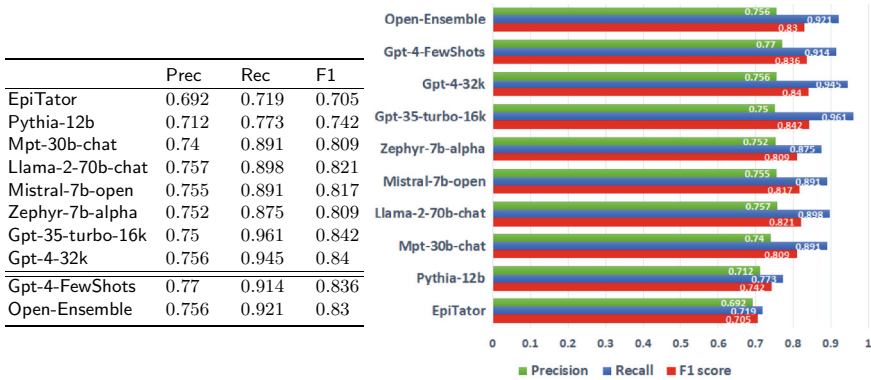


Fig. 1 Comparison of the models for the extraction of the pandemic name

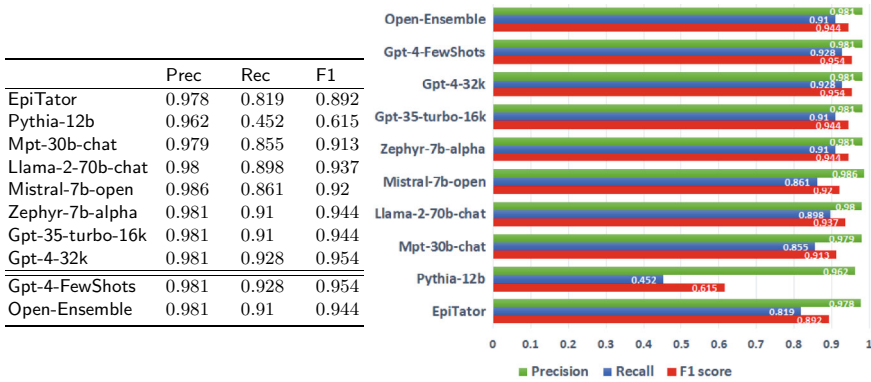


Fig. 2 Comparison of the models for the extraction of the country name

values equally, the harmonic one gives much weight to low values and for this reason is a metric preferred to the classical accuracy in an imbalanced problem setting.

Our computational results are reported in Figs. 1, 2, 3 and 4, which report the comparison of the different algorithms for the extraction, respectively, of the pandemic name, the country where this virus outbreak occurred, the related date and the number of cases associated with the specific outbreak. In particular, each illustration reports the considered metric scores obtained by each method (on the left), also with the corresponding graphical performance illustration, for a better understanding (on the right). Focusing on the performance of the LLMs, two models, namely Pythia-12b and Mpt-30b-chat, have shown underwhelming performance, falling below the standard NLP-based method, EpiTator. However, this was not the case for all LLMs. The majority of them have demonstrated a remarkable ability to outperform EpiTator by a significant margin. Among the LLMs in the evaluation, Gpt-4-32k stood out by achieving the best overall performance, as expected. Another very well performer was Gpt-35-turbo-16k, which also managed to achieve a signif-

	Prec	Rec	F1
EpiTator	0.892	0.576	0.7
Pythia-12b	0.735	0.158	0.26
Mpt-30b-chat	0.814	0.304	0.442
Llama-2-70b-chat	0.916	0.759	0.83
Mistral-7b-open	0.908	0.69	0.784
Zephyr-7b-alpha	0.92	0.804	0.858
Gpt-35-turbo-16k	0.902	0.639	0.748
Gpt-4-32k	0.913	0.734	0.814
Gpt-4-FewShots	0.904	0.658	0.762
Open-Ensemble	0.92	0.804	0.858

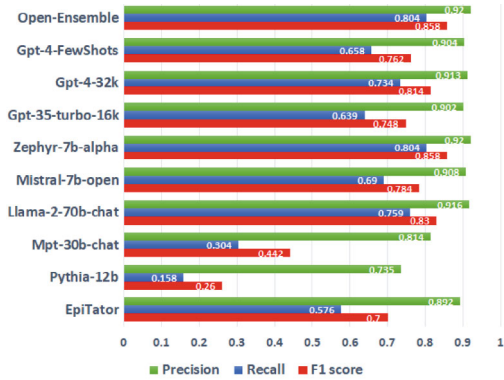


Fig. 3 Comparison of the models for the extraction of the pandemic date

	Prec	Rec	F1
EpiTator	0.387	0.321	0.351
Pythia-12b	0.365	0.277	0.315
Mpt-30b-chat	0.619	0.464	0.531
Llama-2-70b-chat	0.561	0.536	0.548
Mistral-7b-open	0.67	0.527	0.59
Zephyr-7b-alpha	0.615	0.571	0.593
Gpt-35-turbo-16k	0.699	0.455	0.551
Gpt-4-32k	0.673	0.589	0.629
Gpt-4-FewShots	0.733	0.589	0.653
Open-Ensemble	0.625	0.581	0.601

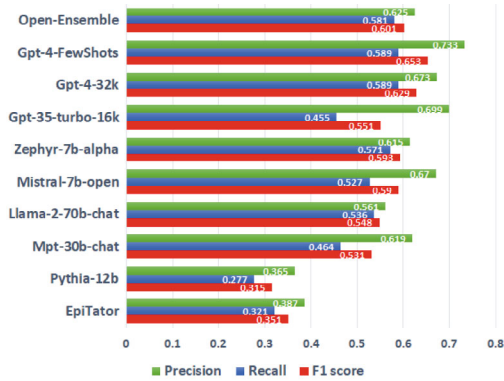


Fig. 4 Comparison of the models for the extraction of the number of cases engendered by the pandemic

icant performance level, as shown in the tables of the results. Furthermore, experimenting with in-context learning via the 3-shots method (Gpt-4-32k-FewShots), we have managed to improve the performance of Gpt-4, although this was not always the case.

Although the superior performance of OpenAI GPT models relative to the other tested methods, their use comes with its own set of challenges. These models are costly to implement, and some of their usage restrictions make them unsuitable in practice for an extensive deployment on a large dataset, like the full ProMED and DONs. On the brighter side, some open-source models performed very well, albeit slightly below the level of the OpenAI GPTs. These include Llama-2-70b-chat, Mistral-7b-openorca and Zephyr-7b-alpha. It's important to note that Llama-2-70b-chat was found to be slower in terms of computational times compared to the rest, given with large number of parameters (70 billion) of its underlying model. Finally, the adopted ensemble approach (OpenLLMs-Ensemble) on these three open-source LLMs produces an overall robust and satisfactory performance on all four epidemic

information extraction tasks, resulting to be comparable to the GPT models. Therefore, OpenLLMs-Ensemble allows for a full deployment on the entire ProMED and DONs data, spanning approximately 70,000 documents, thereby providing a comprehensive and efficient solution for infectious disease epidemic information extraction.

5 Conclusions

This study has demonstrated the significant potential of LLMs in epidemic surveillance, particularly in the extraction of valuable epidemic information from unstructured big data sources. Several LLMs were evaluated and their capabilities in extracting epidemic information were assessed. The findings of this study suggest that LLMs, particularly when used in an ensemble, can significantly enhance the accuracy and timeliness of epidemic modelling and forecasting. This application of LLMs not only streamlines the data-gathering process but also has the potential to improve early warning systems and epidemic response strategies. These models provide a promising tool for managing future pandemic events, reducing their impact on society and economies, and, finally, saving lives.

Acknowledgements The views expressed are purely those of the authors and do not, in any circumstance, be regarded as stating an official position of the European Commission. We would like to acknowledge the GPT@JRC initiative for providing access to the LLMs used in this study. We extend our gratitude also to the JRC Big Data Analytics Platform for providing secure access to various open-source AI models. Finally, we would like to thank the colleagues contributing to the WHO EIOS initiative for the helpful suggestions during the development of this work.

References

1. Abbood A, Ullrich A, Busche R, Ghazzi S (2020) EventEpi-A natural language processing framework for event-based surveillance. *PLoS Comput Biol* 16(11). <https://doi.org/10.1371/journal.pcbi.1008277>
2. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P et al (2020) Language models are few-shot learners. In: *Advances in neural information processing systems*, vol 2020
3. Brownstein JS, Rader B, Astley CM, Tian H (2023) Advances in artificial intelligence for infectious-disease surveillance. *New Engl J Med* 388(17):1597–1607. <https://doi.org/10.1056/NEJMra2119215>
4. Consoli S, Reforgiato Recupero D, Petkovic M (eds) (2019) *Data science for healthcare—methodologies and applications*. Springer. <https://doi.org/10.1007/978-3-030-05249-2>
5. Dong Q, Li L, Dai D, Zheng C, Wu Z, Chang B et al (2023) A survey on in-context learning. *arXiv* 2301:00234
6. Leuba SI, Yaesoubi R, Antillon M, Cohen T, Zimmer C (2020) Tracking and predicting U.S. influenza activity with a real-time surveillance network. *PLoS Comput Biol* 16(11). <https://doi.org/10.1371/journal.pcbi.1008180>

7. Madoff LC, Woodall JP (2005) The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch Med Res* 36(6):724–730
8. McDonald DJ, Bien J, Green A, Hu AJ, DeFries N, Hyun S et al (2021) Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction. *Proc Nat Acad Sci United States of America* 118(51). <https://doi.org/10.1073/pnas.2111453118>
9. Mukherjee S, Mitra A, Jawahar G, Agarwal S, Palangi H, Awadallah A (2023) Orca: progressive learning from complex explanation traces of GPT-4
10. Sagi O, Rokach L (2018) Ensemble learning: a survey. *Wiley Interdisc Rev Data Mining Knowl Discov* 8(4). <https://doi.org/10.1002/widm.1249>
11. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C et al (2012) Digital epidemiology. *PLoS Comput Biol* 8(7):e1002616
12. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: *Advances in neural information processing systems*, pp 3104–3112
13. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y et al (2023) Llama 2: open foundation and fine-tuned chat models
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I (2017) Attention is all you need. In: *Advances in neural information processing systems*, pp 5999–6009
15. Vespignani A (2011) Modelling dynamical processes in complex socio-technical systems. *Nat Phys* 8(1):32–39

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Effects of Inductive Load on Photovoltaic Systems



E. Okpo, P. F. Le Roux , and O. I. Okoro

Abstract The increasing demand for electrical energy, driven by technological advancements in underdeveloped and developing nations, has led to a growing reliance on renewable energy sources. Inductive loads requiring high starting currents can significantly affect power sources. Therefore, it is imperative to investigate the impacts of inductive loads on photovoltaic (PV) systems. This study aims to investigate the major parameters of the asynchronous machine, a typical representation of inductive load rated at 15 kW and 7.5 kW, respectively. With the analysis performed in MATLAB/Simulink, the major parameters investigated include the machine rotor current, rotor speed and rotor electromagnetic torque. The present study will propose strategies to mitigate the impact of inductive loads on PV systems, facilitating the seamless integration of solar PV systems into our energy infrastructure.

Keywords Photovoltaic system · Inductive load · MATLAB/Simulink · Rotor current · Rotor speed · Rotor torque · PV array

1 Introduction

The advent of technology in recent times has driven a surge in energy demand across sectors, including industrial and commercial utilities. This has informed the decision of power engineers to study alternative energy sources to reduce reliance on fossil fuel generators and reduce the expenses of diesel and gas usage. The negative environmental impacts that combustion engines impose on the environment cannot be over-emphasised. Therefore, addressing the environmental consequences of combustion engines that release greenhouse gases is imperative. PV systems emerge as optimal

E. Okpo (✉) · P. F. Le Roux
Tshwane University of Technology, Pretoria, South Africa
e-mail: ekomokpo@aksu.edu.ng

O. I. Okoro
Michael Okpara University of Agriculture, Umudike, Nigeria

solutions in this context due to their abundant availability in nature, environmentally friendly and minimal operational cost [1, 2].

The high demand associated with inductive loads, a crucial element in power systems, significantly influences distribution networks and the overall performance and efficiency of the PV systems deployed to serve such loads [3]. Therefore, the research needs to be comprehensively examine of the effects of inductive loads on PV systems and propose potential mitigation strategies to address these challenges [4, 5]. The variability in the power output from the PV system results from multiple factors, including irradiance levels, environmental temperature, panel ageing, solar panel orientation, and additional environmental variables like humidity and wind speed, all of which impact the performance of PV systems [6, 7].

Although inverter-based grid systems have existed for several decades, controversy regarding their reliability when they become the dominant source compared to conventional sources has not been addressed [8, 9]. The challenges encountered in a power system with high penetration of renewable energy, particularly solar PV systems, have been discussed in [10, 11]. Studies on comparison between the predicted transient torque and speed in a conventional model of a three-phase induction motor were conducted in [12–14], and a model with skin effect was implemented. At the same time, qualitative and analytical methods were used to evaluate the impact of rotor and stator slots with harmonics in the windings. Access to a reliable electricity supply is pivotal in empowering individuals and facilitating personal and economic development. Performance evaluation of asynchronous motor was conducted in [15]. Any nation's economic growth and development are intrinsically tied to the availability of energy [16]. However, it is imperative to consider and work on electric motors' structural analysis and geometry, as discussed in [17].

The present study is significant as it will explore the impact of inductive loads on the performance of PV systems. This exploration involves modelling and simulating a three-phase asynchronous machine under various load conditions when the PV system is interconnected to the system. Major parameters, such as electromagnetic torque, rotor speed, and rotor current of the asynchronous machine rated at 15 kW and 7.5 kW, respectively, have been meticulously examined. Solar cells were interconnected in series and parallel configurations to determine the PV system's power generation potential to meet the desired power output. This paper is structured as follows: Section Two will detail the method employed for assessing PV system performance when interfacing with inductive loads, Section Three will present the simulation results and subsequent discussions, and Section Four will provide the conclusion and recommendations. MATLAB/Simulink software will be used for all modelling and simulation purposes.

2 Method

The following approach was carefully followed during the modelling process of the PV system interfaced with inductive load to analyse the interaction and consequence of inductive load on PV system performance.

2.1 Sizing of the PV Array

The following calculations are deployed to determine the quantity of a PV array arranged in series and parallel to generate the specified power output to drive the inductive load. Assuming the PV system to operate under Standard Test Conditions (STC) with a given solar irradiance of 1 kW/m² and cell temperature of 25 °C, the PV size of the array to be connected in series and parallel, can be determined by the following combination. The number of series string N_s is expressed as follow:

$$N_s = \frac{V_{pp}}{V_{mpp}} \quad (1)$$

where V_{pp} is defined as the peak-to-peak voltage of the PV array as specified by the manufacturer and V_{mpp} is the voltage at the maximum power point. The total number of PV arrays to be connected in parallel is expressed as follows:

$$N_p = \frac{P_{out}}{N_s * P_{max}} \quad (2)$$

where N_p is defined as the number of parallel strings, P_{out} is the output power, N_s is the number of series connected strings and P_{max} is the maximum power.

The open circuit voltage of the PV can be expressed with the following expression:

$$V_{oc}(T) = V_{ocs} + \Delta V_{ocs}(T - T_s) \quad (3)$$

The Fill Factor (FF) that describes the quality of the PV cell is defined as follows:

$$FF = \frac{P_{max \text{ Practical}}}{P_{max \text{ theoretical}}} = \frac{V_{mp} * I_{mp}}{V_{oc} * I_{sc}} \quad (4)$$

The combination of the following parameters can calculate the energy generation potential from the PV system:

$$E_{in} = T_g * \alpha_{sc} * P_{sc} * G * A_{sc} \quad (5)$$

where E_{in} is the total energy from the PV module's top surface, T_g the glass transmissivity, α_{sc} the PV module absorptivity, P_{sc} represents the solar panel packing factor,

Table 1 PV module parameters specifications

S. No.	Component description	Rating
1	Maximum power (P_{max})	10 W
2	Open circuit voltage (V_{oc})	21.5 V
3	Voltage at maximum power (V_{mp})	17.7 V
4	Short circuit current (I_{sc})	0.65 A
5	Current at maximum power (I_{mp})	0.57 A
6	Inverter output voltage	400 V _{rms}

G is the irradiation intensity, and A_{sc} represents the area of the solar cell. The total PV converted to electrical energy (E_e) is expressed as follows:

$$E_e = \eta_{sc} * P_c * G * A_{sc} \tag{6}$$

The remaining portion of the PV output is converted into thermal energy expressed as follows:

$$E_t = U_t(T_{sc} - T_{bs}) \tag{7}$$

with E_l defined as the PV loss energy, E_e the electrical energy and E_t is the thermal energy.

The above equations were considered while modelling a 20.520 kW PV system deployed to serve the asynchronous machine. Shown in Table 1 are the selected parameters used in the modelling process of the PV system in MATLAB/Simulink:

The simulation model of the PV array is shown in Fig. 1 as follows:

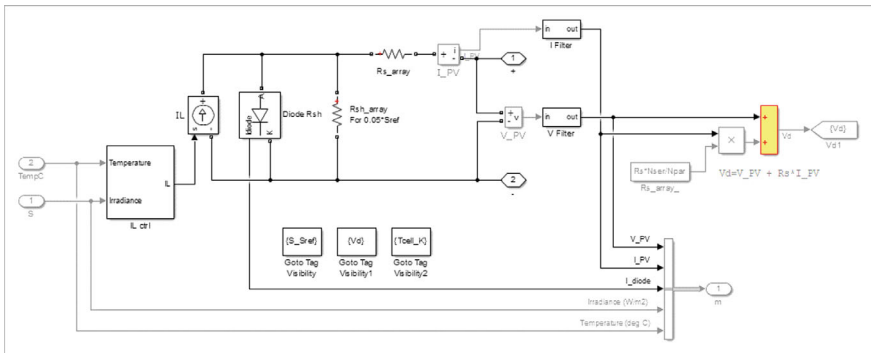


Fig. 1 Simulation model of the PV array

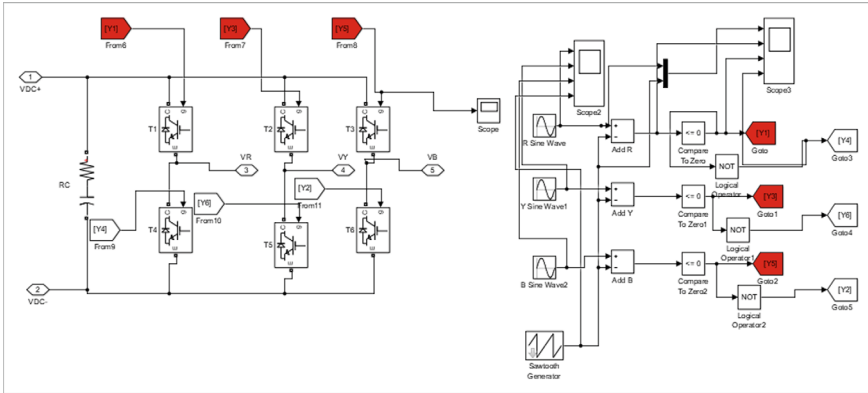


Fig. 2 Inverter simulation model

2.2 Modelling of the Inverter System

The generated output from the PV system is a DC signal and, therefore, requires an inverter to convert it into a three-phase alternating current (AC). The major components used in setting up the inverter system include the following: metal–oxide–semiconductor field-effect transistor (MOSFET), sinewave generator, saw tooth generator, input and output blocks, resistors, capacitors and logic function blocks. The component was linked in MATLAB/Simulink, with the sinewave generator producing a pulse directed to the input of the insulated gate bipolar transistor (IGBT) switches. Subsequently, a simulation was conducted to analyse the resulting waveforms at the inverters’ output. A three-phase L-C filter was implemented to mitigate the ripples emanating from the PWM inverter output. Figure 2 illustrates the simulation model of the inverter system and the sine wave generator.

2.3 Modelling of the Asynchronous Machine

The simulation was initially carried out with the 20,520 W rated PV system interfaced with an asynchronous machine rated at 15 kW and later with the PV system of the same power generation potential interfaced with an asynchronous machine rated at 7.5 kW to enable a novel comparison of the two different scenarios under varying load conditions. The parameters listed in Table 2 were deployed for modelling these asynchronous machines in MATLAB/Simulink version R2016a. This evaluation is aimed to determine the comprehensive performance of the machines.

The load torque is calculated using the following combination as follows:

$$T_l = \frac{P_{output}}{\omega} \tag{8}$$

Table 2 Specifications of the asynchronous machine parameters

S. No.	Parameter	Value
1	Input power of the motor	20 HP or 15 kW
2	Motor input voltage	400 V
3	Frequency	50 Hz
4	Motor speed	1460 RPM
5	Mechanical input	Torque T_m
6	Mechanical power	20,515 W
7	Stator resistance	0.2147 Ω
8	Stator inductance	0.000991 H
9	Rotor resistance	0.2205 Ω
10	Rotor mutual inductance	0.06419 H
11	Inertia (J)	0.102 kg.m ²
12	Friction factor (F)	0.009541 Nms
13	Number of pole pairs	2
14	Initial condition	0000

The following equation defines the motor power:

$$P_{\text{output}} = T_{\text{sh}} * \omega \quad (9)$$

2.4 PV System Interfaced with Asynchronous Machine

The PV system was utilised to power the 15 kW and the 7.5 kW asynchronous machine under different load conditions. The investigation aimed to assess their efficiency and impact on the PV system. The comprehensive model illustrating the interface between the PV system and the induction machine is depicted in Fig. 3 as follows

3 Outcome of Simulation Results

The PV array consists of the following specified input parameters as calculated in Sect. 2, under standard test conditions (STC) as shown in Table 3

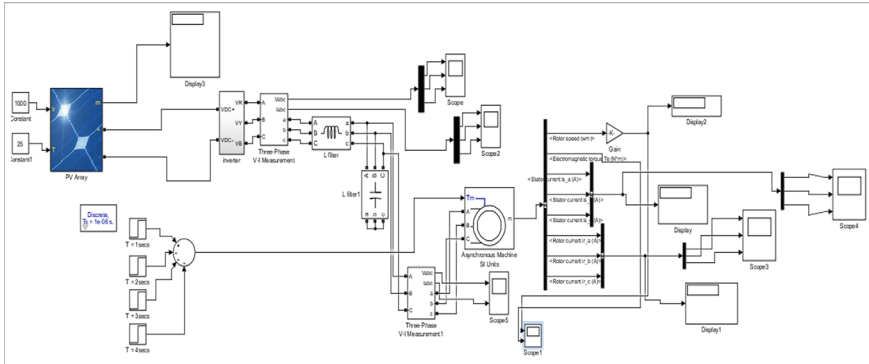


Fig. 3 PV System interfaced with inductive load

Table 3 Required parameters of the modelled PV array

S. No.	Required parameters	Value
1	Irradiance	1000 W/m ²
2	Number of parallel strings	64
3	Number of series-connected modules per string	32
4	Number of cells per module	60
5	Short circuit current (I_{sc})	0.65 A_{sc}
6	Open circuit voltage (V_{oc})	21.5 V_{oc}
7	Current at maximum power (I_{mp})	0.57 A
8	Temperature coefficient of short circuit current (I_{sc})	0.102 °C
9	Temperature coefficient of open circuit voltage (V_{oc})	- 0.36099 °C
10	Voltage at maximum power (V_{mp})	17.7 V
11	Solar cell maximum power	10.089 W

3.1 Asynchronous Machine Interfaced with PV System

The following major parameters of the induction machine were investigated to examine the impacts of inductive loads on the PV system. The investigation was first carried out on a 15 kW machine and then on a 7.5 kW machine. The motor was subjected to intermittent loading for 4 s to determine the response of the PV system under varying load conditions. The calculated values of the load torque for both the 15 kW and 7.5 kW machine deploying (9) in Sect. 2 are presented in Table 4 as follows:

Comparison of the following major parameters of the asynchronous machine rated at 15 kW and 7.5 kW were examined, including the rotor current, rotor electromagnetic torque, and rotor speed, respectively. The observed signal waveforms are presented in Figs. 4 and 5.

Table 4 Motor torque and speed of the asynchronous motor rated at 15 kW under different load conditions

S. No.	Loading condition in seconds	Load torque (T_L) in N-m on the motor rated at 15 kW and speed of 1460 RPM	Load torque (T_L) in N-m on the motor rated at 7.5 kW and speed of 1440 RPM
1	T_L at 1 s = T_L	98	49.7
2	T_L at 2 s = $T_{L/2}$	49	24.9
3	T_L at 3 s = $T_{L/3}$	33	17
4	T_L at 4 s = $T_{L/4}$	25	12.4

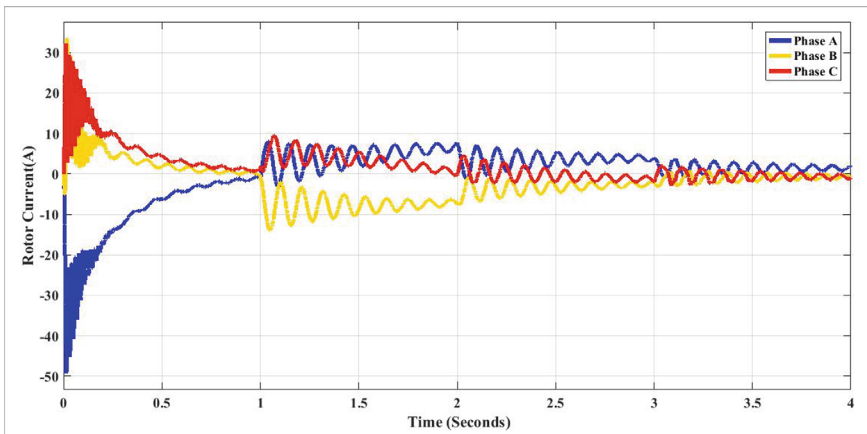


Fig. 4 Three-phase rotor current of the asynchronous machine rated at 15 kW

The outcome showed that a decrease in load capacity results in a decrease in the rotor current, as the peak amplitude of the rotor current was higher when the 15 kW machine was connected to the PV system and notably decreased when it was replaced with a 7.5 kW rated asynchronous machine, this showed that decrease in the load subjected to the PV system invariably improves its performance. The electromagnetic torque of the motor at 15 kW and 7.5 kW, respectively, under varying load conditions, will be compared in Figs. 6 and 7.

The electromagnetic torque of the asynchronous becomes more stable and tends to operate closer to its ideal condition when the 15 kW machine is replaced with a 7.5 kW machine. Figures 8 and 9 show the speed behaviour of the asynchronous machine rated at 15 kW and 7.5 kW, respectively.

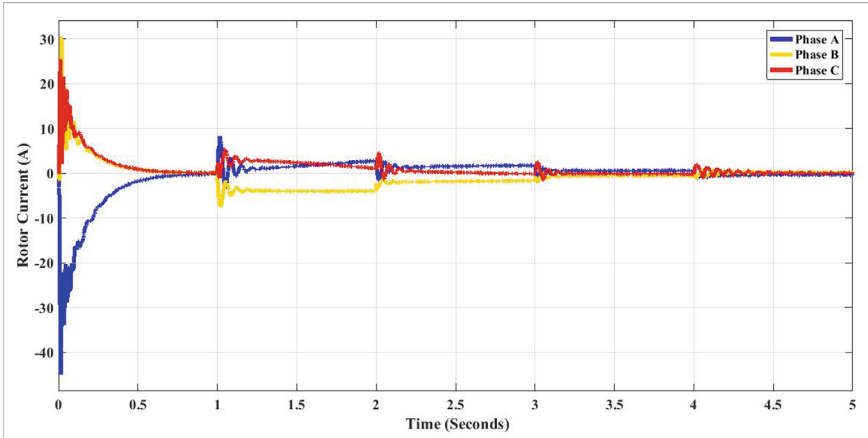


Fig. 5 Three-phase rotor current of the asynchronous machine rated at 7.5 kW

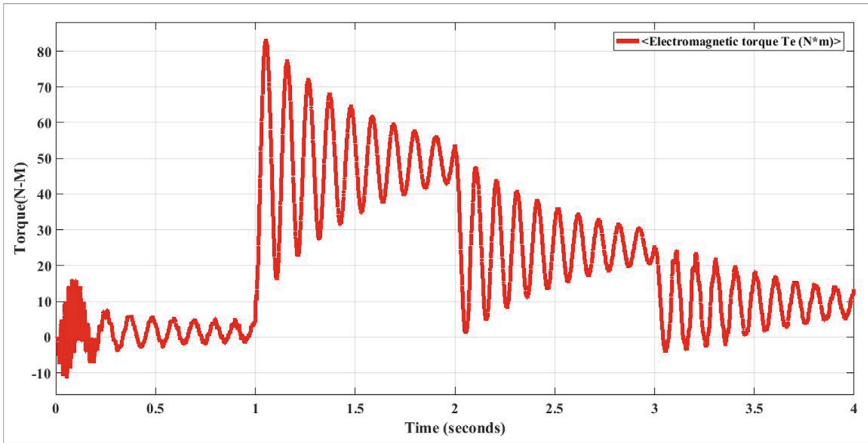


Fig. 6 Rotor electromagnetic torque of the asynchronous machine rated at 15 kW

The stability of the rotor speed significantly improves and responds quickly to load variations. This indicates a significant improvement in the speed as the capacity was reduced. This case scenario can also be tested on the IEEE power systems [18].

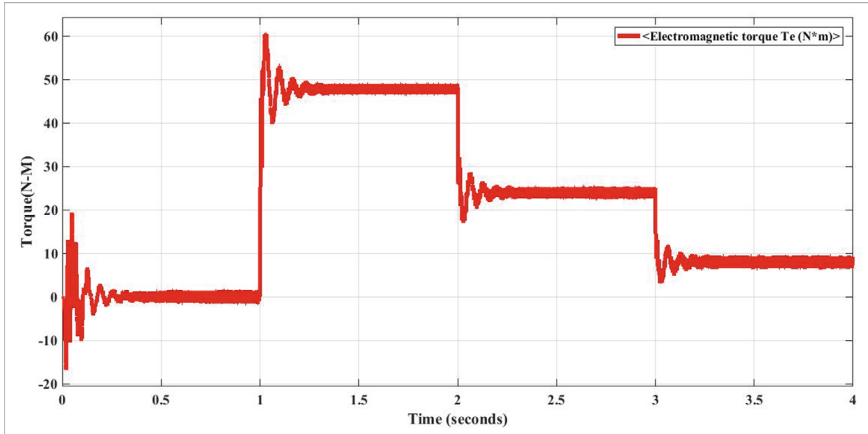


Fig. 7 Rotor electromagnetic torque of the asynchronous machine rated at 7.5 kW

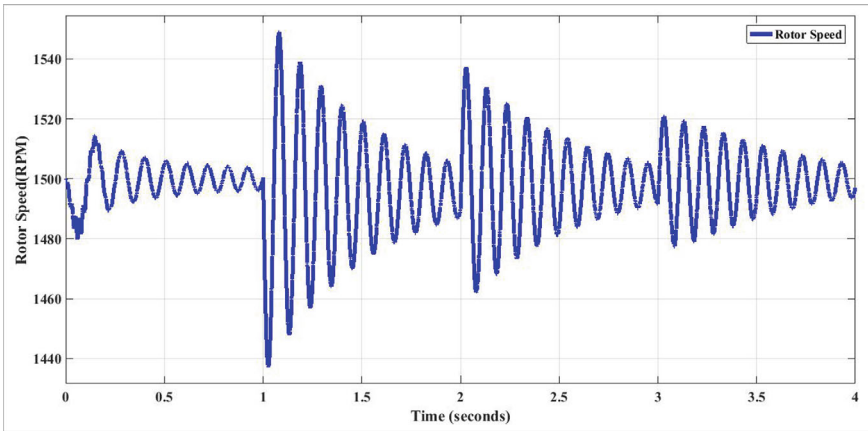


Fig. 8 Rotor speed of the asynchronous machine rated at 15 kW

4 Conclusion

The significant impacts on the performance of PV systems when interfaced with inductive loads have been carried out. The outcome illustrated that a proper load management system is imperative, particularly when power is sourced from a PV system to drive inductive loads such as induction machines due to their high-power requirements and several applications in residential buildings, commercial centres and industries. Adherence to this will mitigate the fluctuation or total failure of the PV system due to its fragility. The study’s outcome showed that significant improvement in the machine stability is recorded when the asynchronous machine rated at 15 kW was replaced with 7.5 kW machine while maintaining the same power supply from

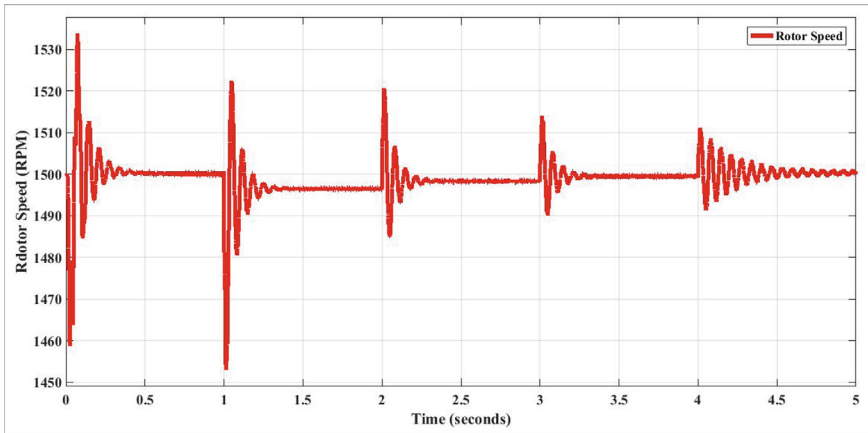


Fig. 9 Rotor speed of the asynchronous machine rated at 7.5 kW

the PV system. The present study also showed that oversizing the PV system to serve a lower capacity load will not only enable the PV system to adjust to varying loads with ease but will also enhance the stability and optimal operation of the induction machine.

References

1. Iwuamadi Obioma C et al (2020) Electric power insecurity: the bane of sustainable industrial development in Nigeria. *Electr Power* 1(1)
2. Yang J et al (2018) Climate, air quality and human health benefits of various solar photovoltaic deployment scenarios in China in 2030. *Environ Res Lett* 13(6):064002
3. Makola C, Le Roux P, Jordaan J (2021) Design, analysis, and operation of photovoltaic power in a microgrid with an EESS. In: 2021 IEEE PES/IAS PowerAfrica. IEEE
4. Masadeh M (2018) Modeling and Emulation of induction machines for renewable energy systems, Concordia University
5. Makola CS, Le Roux PF, Jordaan JA (2023) Comparative analysis of lithium-ion and lead-acid as electrical energy storage systems in a grid-tied microgrid application. *Appl Sci* 13(5):3137
6. Zainal NA, Yusoff AR (2016) Modelling of photovoltaic module using matlab simulink. In: IOP conference series: materials science and engineering. IOP publishing
7. Song Z, Liu J, Yang H (2021) Air pollution and soiling implications for solar photovoltaic power generation: a comprehensive review. *Appl Energy* 298:117247
8. Shakerighadi B et al (2023) An overview of stability challenges for power-electronic-dominated power systems: the grid-forming approach. *IET Gener Transm Distrib* 17(2):284–306
9. Mohammed N, Alhelou HH, Bahrani B, Grid-forming power inverters: control and applications, CRC press
10. Hoke A et al (2021) Island power systems with high levels of inverter-based resources: stability and reliability challenges. *IEEE Electrification Mag* 9(1):74–91
11. Gu Y, Green TC (2022) Power system stability with a high penetration of inverter-based resources. In: Proceedings of the IEEE
12. Okoro OI (2002) Dynamic and thermal modelling of induction machine with non-linear effects, Kassel University Press Kassel

13. Le Roux P, Ngwenyama M (2022) Static and dynamic simulation of an induction motor using matlab/simulink. *Energies* 15(10):3564
14. Mishra A, Rajpurohit B (2014) Modeling of 3-phase induction motor including slot effects and winding harmonics. In: 2014 6th IEEE power India international conference (PIICON). IEEE
15. Okpo EE et al (2019) Performance evaluation of 5.5 kW six-phase asynchronous motor. In: 2019 IEEE PES/IAS PowerAfrica. IEEE
16. Eseosa O, Enefiok OE (2015) PV-diesel hybrid power system for a small village in Nigeria 1(4)
17. Lee J (2010) Structural design optimization of electric motors to improve torque performance, University of Michigan
18. Le Roux P, Ngwenyama M, Aphane T (2022) 14-Bus IEEE electrical network compensated for optimum voltage enhancement using FACTS technologies. In: 2022 3rd International conference for emerging technology (INCET). IEEE

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Low Code Development Cycle Investigation



Małgorzata Pańkowska 

Abstract Technology plays an important role in the advancement of end-user development of software applications. It supports the way requirements are elicited, collected, analyzed, and processed into functionalities and non-functionalities in information systems. Technology enables end-users to create their own application for particular usage. This paper aims to present low code application development methodology resulting from practical experience as well as from the literature survey. The paper findings present that the low code development cycle (LCDC) is different in comparison with traditional or agile software development methods. The LCDC should emphasize the role, competencies, and experience of the end-user, who is a low code developer. This paper consists of two main parts. The first part covers literature surveys on contemporary approaches to citizen software development and low coding. The second part covers the LCDC ecosystem and process, which reveals the main active role of end-user and supportive roles of other project stakeholders, i.e., analysts, software engineers, testers, integrators, and other facilitators.

Keywords Citizen development · Low code · Ecosystem · Process · Life cycle

1 Introduction

Collaboration in the end-user software development community involves a variety of participatory methods and stakeholder objectives that do not necessarily align with those in conventional information system life cycle development collaboration. Particularly, the end-user (i.e., citizen) software development requires creativity of users, their engagement, flexibility, and decision-making throughout a project. Beyond that, an infrastructure is needed that supports communication, tooling, and software application implementation. The approach of involving in the development

M. Pańkowska (✉)

Department of Informatics, University of Economics in Katowice, 1 Maja, Silesia, 40-287, Katowice, Poland

e-mail: pank@ue.katowice.pl

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_19

265

of end-users is increasingly used and is known as citizen development. It requires usage of specialized Computer Aided Software Engineering (CASE) tools, i.e., Low Code Development Platform (LCDP). Citizen software development fosters an open and participatory approach to application implementation, reducing the distance between end-user and software engineer, and contributing to the goals of a business unit. End-user is allowed to construct forms of logic for solving problems and to convert the constructed form of logic into built software. The low-code applications are implemented and exploited for various purposes in different disciplines, however, elaboration of a common methodology for low-code software life cycle development is still a challenge. First, the heterogeneity of citizens, i.e., end-users is particularly evident in various business organizations. They are in big as well as in small business units. They have various competencies, i.e., knowledge, skills, social competencies, experiences, habits, preferences, risk averse, and willingness to learn and to be innovative. Hence, a generalization of their requirements is almost impossible, and by definition low code applications are preferably developed by individuals and for their individual usage. That situation reveals problems of low code software maintenance. However, some end-users can be interested in dissemination of their software development activities for an enrichment of the citizen development as a method. Hence, they provide software documentation and work on integration of their application with other parts of the business information systems, i.e., legacy systems, in a business organization. Taking into account the basic duties of end-users and therefore their limited time for software development, the citizen development calls for inclusivity, which is achieved through selected practices, tools, and methodological procedures. Particularly, the computerized tools that serve the practical implementation of citizen software are to be supported by guidelines and highly intuitive. End-users usually have no time to learn, but they want to realize their requirements as quickly as possible.

Communication in the field of citizen software development means more than just publishing results. Short time success motivates further work, but on the other side—failure results in disappointment and discourages further involvement. Hence, the citizen developers need LCDP facilitators, who help in learning and the application optimization. There are also various attitudes of top managers, who may support, but, on the other side, can discourage business employees to individual application development. Successfully, an increasing number of institutions, including government agencies, and big companies, are showing an interest in the field of citizen software development. Surprisingly, they believe that the low-code software can be treated as a prototype of further professionally developed applications. Through prototyping, end-users are able to reveal their real expectations. Hence, the process of requirement elicitation for professional IT companies is much easier than in a situation, when the end-user does not want to be involved in an interviewing or software customization, and deployment process. Beyond that, small and medium enterprises (SMEs) are interested in citizen software development because of the lower cost of software prepared by end-users, i.e., software prosumers, who are doing the applications for themselves and by themselves. Frequently, SMEs financial conditions exclude

investment in professional software available in cloud or on-premise, but creativity of end-users and access to open source tools permit to fulfill the requirements.

This study aims to support a low code software development methodology. Author focuses on identification of low code application ecosystems as well as on modeling processes covering low-code software development life cycle. In this contribution, the author emphasizes the particular role of the end-user. The rest of that paper includes the following sections. First, a literature survey has been done to present actual justification of low-code software development. Next, the low-code software development ecosystem in Archimate notation is presented, and finally, the Business Process Model and Notation (BPMN) process of Low Code Development Life Cycle (LCDLC) is included.

2 Theoretical Background of Low Code Development

The Internet democratization refers to the rapid growth of access to Information Communication Technologies (ICTs). The Internet users have access to information as well as to software development tools. According to Haklay et al. [8], citizen science means public participation in a scientific research activity. By analogy, citizen software development is defined as a software project conducted by amateurs or nonprofessional designers. Citizen software development creates opportunities for conducting investigations on requirements elicitation, co-designing, and co-testing of low-cost sensors, as well as on collecting and co-analysing data, and co-developing data applications. Volunteerism is important to the understanding of citizen software development, which presents a step forward in the democratization of application implementation. That approach brings new knowledge to particular domains as well as to ICT. Citizen science has been for years developed in the do-it-yourself (DIY) culture for pragmatic reasons. For example, over the past 30 years, Visual Basic applications have been implemented by laypersons in business for their own usage. That approach to software development is also known as information system (IS) prosumption, i.e., production for own exploitation.

According to Senabre Hidalgo et al. [26], public participation in science includes participatory action research (PAR) and the involvement of various social organizations, i.e., chamber of commerce, consumer associations, non-profit organizations, social groups, and environmentalists' associations. Some of them are low-budget business units, looking for relatively low-cost investments in the ICTs. Although traditionally, an IT professional is a software project manager and end-users work as data gatherers and requirements' providers, in the citizen software development approach end-users are able to reveal their real problems, because they take active role and software engineers are co-designers and facilitators in that process. Facilitators are expected to create the necessary conditions for equitable and free speaking, and to support collective decision-making during project meetings. Senabre Hidalgo et al. [26] argue that a facilitator is in charge of suggesting the materials and ensuring discussions among project partners. End-users' participation in software projects has

been defined as the degree to which the user is involved in producing and exploitation of the project products. User involvement refers to the subjective psychological state reflecting the importance of a given system.

Participatory Design (PD) has focused on the designing of user applications or the co-realization of information systems. In this approach, users are the best to determine how to automate their work, and the designers are just consultants. PD is realized through workshops where users can feel free to express their ideas concerning the designed applications [28]. User-Centered Design (UCD) is a philosophy that is based on studying the end-users' needs and interests. Not only the profiles, activities and environments of users are being investigated, but also their goals and values [20]. UCD is a process focusing on usability throughout the entire development process and further throughout the system life cycle. This philosophy highlights the role of prototyping, which must be used to visualize and evaluate ideas and design solutions in cooperation with the end users in a real context [7]. The method of User Experience (UX) Design is defined as the creation of the subjective relationship between user and application or device. That design is about considering the user, the task, and the context, i.e., environment of work within the company culture. It aims to enhance the entire experience and emotions resulting from the use of a product, system, or service [1]. Actor Network Theory (ANT) was developed by Callon and Latour to describe the creation and evolution of socio-technical systems and to focus on the dynamics of relationships among main actors (i.e., end users) and networks (i.e., business managers and ICT people) [3]. Those theories, although useful and supportive for software project management, have not emphasized the main role of end-users in the application implementation process.

Lately developed agile methods and DevOps approach are also combined with the LCDP usage. Lebens and Finnegan [12] have proposed to use the LCDPs to teach Agile Methodology. This approach was rational, because the agile methods focus on the end user involvement in fast result generating, prototyping, and software validation. University students can use the Microsoft Power Apps platform to develop their competences in agile software development. Rather old, but still valid approach can be the Rapid Application Development (RAD), which is a software development methodology promoting the rapid release of a software prototype. This agile approach utilizes the user feedback from the prototype for further improvement of the product. Therefore, the low-code prototypes can be included in projects, where developers use the RAD tools, i.e., LCDPs, Joint Requirement Planning (JRP), and the Joint Application Development (JAD). JAD uses organized and intensive workshops to bring together project stakeholders, i.e., sponsors, users, analysts, and programmers to jointly define and design information systems. Low code prototyping aims to make end-user ideas tangible, realizable by creating a first implementation. The low code development platform can also be used in the Design Thinking (DT) methodology. Although it is still a challenge, the DT is a methodology that uses co-designing and various problem solving techniques to ensure an alignment of business needs with what is technologically feasible. The DT methodology incorporates an empathy for the user, rapid prototyping and abductive reasoning [4].

Taking into account that citizen software development is a way to democratize the application implementation, this study has formulated two research questions, i.e., What are the Low Code Development (LCD) Methodologies? What are the Citizen Development (CD) methodologies? The author has proposed the research structure consisting of two parts, i.e., a literature survey, and the LCDLC proposal. Answers that research questions have been received through the literature surveys. Therefore, the author reviewed the following repositories: Scopus, Association for Information Systems Electronic Library (AISELib), Springer Link, IEEEExplore, and Sage Journals. The search phrase was as follows: “citizen development” AND “low code development”. In general, the author has found 1272 publications released in 2010–2023. The publication sources were not restricted. Articles were identified and screened for relevance through abstract, title, and keywords. Next, 700 publications were excluded because of lack of access to the full text and too less valuable information about the research findings in the abstracts. Hence, 572 publications were reviewed. However, that number of publications was further reduced, because mostly, they were not strictly about the citizen software development methodology, nor about the low code development methodology. In the next step, 52 publications were studied. That publications cover mostly discussions on the low code platforms, their applicability strengths and weaknesses, as well as on the LCDP’s experimental usage results.

As the LCDPs provide a graphical user interface to drag and drop with little effort and in a user-friendly way, they are easy to learn and speed up the application development. Martinez-Lasaca et al. [15] argue that the LCDPs enable prototyping, hence project stakeholders have an opportunity to find a compromise solution in a conflicting situation. The LCDPs are particularly preferred in domains that have quite well identified procedures and processes, and in which stakeholders are interested in the business process automation. The LCDPs enable data handling and workflow automation based on business processes modeled in the Business Process Model and Notation (BPMN) as a reference modeling language [24, 29]. The survey results show that the domains where low code development is mostly used are agriculture [6, 18, 19], Internet of Things (IoT) application exploitation [5], manufacturing industry [23, 25], pharmaceutical enterprises [17], high school education [14], and Enterprise Resource Planning (ERP) cloud services and components [27]. Brouzos et al. [2] discuss the low code development of cyber-physical systems such as robots and sensors. Martins et al. [16] present a model designed to transform ChatGPT into a low code developer. Leading vendors in the LCDP field are Appian, Mendix, Microsoft, OutSystems, Salesforce, and Service Now [9].

3 Low Code Development Life Cycle Model

The LCD literature survey allows for an identification of a gap concerning the low code development life cycle modeling from the end user perspective. Rokis and Kirikova [21] have identified the LCD principles, i.e., selection of the right platform, visual application outline, usage of predefined components, empowerment of citizen developers, establishing a project team and governance, and iterative development life cycle. They have noticed some necessities, i.e., feasibility analysis, data modeling, defining of the user interfaces, implementation of business logic, integration of external services, testing and deployment. Some of those necessities result from the LCDP guidelines, others are included in the traditional system development life cycle approach. Kuruoglu et al. [10] have presented the low code development methodology based on process modeling and prototyping. Ruscio and Pierantonio [22] have proposed a low code development life cycle consisting of some stages, i.e., data modeling, user interface design, business logic implementation, integration of external services, testing and deployment, and customer feedback. Krejci et al. [11] have recognized that software development on a LCDP consists of similar process phases as in traditional projects. They have identified the following LCD project phases: preparation, generation, improvement and implementation, evaluation, deployment, and improvement in next iterations. As presented in the literature, low-code software development stages are discussed on a high level of preliminary investigation. Unfortunately, the authors do not emphasize the unique role of end-users. However, this study focuses on the LCD process, in which the end-user plays the main roles (see Fig. 2). That LCD process is included in the LCD ecosystem modeled in the Archimate language (see Fig. 1).

The complexity and importance of the LCD software projects, which include using hardware equipment and the LCDP software resource to create products and services encourage the development of the whole LCD ecosystem (see Fig. 1). The concept of ecosystem is quite comfortable, because the developers need not to concentrate on a particular enterprise, but they can focus on identification of the project goals, drivers, stakeholders, principles, business actors, processes, material resources, software and hardware components. The ecosystem planning is always located in a certain strategic motivation context, which is necessary to justify the business requirements and the need to design and implement the software application. That context includes business justification of the LCDP usage. In Fig. 1, the context is limited to goals, drivers, and principles. The non-professional may collaborate with professionals in all stages, but mostly they alone contribute to software implementation. Their goals are as follows (based on [13]):

1. Self-direction, competencies enrichment—"I want to learn".
2. Power through exercising control over people, data, and material resources—"I want to gain recognition and status".
3. Commitment to realization of the business unit strategy—"I want to improve my organization works".
4. Efficiency, reduction of the ICT costs—"I can implement cheaper application".

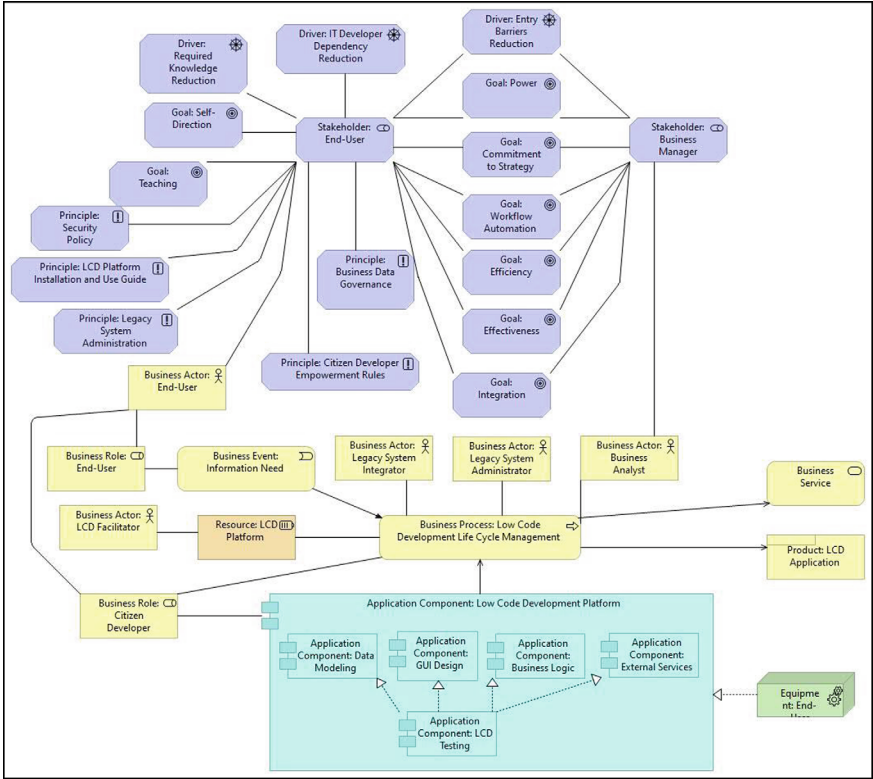


Fig. 1 The low code development ecosystem model in archimate notation

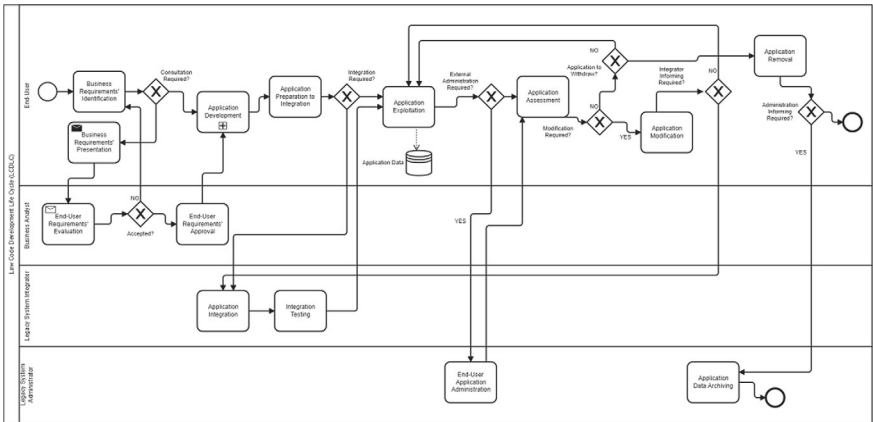


Fig. 2 The low code development life cycle process in notation BPMN

5. Workflow automation—"I want to do it automatically".
6. Effectiveness, elicitation of unique requirements—"I can implement an unique functionality, which is not available in a Commercial of The Shelf (COTS) software".
7. Teaching and providing an educational opportunity to others—"I want to share my knowledge and experience".
8. Integration—"I want to have my application integrated with others".

The identified goals and drivers are linked to the stakeholders, i.e., End-User and partially to the Business Manager. The drivers stimulating low-code development are the following: reduction of entry barriers for software development, reduction of required knowledge for application development, and reduction of dependence on IT developers [9]. Some principles are assumed to be important to the End-user. They are as follows: Security Policy, LCDP Installation and Use Guide, Legacy System Administration, Business Data Governance, and Citizen Developer Empowerment Rules. The LCD ecosystem model takes into account two general scenarios. The first scenario assumes that a citizen developer is working for a separate business unit, e.g., SME, micro company. Then the developer is the only person responsible for data governance and application implementation, maintenance, and administration. The second option covers situations, when a citizen developer is working for a SME or big company. Then the application can be combined with other business software, in this ecosystem model named legacy systems. In the second case, stakeholders have to solve problems of the application integration and administration, as well as data governance. Generally, the LCD ecosystem model covers just one process named Low Code Development Life Cycle Management. The Archimate models allow for high-level strategic modeling of technology implementation issues. Details are included in Figs. 2 and 3. The detailed design should be elaborated in modeling languages, e.g., SysML, UML. Implementation can be done in a LCDP, e.g., OutSystems, Webcon (webcon.com). The LCD ecosystem model assumes that a business event triggers a process, which results in providing business service and product, i.e., the low code application. The process realization depends on the selected LCDP, which includes components for data modeling, user interface design, business logic modeling and implementation, and external services integration. In both scenarios, the LCDP is assumed to be located on the End-User Computer. In this way, the author emphasizes the specific role of citizen developer and meaning of the LCDP, which is not widely used by professional software engineers. That model includes another actor, who is the LCDP Facilitator responsible for helping the End-User in learning, designing, and maintenance of the low code application. The End-User as an actor has two roles: End-User and Citizen Developer. The LCDLC process is included in Figs. 2 and 3.

The LCDLC process is initiated by the End-User, who is inspired by the personal information needed to develop an LCD application. That requirement can be consulted with the Business Analyst, but in another option, it can be developed independently. The same approach is applied to all other tasks in the LCD cycle. The important issue, which arrives in that process, is the governance and retention

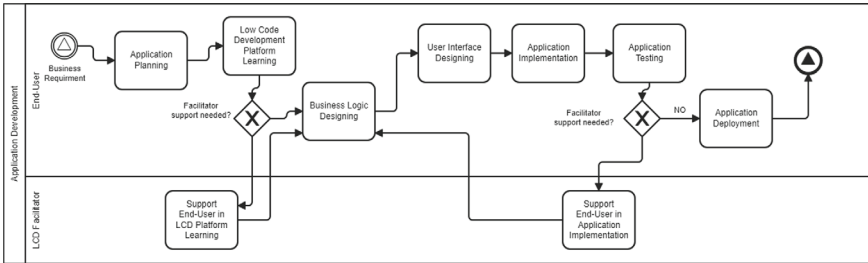


Fig. 3 The application development sub-process in the LCDLC process in notation BPMN

of the application data. The business data should be eventually secured and archived by the business representative, e.g., Administrator. Hence, the company should have elaborated the Citizen Developer Empowerment Rules (see Fig. 1), which explains the data governance principles. For SMEs and micro companies, when the Citizen Developer is the business owner and administrator, they establish the data governance rules. However, the situation when the application and data are just for one End-User is also possible. The LCDLC process (see Fig. 2) includes one sub-process named Application Development (see Fig. 3).

Through this LCDLC process model, the author wants to emphasize the main role of the End-User in the application development process. This opportunity was created just because of the LCDPs availability. This process encourages further questions, e.g., the data governance, involvement of the business analyst, the professional software engineers, system administrators, system integrators, and testers in the application development, as well as the LCDP maintenance costs and procedures. Beyond the lack of governance rules, there are some other inhibitors, i.e., unsolved issues of scalability, limited portability significantly inhibit the LCDP adoption, as well as high prices of commercial tools, lack of customization, vendor lock-on, and lack of knowledge about LCDPs [9].

4 Conclusions

Although end-users have opportunities to use various low code development platforms, the different technical capabilities and multiple scenarios of usage still create new challenges. The literature survey revealed that IT professionals use the LCDPs under control of IT companies. In this paper, the author presents another process of the LCDP usage, which is less formalized, and its heterogeneity hinders its generalization.

References

1. Beccari MN, Oliveira TL (2011) A philosophical approach about user experience methodology. In: Markus A (ed) *Design, user experience and usability, theory, methods, tools and practice*. Springer, Heidelberg, pp 13–22
2. Brouzos R, Panayiotou K, Tsaoudoulas E, Symeonidis A (2023) A low-code approach for connected robots. *J Intell Rob Syst Theory Appl* 108(2):28
3. Callon M, Latour B (1981) Unscrewing the big leviathan: how actors macro-structure reality and how sociologists help them to do so. In: Knorr-Cetina KD, Cicourel AV (eds) *Advances in social theory and methodology: towards an integration of micro and macro-sociologies*. Routledge and Kegan Paul, London, pp 277–283
4. Chasanidou D, Gasparini AA, Lee E (2015) Design thinking methods and tools for innovation. In: Marcus A (ed) *Design, user experience, and usability, design discourse*. Springer, Cham, pp 12–23
5. Chaudhary HAA, Guevara I, Singh A, Schieweck A, John J, Margaria T, Pesch D (2023) Efficient model-driven prototyping for edge analytics. *Electron (Switzerland)* 12(19), Article number 3881
6. Fatouras G, Kousiouris G, Lohier T, Makridis G, Polyviou A, Soldatos J, Kyriazis D (2023) Enhancing smart agriculture scenarios with low-code, pattern-oriented functionalities for cloud/edge collaboration. In: *Proceedings—19th International conference on distributed computing in smart systems and the internet of things, DCOSS-IoT 2023*, pp 285–292, IEEE, Pafos
7. Gulliksen J, Goransson B, Boivie I, Persson J, Blomkvist S, Cajander A (2005) Key principles for user-centered systems design. In: Seffah A, Gulliksen J, Desmarais MC (eds) *Human-centered software engineering—integrating usability in the software development lifecycle*. Springer, Berlin, pp 17–37
8. Haklay MM, Dorler D, Heigl F, Manzoni M, Hecker S, Vohland K (2021) What is citizen science? The challenges of definition. In: Vohland K, Land-Zandstra A, Ceccaroni L, Lemmens R, Perello J, Ponti M, Samson R, Wagenknecht K (eds) *The science of citizen science*, pp 13–34. Springer, Cham
9. Kass S, Strahinger S, Westner M (2023) Practitioners’ perceptions on the adoption of low code development platforms. *IEEE Access* 11:29009–29034
10. Kuruoglu Dolu B, Cetinkaya A, Kaya MC, Nazlioglu S, Dogru AH (2022) MSDeveloper: a variability-guided methodology for microservice-based development. *Appl Sci (Switzerland)* 12(22):11439
11. Krejci D, Iho S, Missonier S (2021) Innovating with employees: an exploratory study of idea development on low-code development platforms. In: *ECIS 2021 Research papers*. 118. https://aisel.aisnet.org/ecis2021_rp/118. Accessed 21 Nov 2023
12. Lebens M, Finnegan R (2021) Using a low code development environment to teach the agile methodology. In: Gregory P, Lassenius C, Wang X, Kruchten P (eds) *XP 2021*, vol 419. LNBIIP. Springer, Cham, pp 191–199
13. Land-Zandstra A, Agnello G, Selman Gultekin Y (2021) Participants in citizen science. In: Vohland K, Land-Zandstra A, Ceccaroni L, Lemmens R, Perello J, Ponti M, Samson R, Wagenknecht K (eds) *The science of citizen science*. Springer, Cham, pp 243–260
14. Luo J, Hou M (2023) Innovation of the higher education grassroots statistical reports system based on low-code development. In: *7th International conference on management engineering, software engineering and service sciences, ICMSS*, pp 36–40
15. Martinez-Lasaca F, Diez P, Guerra E, de Lara J (2023) Dandelion: a scalable, cloud-based graphical language workbench for industrial low-code development. *J Comput Lang* 76:101217
16. Martins J, Branco F, Mamede H (2023) Combining low-code development with ChatGPT to novel no-code approaches: a focus-group study. *Intelli Syst Appl* 20:200289
17. Miyake T, Masuda Y, Oguchi A, Ishida A (2023) Strategic risk management for low-code development platforms with enterprise architecture approach: case of global pharmaceutical enterprise. *Smart Innovation Systems Technol. SIST* 357:88–100

18. Novales A, Mancha R (2023) How hortilux used low-code to develop its IoT digital service. *Commun Assoc Inf Syst* 53:924–937, 38
19. Oteyo IN, Scull Pupo AL, Zaman J, Kimani S, De Meuter W, Gonzalez Boix E (2023) Easing construction of smart agriculture applications using low code development tools. In: Longfei S, Bodhi P (eds) *Mobile and ubiquitous systems: computing, networking and services. MobiQuitous 2022*, pp 21–43. Springer, Cham
20. Raatikainen P, Pekkola S (2023) User-centredness in large-scale information systems implementation. In: 12th Scandinavian conference on information systems. 3. <https://aisel.aisnet.org/scis2021/3>. Accessed 23 Dec 2023
21. Rokis K, Kirikova M (2023) Exploring low-code development: a comprehensive literature review. *Complex Syst Inf Model Q* (36):68–86, 200
22. Ruscio, D., Pierantonio. A.: Supporting the understanding and comparison of low-code development platforms. In: 2020 46th Euromicro conference on software engineering and advanced applications (SEAA), pp.171–178. IEEE, Portoroz (2020).
23. Qu D, Zhang Y, Hu X, Dai W (2023) Contract-based design for low-code development in industrial edge applications. In: 2023 IEEE 32nd International symposium on industrial electronics (ISIE), pp 1–6. IEEE, Helsinki
24. Sahay A, Di Ruscio D, Iovino L, Pierantonio A (2023) Analyzing business process management capabilities of low-code development platforms. *Softw Pract Experience* 53(4):1036–1060
25. Schenkenfelder B, Salomon C, Buchgeher G, Schossleitner R, Kerl C (2023) The potential of low-code development in the manufacturing industry. In: IEEE international conference on emerging technologies and factory automation, ETFA, Code 193521. IEEE, Sinaia
26. Senabre Hidalgo E, Perello J, Becker F, Bonhoure I, Legris M, Cigarini A (2021) Participation and co-creation in citizen science. In: Vohland K, Land-Zandstra A, Ceccaroni L, Lemmens R, Perello J, Ponti M, Samson R, Wagenknecht K (eds) *The science of citizen science*. Springer, Cham, pp 199–218
27. Tang L (2022) ERP low-code cloud development. In: *Proceedings of the IEEE international conference on software engineering and service sciences, ICSESS*, pp 319–323
28. Torpel B, Voss A, Hartswood M, Procter R (2009) Participatory design: issues and approaches in dynamic constellations of use, design and research. In: Voss A, Hartswood M, Procter R, Rouncefield M, Slack RS, Buscher M (eds) *Configuring user-designer relations*. Springer, London, pp 13–30
29. Wang J, Qi B, Zhang W, Sun H (2021) A low-code development framework for constructing industrial apps. In: *Communications in computer and information science*. 1330 CCIS, pp 237–250

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



TermX—Bridging the Gap: Implementing CTS2 and FHIR Compatible Terminology Server



Marina Ivanova^{ID}, Igor Bossenko^{ID}, and Gunnar Piho^{ID}

Abstract As healthcare systems evolve, achieving interoperability and seamless data exchange becomes paramount. The key component of the data exchange is terminology managed by the terminology server. During “*Comparative Analysis of Clinical Terminology Servers: A Quest for an Improved Solution*” research, we did not find a terminology server suitable for our criteria and decided to develop our own terminology server. This article delves into implementing the Common Terminology Services 2 (CTS2) standard on PostgreSQL, a robust open-source relational database. The discussion encompasses the intricacies of integrating CTS2 with PostgreSQL, exploring the challenges and solutions encountered during the process. Additionally, the article comprehensively maps this implementation to HL7 Fast Healthcare Interoperability Resources (FHIR), a widely adopted standard for healthcare data exchange. Readers will gain insights into the technical aspects of the implementation, practical considerations, and the potential impact on healthcare data management. This exploration aims to guide developers, healthcare IT professionals, and stakeholders in leveraging CTS2 on PostgreSQL while ensuring compatibility with HL7 FHIR resources, ultimately fostering improved interoperability in the dynamic landscape of healthcare data systems.

Keywords TermX · Terminology server · PostgreSQL · CTS2 · HL7 FHIR · Healthcare interoperability · Clinical data exchange

M. Ivanova (✉) · I. Bossenko · G. Piho
Department of Software Science, Tallinn University of Technology,
Akadeemia 15A, 12618 Tallinn, Estonia
e-mail: mariiv@taltech.ee

I. Bossenko
e-mail: igor.bossenko@taltech.ee

G. Piho
e-mail: gunnar.piho@taltech.ee

1 Introduction

In the dynamic healthcare landscape, combining technological innovation and robust clinical terminology management is a cornerstone for achieving comprehensive interoperability and data standardisation [1]. The healthcare sector is undergoing a paradigm shift driven by the imperative recognition of the pivotal role played by clinical terminologies, encompassing Code Systems, ValueSets, and ConceptMaps. Unified and standardised internal data models are essential for achieving interoperability transcending the complexities of storing, retrieving, and exchanging clinical information [1]. This research covers the design of the code system and its sub-components.

1.1 Rationale for CTS2 Implementation

Common Terminology Services 2 (CTS2) is a standard that defines different code systems, value sets, and map sets, focusing on terminology services, including terminology mapping, versioning, and access. CTS2 was developed by HL7 in 2015 and revised in 2022. CTS published as “Service Functional Model Specification” for HL7 Common Terminology Services [2].

Choosing CTS2 as a data model can benefit certain contexts, particularly in domains where standardised representation and management of terminologies are essential [3, 4]. Here are some reasons why CTS2 might be chosen as a data model:

- **Standardisation:** CTS2 is a standard developed in the cooperation of the Object Management Group (OMG) [5] and Health Level 7 (HL7). Choosing CTS2 as a data model ensures adherence to a widely recognised and accepted standard for representing terminologies. Standardisation can promote consistency and interoperability across different systems and applications [6].
- **Interoperability:** CTS2 design supports interoperability by providing a common framework for representing and accessing terminological content. Choosing CTS2 can facilitate seamless integration and data exchange between healthcare systems and applications that need to work with terminologies [4, 6].
- **Flexibility:** CTS2 design is flexible and accommodates various terminologies and code systems. This flexibility is valuable when different organisations or systems use different terminologies and a unified approach is needed to integrate and manage them [6].
- **Mapping and Cross-referencing:** CTS2 supports mapping and cross-referencing of concepts across different terminologies. If your data model requires the ability to link or translate between different coding schemes, CTS2 provides mechanisms for achieving this [6].
- **Versioning:** CTS2 supports versioning of terminologies, allowing for tracking changes over time. Versioning is crucial when terminologies are regularly updated or revised, ensuring data remains accurate and aligned with the latest standards [6].

- **Community Adoption:** If your organisation is part of a community or industry that has widely adopted CTS2, choosing it as a data model can facilitate collaboration and interoperability with other stakeholders in the same domain [5].
- **Integration with Health Information Systems:** CTS2 design integrates with health information systems. If your data model needs to align with healthcare standards and terminology requirements, CTS2 can be a suitable choice [4].

1.2 PostgreSQL as the Database Management Platform

PostgreSQL is a leader among open-source relational database management systems (RDBMS). It provides data integrity, security, advanced search options, and JSONB support. PostgreSQL is a compelling option for healthcare institutions seeking a stable and scalable solution for managing clinical terminology [7, 8].

1.3 TermX

TermX is a unique open-source platform for terminology and knowledge management. The main goals of developing TermX were adhering to the latest standards, enhancing import capabilities for unmatched flexibility, and supporting FHIR. TermX is the only terminology server that supports multilingual clinical terminology, multilingual resource descriptions, and a multilingual web user interface. Furthermore, internal data models are meticulously aligned with widely accepted standards, like CTS2, to ensure seamless integration and compatibility.

1.4 FHIR Terminology Module

The FHIR terminology model defines how codes and terminologies are used within the FHIR specification [2]. The HL7 FHIR terminology module provides a common language for terminology representation. By adopting FHIR resources, organisations can overcome the heterogeneity of internal data models and establish a uniform framework for information exchange. This module includes a comprehensive set of resources, such as CodeSystem, ValueSet, and ConceptMap, which facilitate semantic interoperability and data consistency. Transforming internal data models to FHIR resources promotes semantic consistency across healthcare systems. This consistency ensures that data elements are interpreted uniformly. While adopting FHIR terminology modules offers significant benefits, the transformation process has challenges in data mapping complexity, resource versioning, updates, and customisation.

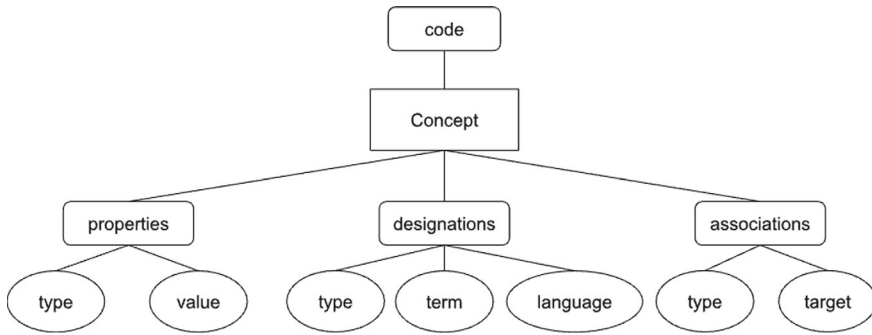


Fig. 1 Concept diagram

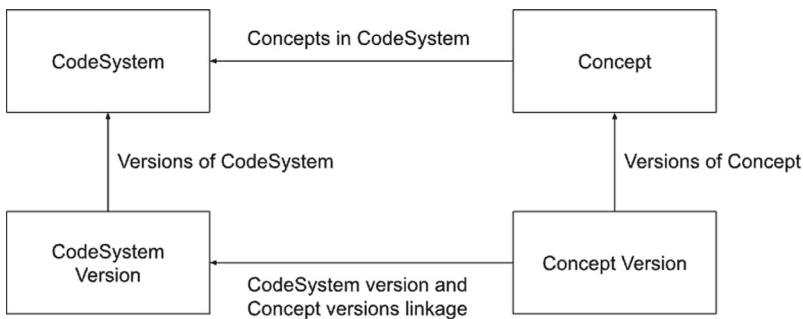


Fig. 2 Versioning model in a code system

1.5 Building Blocks of the Terminology Module

A **concept** is a unique and distinct idea or entity within the realm of health and medicine that can be precisely defined and represented. Concepts in clinical terminology are often abstract and cover a wide range of topics, including diseases, symptoms, procedures, medications, observations, and other healthcare-related elements. Figure 1 represents necessary information about the use of concept entities like code, properties, designations, and associations.

A **code system** is a standardised set of codes or symbols representing and communicating various concepts. These codes are essential for organising and categorising health data in a consistent and structured manner, facilitating the exchange of information between healthcare professionals, systems, and organisations.

Versioning involves managing and tracking changes to individual concepts and entire code systems. Reasons for concept versioning can be updates in medical knowledge, improvements, clarifications, or incorporating new concepts. Figure 2 shows how the concept versions will be associated with the code system versions.

2 Methodology

In adherence to the principles of Design Science Research Methodology (DSRM), we initiated the TermX Server’s design and development process. Meticulously crafting the server, we leveraged standards such as Fast Healthcare Interoperability Resources (HL7 FHIR) and HL7 Common Terminology Services 2 (CTS2) to ensure compatibility and interoperability.

A crucial step involved selecting an appropriate database management system, leading us to choose the relational database management system PostgreSQL for its compatibility with CTS2 and FHIR.

While integrating the CTS2 conceptual model for PostgreSQL, we focused on incorporating the requirements and constraints of the FHIR terminology module to meet established standards.

Upon integrating the CTS2 conceptual model, we compared it to the FHIR conceptual model and developed a transformation process. We actively executed testing and validation processes to guarantee the seamless integration of CTS2 and FHIR.

Throughout the development process, we proactively maintained comprehensive documentation of design knowledge. This documentation encapsulated crucial design decisions and challenges encountered.

3 Results

3.1 *CTS2 Conceptual Model*

The conceptual model of the CTS2 contains 30 entities. Figure 3 demonstrates the conceptual model of CTS2. The entities marked with bold frames relate to the code system.

3.2 *FHIR Terminology Conceptual Model*

The conceptual model of the FHIR Terminology contains 4 entities. Figure 4 demonstrates the conceptual model of the FHIR terminology model.

3.3 *The Comparison of CTS2 and FHIR Conceptual Models*

In examining the conceptual models of CTS2 and FHIR, notable distinctions that underscore their respective strengths and applications emerge. In FHIR, the absence of divisions into code systems and versions is a distinctive feature. Unlike CTS2,

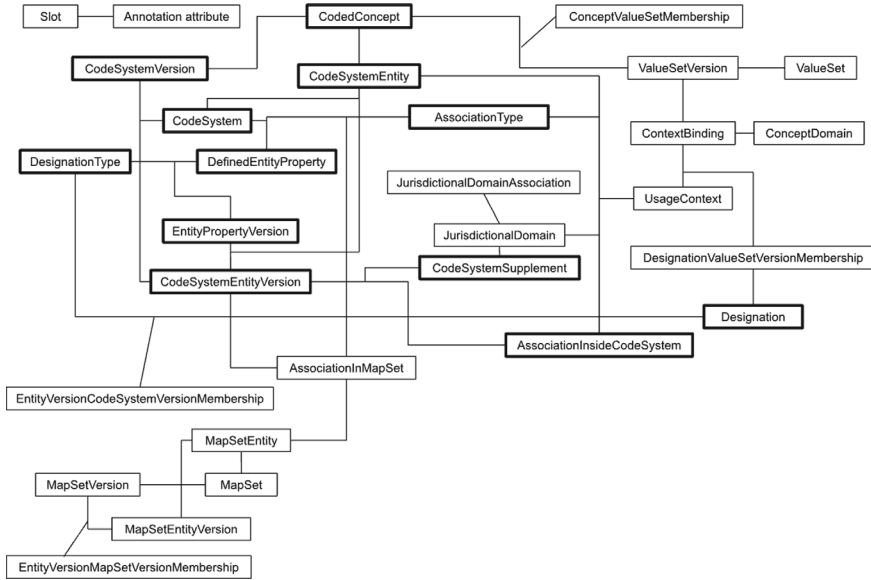
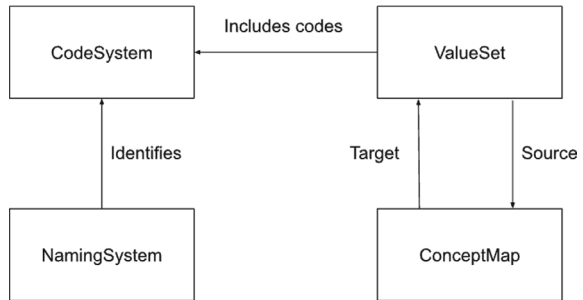


Fig. 3 Entities of the CTS2 conceptual model

Fig. 4 FHIR terminology conceptual model



where each version is treated as a separate resource with dedicated organisational structures, FHIR opts for a model where each version exists as an independent resource. Additionally, FHIR needs a dedicated concept resource and needs to incorporate the notion of concept versioning. Consequently, CTS2 stands out as a more comprehensive system, providing a robust version and full authoring capabilities support. The consequence of these differences is evident

- CTS2 is a more complete system suitable for tasks involving versioning and comprehensive authoring. At the same time, FHIR is predominantly utilised for the streamlined publication of finished versions.

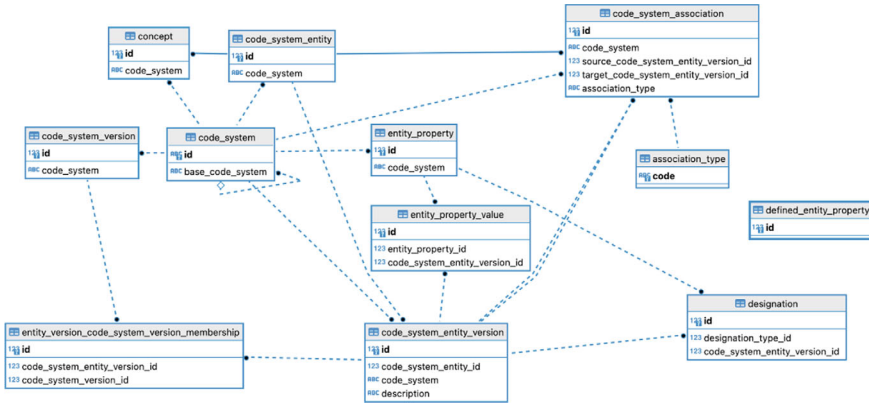


Fig. 5 TermX database model of code system entity

3.4 Database Design

Choosing the CTS2 model as the foundation for our database design was a strategic decision based on a thorough comparison. Illustrated in Fig. 5, our database structure mirrors the design of code system entities in the relational database. We meticulously incorporated the CTS2 inherent structure, relationships, and entity names into our database design process. Although our use cases did not necessitate the complete implementation of the CTS2 concept model, we deliberately omitted certain entities (JurisdictionalDomain, UsageContext, ConceptDomain, etc.). At the same time, all unused entities can be added as needed. To enhance comprehensibility and readability, we streamlined some of the names, as the intricacies of the CTS2 concept model can be overwhelming. For a more robust querying experience, we introduced new columns and relationships. Moreover, we incorporated additional columns to align with FHIR resources, ensuring a seamless correspondence and meeting the requirements of essential fields.

Figure 6 shows an example of database data based on the concept from the Logical Observation Identifiers, Names, and Codes (LOINC) code system.

3.5 CTS2 Data Model to FHIR Terminology Resources

As we have shown in comparing CTS2 and FHIR conceptual modules, the FHIR terminology module is more primitive. During the decision-making process for mapping attributes from CTS2 to FHIR, our approach was guided by the recognition that the FHIR model can be considered a subset of the CTS2 model. We aimed to identify the most suitable CTS2 attributes that align with the contextual meaning of the corresponding FHIR attributes. In cases where a direct match was not found, the flexibility of the CTS2 model allowed us to add new attributes specifically tailored

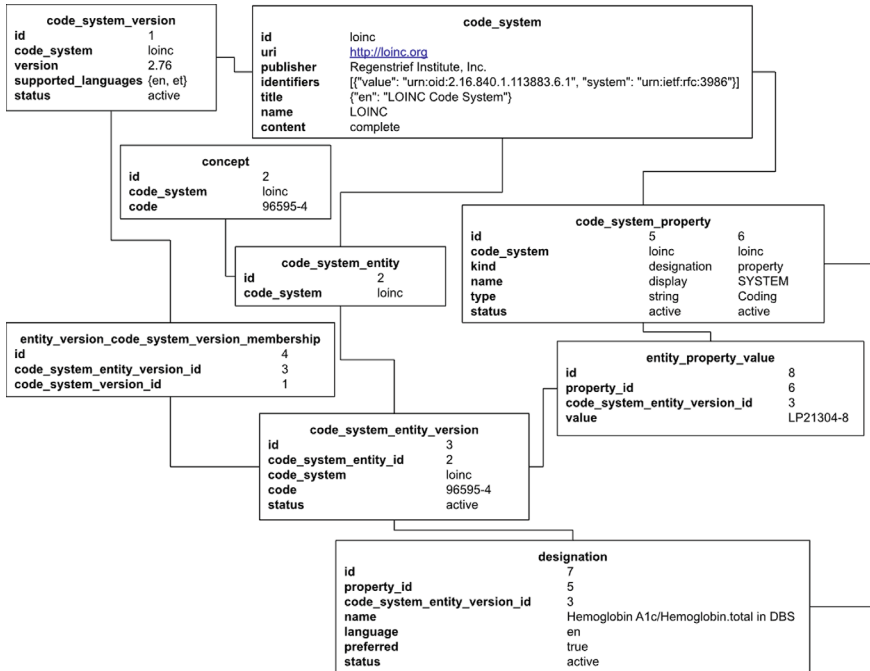


Fig. 6 Code system entity example based on LOINC

to meet the requirements of the FHIR specification. Such attributes include resource URI, both computer- and human-friendly names of the resource, resource authoring information (author, editor, reviewer, endorser), and purpose description of resources. During CTS2 conceptual module implementation in the database and FHIR resource analysis, we added all attributes required by FHIR specification. Model extension methodology works unidirectionally, as the FHIR model is strictly defined. As a result, there were not problems with adding FHIR interfaces on top of the extended CTS2 database model. The implemented interfaces for code systems include:

- GET, POST, PUT, DELETE actions for CodeSystem
- HL7 FHIR terminology module operations for CodeSystem: *lookup, validate-code, subsumes, find-matches*

4 Discussion

This discussion segment ventures into the nuanced landscape of challenges encountered during the development of TermX. In this chapter, we unravel specific considerations that have emerged during our journey, addressing intricacies such as resource

deletion strategies, versioning complexities between CTS2 and FHIR, flexible data handling capabilities inherent in the CTS2 model, and integrating FHIR constraints.

Implementing resource deletion, particularly for Code Systems, is crucial to ensuring data integrity. The decision to opt for logical deletion, where the resource is marked as deleted but retains its data in the database, is a thoughtful approach. This strategy mitigates the risk of accidental data loss and aligns with best practices in data management. Allowing users to reuse the code of a deleted Code System introduces a layer of flexibility and error prevention.

The disparities between FHIR and CTS2 data models pose challenges, particularly in handling versioning. While FHIR treats each version as a separate resource, CTS2 considers version resources independent entities from the Code System resource. The need to map version codes to FHIR introduces complexities, raising questions about the efficiency of this mapping process and its impact on data consistency.

The discussion on CTS2's flexibility in accommodating data in different formats, such as JSON structures, is noteworthy. The ability to handle diverse languages for fields like title, purpose, and description enhances usability, mainly when the terminology server caters to a wide range of specialists speaking different languages.

Identifying and adding attributes required by FHIR to the CTS2 model for appropriate conversion is critical in ensuring seamless interoperability between the two standards. Also, incorporating FHIR constraints, such as warnings and rules, into the management processes of terminology resources contributes to the overall robustness and reliability of the terminology server.

In conclusion, the challenges and solutions discussed above highlight the intricate nature of implementing a terminology server that aligns with CTS2 and FHIR standards. Future considerations involve refining these implementations based on ongoing feedback, evolving standards, and emerging requirements in the dynamic landscape of healthcare interoperability.

5 Conclusion

In the dynamic realm of healthcare data management, the development and implementation of TermX, an open-source terminology server, marks a significant stride towards achieving comprehensive interoperability and data standardisation. This article has delved into the intricacies of implementing the Common Terminology Services 2 (CTS2) standard on PostgreSQL while integrating with HL7 Fast Healthcare Interoperability Resources (FHIR). As we conclude this exploration, several vital takeaways emerge:

- *Standardisation and Interoperability:* Implementing CTS2 as a data model ensures adherence to widely recognised standards, fostering consistency and interoperability across diverse healthcare systems and applications. CTS2's flexibility and support for mapping contribute to seamless integration, addressing the complexities of varied coding schemes.

- *Database Management with PostgreSQL*: The strategic choice of PostgreSQL as the database management platform underscores the commitment to data integrity, security, and accessibility in healthcare settings. Using PostgreSQL's advanced features, including indexing options, complex query support, and JSONB capabilities, enhances the efficiency and flexibility of clinical terminology management.
- *TermX Innovation and FHIR Integration*: TermX stands out as an innovative platform, embracing the latest standards, enhancing import capabilities, and supporting FHIR to meet the evolving demands of modern healthcare. Integrating FHIR resources, including CodeSystem, ValueSet, and ConceptMap, underscores TermX's commitment to semantic interoperability and data consistency.
- *Addressing Versioning Challenges*: The discussion on versioning complexities between CTS2 and FHIR models highlights the nuanced nature of terminology management. Considering logical deletion, code system reuse, and the challenges in mapping version codes between standards reflects a commitment to comprehensive data governance and error prevention.
- *CTS2 Flexibility and Attribute Mapping*: The recognition of CTS2's flexibility in handling data in different formats and languages showcases a forward-looking approach to catering to a diverse user base. Attribute mapping to accommodate FHIR requirements demonstrates a meticulous effort to align with standards and ensure seamless conversion between data models.
- *FHIR Constraints for Robust Resource Management*: Incorporating FHIR constraints, including warnings and rules, in the management processes of terminology resources adds an extra layer of reliability. This proactive approach enhances the robustness of the terminology server, contributing to the overall quality and consistency of healthcare data.

TermX emerges as a technological solution and a testament to the commitment to advancing healthcare interoperability. Thus, TermX can synchronise its resources via GitHub or directly with other terminology servers using the FHIR API. The comprehensive exploration of its implementation, from the choice of standards to database design and FHIR integration, lays the foundation for future innovations in clinical terminology management. Today, the centres for the development of medical standards in Lithuania and Estonia have become convinced of the effectiveness of the TermX solution and have adopted it as the primary means of describing and publishing terminology. As the healthcare landscape evolves, TermX stands ready to adapt, ensuring it remains at the forefront of facilitating a seamless, interoperable, standardised healthcare ecosystem.

Code Availability The TermX terminology server has been published as an open-source software on GitLab [9] and integrated into the TermX interoperability platform [10].

Acknowledgements M.I. and I.B. designed the idea for the manuscript, and M.I. wrote it with support from I.B. All authors contributed to the final version. G.P. supervised the project. M.I. and I.B. designed the solution and architecture of the TermX terminology server, which M.I. developed with the support of Daniel Dubrovski and I.B.

References

1. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S (2019) Why digital medicine depends on interoperability. *NPJ Digital Med* 2(1):79
2. HL7, “Fhir terminology module” (2023). <http://hl7.org/fhir/terminology-module.html>
3. Blobel B et al (2022) Cts2 owl: mapping owl ontologies to cts2 terminology resources. In: *PHealth 2022: proceedings of the 19th international conference on wearable micro and nano technologies for personalized health*, vol 299, IOS Press, p 44
4. Gazzarata R, Maggi N, Magnoni LD, Monteverde ME, Ruggiero C, Giacomini M (2021) Semantics management for a regional health information system in italy by cts2 and fhir. In: *Applying the FAIR principles to accelerate health research in Europe in the Post COVID-19 era: proceedings of the 2021 EFMI special topic conference*, vol 287, IOS Press, p 119
5. OMG (2023) Common terminology services 2. <https://www.omg.org/spec/CTS2>
6. HL7 (2022) H17 common terminology services. service functional model specification. release 2. http://www.hl7.org/documentcenter/private/standards/CTS/V3/CTS2_r2_2015FEB_R2022.pdf
7. Group TPGD (2023) About postgresql. <https://www.postgresql.org/about>
8. Austin T, Sun S, Lim YS, Nguyen D, Lea N, Tapuria A, Kalra D et al (2015) An electronic healthcare record server implemented in postgresql. *J Healthcare Eng* 6:325–344
9. Kodality, “Termx.” (2024). <https://gitlab.com/kodality/terminology>
10. Kodality, “Termx tutorial” (2023). <https://termx.kodality.dev/wiki/termx-tutorial/about>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



An Improved Technique for Generating Effective Noises of Adversarial Camera Stickers



Satoshi Okada and Takuho Mitsunaga

Abstract Cyber-physical systems (CPS) represent the integration of the physical world with digital technologies and are expected to change our everyday lives significantly. With the rapid development of CPS, the importance of artificial intelligence (AI) has been increasingly recognized. Concurrently, adversarial attacks that cause incorrect predictions in AI models have emerged as a new risk. They are no longer limited to digital data and now extend to the physical environment. Thus, they are pointed out to pose serious practical threats to CPS. In this paper, we focus on the “adversarial camera stickers attack,” a type of physical adversarial attack. This attack directly affixes adversarial noise to a camera lens. Since the adversarial noise perturbs various images captured by the camera, it must be universal. To realize more effective adversarial camera stickers, we propose a new method for generating more universal adversarial noise compared to previous research. We first reveal that the existing method for generating noise of adversarial camera stickers does not always lead to the creation of universal perturbations. Then, we address this drawback by improving the optimization problem. Furthermore, we implement our proposed method, achieving an attack success rate 2.5 times higher than existing methods. Our experiments prove the capability of our proposed method to generate more universal adversarial noises, highlighting its potential effectiveness in enhancing security measures against adversarial attacks in CPS.

Keywords Adversarial example · Machine learning · Cybersecurity

1 Introduction

Machine learning technologies have become more and more popular not only in research fields but also in the practical world. They are being implemented in

S. Okada (✉) · T. Mitsunaga
INIAD, Toyo University, Tokyo, Japan
e-mail: satoshi.okada@iniad.org

T. Mitsunaga
e-mail: takuho.mitsunaga@iniad.org

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_21

various applications such as image recognition, anomaly detection, text mining, and malware detection [1–4]. Among these machine learning algorithms, “deep learning,” in particular, has gathered significant attention due to its impressive capabilities, which are comparable or even superior to human abilities in tasks such as natural language processing and decision-making. Furthermore, this technology continues to evolve rapidly. The advancements have been made possible by the availability of large datasets for training neural networks and the remarkable progress in hardware technology [5].

However, we must be careful about the real-world applications of such deep learning techniques. This is because deep learning technology has not yet gained complete trust in terms of security [6–8]. Recent studies have pointed out the vulnerability of machine learning models to adversarial attacks, where attackers induce incorrect predictions by perturbing the input data. Most of these adversarial attacks involve directly adding noise to the input digital data of the machine learning models. However, recently, several attacks have been proposed that extend these attacks to the physical world [9–12], such as stickers, to deceive visual sensors like webcams. Such physical adversarial examples are easier to apply in the real-world compared to digital manipulations and could be used to deceive AI systems in actual environments. Therefore, developing defense mechanisms against physical adversarial examples is an important research area for ensuring the security of AI systems.

One type of physical adversarial attack, the adversarial camera stickers attack, has been proposed by Li et al. [12]. The stickers can be directly affixed to the camera lens itself and contain a pattern of elaborately crafted dots that induce the misclassification of all images viewed by the camera. In other words, this attack is expected to consistently cause the misclassification of various types of inputs (images) belonging to a specific class with a fixed noise. If this attack is conducted on a factory automation system and the sticker is affixed to a camera monitoring a production line, there is a risk that a defective product will be misclassified as a normal product. Such a kind of attack poses a more serious threat if the perturbation is more universal. Li et al. proposed an optimization problem to be solved to construct universal noise. Furthermore, they implemented the optimized noise and measured the impact of their proposed method on a machine learning model (ResNet-50 [13]).

1.1 Contribution

In this paper, we propose a new method for generating more universal noise of adversarial camera stickers than Li et al. We first point out the drawbacks of Li et al.’s optimization problem. Specifically, we logically point out that solving the optimization problem proposed by Li et al. does not necessarily lead to the generation of truly universal noise. To address this issue, we modify the objective function of the optimization problem and propose a new method for generating more universal noise. Furthermore, we implement our proposed method and conduct a comparative analysis with the adversarial noise generated by Li et al. The results of this comparison reveal

that our proposed method achieves an attack success rate that is five times higher than that of the previous research. This improvement indicates that our method is capable of generating more universally effective noise.

1.2 Organization of the Paper

We explain adversarial attacks and introduce related research in Sect. 2. Section 3 describes our proposals. We provide the experimental results and discussion in Sect. 4. Section 5 concludes this paper.

2 Adversarial Attacks

In this section, we first explain adversarial attacks, classify them into digital and physical attacks, and introduce previous studies on adversarial attacks. We then summarize adversarial camera sticker attacks, closely related to our research.

2.1 Explanations of Adversarial Attacks

Adversarial attack causes misclassification in a machine learning model by manipulating input data. The attacker successfully manipulates the input data to cross a decision boundary, causing that input data to be misclassified or controlling for unintended predictions. This attack formula can be formulated as follows [14]:

$$\begin{aligned} & \text{minimize} && \|x' - x\| \\ & \text{subject to} && f(x') = l', \\ & && f(x) = l, \\ & && l \neq l', \\ & && x' \in [0, 1]^m, \end{aligned}$$

where $x \in [0, 1]^m$ is an input to a classifier f , l is correct predicted class for x , and $l' \neq l$ is target class for $x + r$, $r \in [0, 1]^m$ - small perturbation to x .

If attackers intend to misguide the model to a specific class other than the true class, this attack is called a “targeted attack.ad.” If attackers try to cause the model to predict any of the incorrect classes, it is called an “untargeted attack.” Additionally, adversarial attacks are also classified based on the information available to the attacker. If an attacker knows all information, including input and output data, as well as the weights and classification labels of the target model, the attack is called a

“white-box attack.” On the other hand, an attack conducted under conditions where the attacker only has access to information about the input/output data is called a “black-box attack.”

Adversarial attacks are typically categorized into two types: digital world attacks and physical world attacks. Digital world attacks directly injected perturbations into the input data. Meanwhile, physical world attacks manipulate actual objects to induce misjudgments in AI systems [15].

2.2 *Related Research 1: Digital World Attacks*

Digital adversarial attacks, particularly in the realm of deep learning, have been extensively examined. This was initially highlighted by Szegedy et al. [14] in 2013, who employed the box-constrained L-BFGS method to identify perturbations. Following this, Goodfellow et al. [16] introduced the FGSM, a straightforward attack strategy. This method applies perturbations in the direction within the image space that maximizes the increase of the linearized cost. Building on this concept, Kurakin [17] developed an iterative, multi-step variation of this approach. This idea was expanded into a different norm space [18]. These methodologies, alongside variants like those explored by Moosavi-Dezfooli et al. [19], represented the forefront of current digital attack strategies.

2.3 *Related Research 2: Physical World Attacks*

In this section, we discuss various existing studies on physical adversarial attacks. For example, there have been some strategies to deceive detectors using carefully crafted items such as glasses [20] and T-shirts [21]. Other research created adversarial objects in reality, including printed images [17], 3D objects [11], and roadside objects [22]. Others caused misjudgments by affixing physical noise to the objects recognized by AI [10, 23–25].

Adversarial Camera Stickers Attacks In this study, we especially focused on the adversarial camera stickers proposed by Li et al. [12] among physical world attacks. In this attack, perturbations are calculated to cause the misclassification of AI. Then, they are printed as an adversarial sticker and directly affixed to the camera lens. Thus, the perturbations are able to cause the targeted deep learning model to incorrectly classify all the images the camera captures through its lens.

The proposed perturbations consist of small dots on a sticker. A single dot π_0 is formulated in the following equations:

$$\begin{aligned}\pi_0(x; \theta)(i, j) &= (1 - \alpha(i, j)) \cdot x(i, j) + \alpha(i, j) \cdot \gamma \\ \alpha(i, j) &= \alpha_{\max} \cdot \exp(-d(i, j)^\beta) \\ d(i, j) &= \frac{(i - i^{(c)})^2 + (j - j^{(c)})^2}{r^2}.\end{aligned}$$

There are five types of parameters which decide the above equations:

$$\theta = (\gamma, (i^{(c)}, j^{(c)}), r, \alpha_{\max}, \beta) \quad (1)$$

The explanations of them are the following:

- $\gamma \in [0, 1]^3$: RGB color
- $(i^{(c)}, j^{(c)}) \in \mathbb{R}^2$: center location (pixel coordinates)
- $r \in \mathbb{R}_+$: effective radius
- $\alpha_{\max} \in [0, 1]$: maximum alpha blending value
- $\beta \in \mathbb{R}_+$: exponential dropoff of alpha value

Adversarial camera stickers induce consistent misclassification across different inputs (images) that belong to a particular target class through a fixed adversarial perturbation. This means that adversarial stickers must keep a high attack success rate regardless of the diverse angles, scales, and lighting conditions of captured images. The more a certain noise leads to the misclassification of a variety of images belonging to the target class, the more universal—and thus effective—that noise is. However, the optimization methods for noise proposed by Li et al. are not always suitable for generating universal noise. In this paper, we point out the drawbacks of the previous work’s method and propose a new approach to find better (more universal) noise.

3 Our Proposed Method

In this section, first, we identify the drawbacks of the existing methods. Subsequently, we introduce how we address them and propose a new method for generating more universal noise for adversarial camera stickers.

3.1 Drawback of Previous Work

Li et al. proposed an optimization problem to find a “universal” adversarial noise of adversarial camera stickers. Let ℓ represent the loss function, f the classifier, and π the adversarial noise. If $x^{(1)}, \dots, x^{(M)}$ as M images of the y^* class are perturbed to be misclassified as y_{targ} , the optimization problem is formulated as follows.

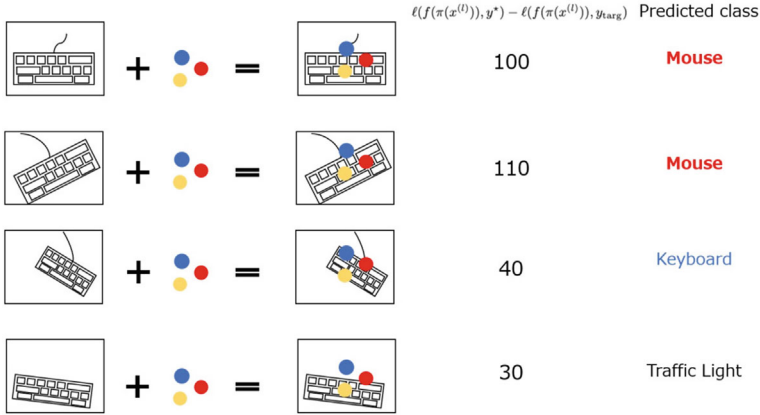


Fig. 1 Performance of noise under optimization

$$\max_{\pi} \sum_{l=1}^M (\ell(f(\pi(x^{(l)})), y^*) - \ell(f(\pi(x^{(l)})), y_{\text{targ}})) \quad (2)$$

We point out that this optimization problem is insufficient. For instance, consider the creation of adversarial camera stickers that misclassify images of the “computer keyboard” class as “mouse.” For simplicity, we assume $M = 4$. While maximizing Eq. 2, a noise like the one in Fig. 1 is generated. At this point, the value of Eq. 2 would be $280 (= 100 + 110 + 40 + 30)$, and the number of successful target attacks (images misclassified as “mouse”) would be two. Since the value of Eq. 2 has not yet reached its maximum, the optimization process continues. Eventually, the noise evolved as shown in Fig. 2, making Eq. 2 reach $300 (= 150 + 80 + 30 + 40)$, an increase of 20 from the previous. Meanwhile, the number of successful target attacks becomes one. This shows that the attack success rate decreases despite optimizing Eq. 2. That is because the greedily increasing the value of Eq. 2 led to the noise being more optimized for images whose value of $\ell(f(\pi(x^{(l)})), y^*) - \ell(f(\pi(x^{(l)})), y_{\text{targ}})$ is more likely to get large (in this example, the top keyboard image in Fig. 2) than others. Through the analyses, we hypothesize that optimizing Eq. 2 does not necessarily lead to the generation of better noise.

3.2 Details of Our Proposed Method

To address the problem of the previous work, we propose a new optimization problem for generating more universal adversarial noise. In the context of adversarial camera stickers, more “universal” noise means that they can cause intended misclassification against a larger number of images in the targeted class (y^* in Eq. 2). Considering it, we establish the following new optimization problem:

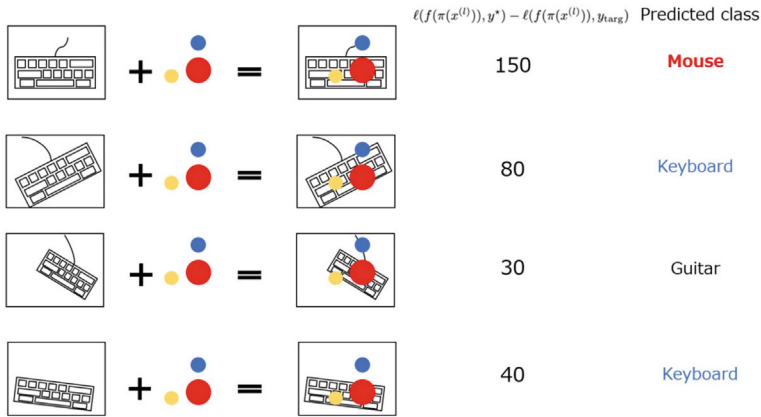


Fig. 2 After optimization: performance of adversarial examples

$$\max_{\pi} \sum_{l=1}^M \mathbb{1}[f(\pi(x^{(l)})) = y_{\text{targ}}] \tag{3}$$

$\mathbb{1}[\cdot]$ represents the indicator function, yielding a value of 1 if the condition within is satisfied. Specifically, it outputs 1 if the perturbed image is classified as the target class y_{targ} and 0 otherwise. The goal of our proposed method, therefore, is to identify a perturbation π that maximizes the count of such successful misclassifications, effectively maximizing the number of images that are transformed into the target class under the perturbation. By altering the optimization problem in this way, we can prevent the noise from becoming excessively specialized to certain specific images and thus becoming less universal.

In our proposed method, we first try to solve the optimization problem 2. Then, during the intermediate stages of this optimization, we select the noise that achieved the highest number of successful target attacks on the training images.

4 Evaluation Results and Discussion

We implemented our proposed method and assessed whether it could generate more effective (and more universal) adversarial noises. In this section, we first explain the methodology of our experimental setup and flow. Following this, we present our experimental results and discuss them.

4.1 Experimental Setup

In this experiment, we simulated the adversarial camera stickers in the digital world fooling a ResNet-50 [13]. To maintain consistency with the fooled model, we used validation images of ILSVRC2012 [26] when we optimized adversarial noises and tested their performances. We implemented our proposed method based on the code available on GitHub [27], which is the essential version of the previous research method [12]. We chose it because the complete original codes of the previous work were not available. Furthermore, since some important parameter settings were not available either, we set the parameters of noise (Eq. 1) as follows:

- Parameters to be optimized: $\gamma, (i^{(c)}, j^{(c)})$
- Fixed parameters that are not optimized: $(r, \alpha_{\max}, \beta) = (25, 1.9, 1.9)$.

4.2 Experimental Flow

In this experiment, we dealt with the targeted attack that caused misclassification from “computer keyboard” to “mouse.” Initially, we equally divided the dataset of the “computer keyboard” class from ILSVRC2012 into two random sets: training data and test data. Each of them was utilized for optimizing and validating the adversarial noise. Following the method proposed by Li et al. [12], we optimized the adversarial noise. Specifically, we continued the optimization process until the value of Eq. 2 (called “total loss”) saturated—when the difference between the loss values from one epoch to the next was less than 0.001. After the whole optimization process, we obtained the noise whose total loss was minimized. It was regarded as the noise optimized with Li et al.’s method. During this optimization process, we also identified the noise that caused most samples of training data to be misclassified as “mouse.” Since it satisfies our proposed optimization problem (Eq. 3), we selected it as our proposed adversarial noise. We then used the test data to compare the attack success rate between the noise optimized by Li et al.’s method and our proposed method.

4.3 Results

During the optimization process of the adversarial noise, we tracked both the transition of values defined by Eq. 2 and the number of training data samples that were misclassified as “mouse” due to the adversarial noise (called “mouse count”). The transitions of these values are depicted in Fig. 3. The x -axis represents the epoch number of the optimization process. The y -axis on the left, corresponding to the red line, represents total loss. Meanwhile, the y -axis on the right, corresponding to the blue line, shows mouse count. The optimization process was completed in 385 epochs, and the mouse count reached its maximum when the number of epochs was

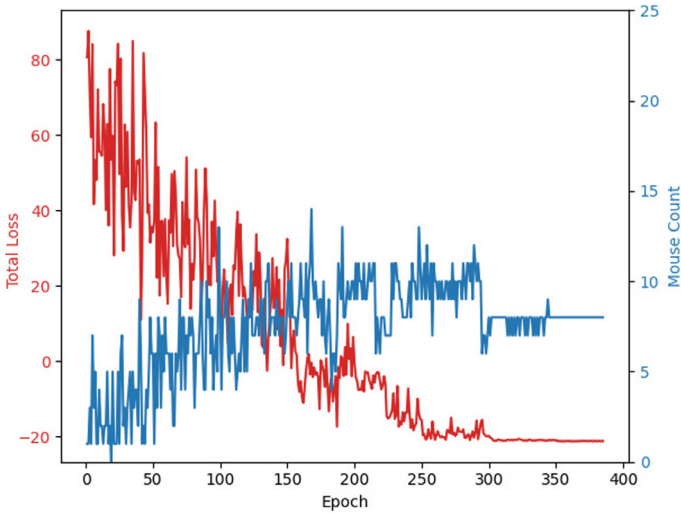


Fig. 3 Transition of total loss and mouse count

Table 1 Performance of each adversarial noise

Data	Adversarial noise	Prediction		
		Correct (%)	Target (%)	Other (%)
train	No	76	8	20
	Li et al. [12]	24	32	44
	Our proposed method	28	56	16
test	No	64	4	32
	Li et al. [12]	28	8	64
	Our proposed method	44	20	36

Bold shows author’s proposed method

168. At each of the epochs (385 and 168), we obtained the adversarial noise by Li et al. [12] and by our proposed method. The comparison of each noise’s performance is summarized in Table 1. The “correct” column shows how many samples the targeted model (ResNet-50) predicts correctly as “computer keyboard.” On the other hand, the “target” column shows how many samples are misclassified as “mouse.” The “other” means how many samples are classified as other classes than the correct and target label. We also described the evaluation results of each noise’s performance when using the test data in Table 1.

4.4 Discussion

As shown in Table 1, the noise generated by the proposed method achieved a higher attack success rate compared to previous studies on both training and test data. When

utilizing training data, the attack success rate was about 1.8 times higher than that of the existing methods, and when validated using test data, it was 2.5 times higher. This means that our proposed method generates more universal noises than Li et al.

Regarding computational complexity, the proposed method is the same as the existing method because the range of perturbations to be searched is identical. Since the existing method was intended to optimize the sum of loss functions (Eq. 2), more universal noise found during the optimization process is not highly valued. In contrast, the approach of our proposed method is intended to select universal noise. In other words, we achieved the generation of more universal noises without increasing computational complexity.

However, our study has room for improvement. A limitation of our study is that the experiments have been conducted in the digital world. Adversarial camera sticker attacks are physical world adversarial attacks. To ascertain the real-world applicability and effects of our proposed method, it is essential to physically create our proposed noises and measure their performance in actual environments. Adversarial training is generally effective as a countermeasure against such adversarial attacks. Since the noise in adversarial training should be universal, this research contributes to the development of robust AI models.

5 Conclusion

This paper proposes a new method for generating more effective (universal) noises for adversarial camera sticker attacks than previous research [12]. We pointed out that the previous method greedily minimizes loss in the adversarial noise optimization process, and it does not always lead to effective noise. Thus, we improved the optimization problem by focusing on maximizing the number of attack success cases during the optimization process. Furthermore, we implemented our proposed method and compared the attack success rate of our proposed noise with that of the previous work. As a result, it was clear that our proposed method generated more effective noise of adversarial camera sticker attacks.

References

1. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, IEEE Computer Society, pp 1026–1034
2. Oshio T, Okada S, Mitsunaga T (2022) Machine learning-based anomaly detection in zigbee networks. In: 2022 IEEE international conference on computing (ICOCO), IEEE Computer Society, pp 259–263
3. Amrit C, Paauw T, Aly R, Lavric M (2017) Identifying child abuse through text mining and machine learning. *Expert Syst Appl* 88:402–418

4. Okada S, Fujimoto M, Matsuda W, Mitsunaga T (2021) Malware detection method using tree-based machine learning algorithms. In: 2021 IEEE international conference on computing (ICOCO), IEEE Computer Society, pp 103–108
5. Potok TE, Schuman CD, Young SR, Patton RM, Spedalieri FM, Liu J, Yao K, Rose GS, Chakma G (2018) A study of complex deep learning networks on high-performance, neuromorphic, and quantum computers. *ACM J Emerg Technol Comput Syst* 14(2):19:1–19:21 (2018)
6. Li C, Guo W, Sun SC, Al-Rubaye S, Tsourdos A (2020) Trustworthy deep learning in 6g-enabled mass autonomy: from concept to quality-of-trust key performance indicators. *IEEE Veh Technol Mag* 15(4):112–121
7. Karmakar GC, Chowdhury A, Das R, Kamruzzaman J, Islam SM (2021) Assessing trust level of a driverless car using deep learning. *IEEE Trans Intell Transp Syst* 22(7):4457–4466
8. Hassan MM, Hassan MR, Huda MS, de Albuquerque VHC (2021) A robust deep-learning-enabled trust-boundary protection for adversarial industrial iot environment. *IEEE Internet Things J* 8(12):9611–9621
9. Brown TB, Mané D, Roy A, Abadi M, Gilmer J (2017) Adversarial patch CoRR, vol abs/1712.09665
10. Eykholt K, Evtimov I, Fernandes E, Li B, Rahmati A, Xiao C, Prakash A, Kohno T, Song D (2018) Robust physical-world attacks on deep learning visual classification. In: 2018 IEEE conference on computer vision and pattern recognition, CVPR 2018, Salt Lake City, UT, USA, June 18–22, Computer Vision Foundation/IEEE Computer Society, pp 1625–1634
11. Athalye A, Engstrom L, Ilyas A, Kwok K (2018) Synthesizing robust adversarial examples. In: Dy JG, Krause A (eds) Proceedings of the 35th international conference on machine learning, ICML 2018, Stockholm, Sweden, July 10–15, vol 80 of Proceedings of Machine Learning Research, PMLR, 2018, pp 284–293
12. Li J, Schmidt FR, Kolter JZ (2019) Adversarial camera stickers: a physical camera-based attack on deep learning systems. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th international conference on machine learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA, vol. 97 of Proceedings of machine learning research, PMLR, pp 3896–3904
13. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30. IEEE Computer Society pp 770–778
14. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow IJ, Fergus R (2014) Intriguing properties of neural networks. In: Bengio Y, LeCun Y (eds) 2nd international conference on learning representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014, Conference Track Proceedings
15. Bajaj A, Vishwakarma DK (2023) A state-of-the-art review on adversarial machine learning in image classification. *Multimedia Tools Appl* 1–66
16. Goodfellow IJ, Shlens J, Szegedy C (2015) Explaining and harnessing adversarial examples. In: Bengio Y, LeCun Y (eds) 3rd international conference on learning representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference track proceedings
17. Kurakin A, Goodfellow IJ, Bengio S (2017) Adversarial examples in the physical world. In: 5th international conference on learning representations, ICLR 2017, Toulon, France, April 24–26, 2017, Workshop Track Proceedings, OpenReview.net
18. Madry A, Makelov A, Schmidt L, Tsipras D, Vladu A (2018) Towards deep learning models resistant to adversarial attacks. In 6th international conference on learning representations, ICLR 2018, Vancouver, BC, Canada, April 30 to May 3, Conference Track Proceedings, OpenReview.net
19. Moosavi-Dezfooli S, Fawzi A, Frossard P (2016) Deepfool: a simple and accurate method to fool deep neural networks. In: 2016 IEEE conference on computer vision and pattern recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, IEEE Computer Society, pp 2574–2582
20. Sharif M, Bhagavatula S, Bauer L, Reiter MK (2016) Accessorize to a crime: real and stealthy attacks on state-of-the-art face recognition. In: Weippl ER, Katzenbeisser S, Kruegel C, Myers AC, Halevi S (eds) Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, Vienna, Austria, October 24–28, ACM, pp 1528–1540

21. Xu K, Zhang G, Liu S, Fan Q, Sun M, Chen H, Chen P, Wang Y, Lin X (2020) Adversarial t-shirt! evading person detectors in a physical world. In: Vedaldi A, Bischof H, Brox T, Frahm J (eds) Computer vision–ECCV 2020–16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V, vol 12350. Lecture Notes in Computer Science. Springer, pp 665–681
22. Kong Z, Guo J, Li A, Liu C (2020) Physgan: generating physical-world-resilient adversarial examples for autonomous driving. In: 2020 IEEE/CVF conference on computer vision and pattern recognition, CVPR 2020, Seattle, WA, USA, June 13–19, Computer Vision Foundation/IEEE, pp 14242–14251
23. Ryu G, Park H, Choi D (2021) Adversarial attacks by attaching noise markers on the face against deep face recognition. *J Inf Secur Appl* 60:102874
24. Wang Y, Lv H, Kuang X, Zhao G, Tan Y, Zhang Q, Hu J (2021) Towards a physical-world adversarial patch for blinding object detection models. *Inf Sci* 556:459–471
25. Kantaros Y, Carpenter TJ, Sridhar K, Yang Y, Lee I, Weimer J (2021) Real-time detectors for digital and physical adversarial inputs to perception systems. In: Maggio M, Weimer J, Farque MA, Oishi M (eds) ICCPS '21: ACM/IEEE 12th international conference on cyber-physical systems, Nashville, Tennessee, USA, May 19–21, ACM, pp 67–76
26. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis (IJCV)* 115(3):211–252
27. Github–yoheikikuta/adversarial-camera-stickers: a very limited implementation of [arxiv:1904.00759](https://arxiv.org/abs/1904.00759)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Artificial Intelligence in Breast Cancer Diagnosis: “Synergy-Net” in Campania FESR-POR (European Fund of Regional Development—Regional Operative Program) Research Project



Domenico Parmeggiani, Giancarlo Moccia, Pasquale Luongo, Francesco Miele, Alfredo Allaria, Francesco Torelli, Stefano Marrone, Michela Gravina, Carlo Sansone, Ruggiero Bollino, Roberto Ruggiero, Paola Bassi, Antonella Sciarra, Simona Parisi, Francesca Fisone, Maddalena Claudia Donnarumma, Chiara Colonnese, Paola Della Monica, Marina Di Domenico, Ludovico Docimo, and Massimo Agresti

Abstract *Background:* Nowadays, mammography and DCE-MRI are the gold standard for breast cancer screening. Scientific experience demonstrate an increasing application of Echo elastosonography in breast cancer diagnosis, for this reason, within the Synergy-Net project the objective is to create a system based on machine

D. Parmeggiani (✉) · G. Moccia · P. Luongo · F. Miele · A. Allaria · F. Torelli · P. Bassi · A. Sciarra · S. Parisi · M. C. Donnarumma · C. Colonnese · M. Agresti
Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy
e-mail: domenico.parmeggiani@unicampania.it

S. Marrone · L. Docimo
Department of Woman, Child and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy

M. Gravina · C. Sansone
Department of Electrical Engineering and Information Technology (DIETI), University “Federico II”, Naples, Italy

C. Sansone
CINI, ITEM Laboratory “C.Savy”, Via Cintia 21, Naples, Italy

R. Bollino · R. Ruggiero
Bollino.It SpA, Via Delle Ustrie 31, Naples, Italy

R. Ruggiero · S. Parisi · F. Fisone
Division of General, Mini-Invasive and Obesity Surgery, University of Campania “Luigi Vanvitelli”, 80131 OncologicalNaples, Italy

P. D. Monica · M. Di Domenico
Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, Naples, Italy

learning algorithms, a CAD developed with CNNs, capable of representing a decision support in the analysis of echo-elastic sonographic images. *Results*: 315 female patients at the “Vanvitelli” Breast Unit were subjected to digital 3D mammography tomosynthesis and advanced ultrasound (quantitative Echo elastosonography) and comparative analysis of the two methods. We collected 70 breast cancer, 200 benign pathologies (dropout of 11.1%) and we compared two different homogeneous by composition subgroups (35 with cancer and 100 with benign pathology 20.5%), relating to the diagnostic predictive capabilities of 3D digital mammography and quantitative elastosonography: Group A age 40–50 years, Group B age 50–60 years. Results demonstrate fair predictive performances of echoelastosonography versus traditional mammography, especially in the first group (40–50 years), $p < 0.05$ in favor of elastosonography regarding sensitivity and accuracy; performances that appear substantially comparable in the second group (50–60 years). We also tested the diagnostic predictive capabilities of the CAD Synergy-Breast-Net achieving encouraging results: Sensitivity 80%, Specificity 72%, Accuracy 74%, Negative Predictive Value 84.7% and Positive Predictive Value 83.3%. *Conclusion*: From this perspective, the screening of patients aged < 40 years and integrate the main screening activities in patients aged between 40 and 50 years integrated with increasingly high-performance machine learning systems, could represent a valid alternative in the future.

Keywords Breast cancer · Elastography · FibroScan · CNNs · Artificial intelligence

1 Introduction

This study is part of a ReS project (Synergy-Net: Research and Digital Solutions in the fight against oncological diseases). Its objective is the development of a Synergy-Net technological platform aimed at improving and strengthening oncological cancer prevention. In this study, we refer exclusively to experimental developments produced in the diagnosis of breast cancer.

Breast cancer is the most common neoplasm in women. In 2019, more than 50,000 women fell ill with this neoplasm in Italy. The reduction in the average age of onset of this neoplasm is also certain. This data affects screening diagnostics in two different ways: The first is that young women are not enrolled in screening programs and the second is that in young breasts clinical and instrumental diagnostics are more difficult due to the density of the breast. The reduction in sensitivity of mammography in dense breasts (classes C and D of mammographic density) is certain. Furthermore, if we consider that early diagnosis is an imperative to achieve patient recovery, we understand how breast diseases constitute a challenge for diagnostics. Added to this is the expansion of the audience of women to be called for screening also to the 40–50 years age group, with risk factors, in which the possibility of dense breasts is significantly greater. Hence, increasingly refined technologies in the study of the

dense breast, such as mammography with 3D digital tomosynthesis and mammography with contrast medium. The need therefore emerges that, in screening programs, reference cannot be made to mammography alone but that a complex strategy is necessary that can lead to personalized diagnostic paths based on different factors and also using different diagnostic techniques. Hence, increasingly refined technologies in the study of the dense breast, such as mammography with 3D digital tomosynthesis and mammography with contrast medium. The need therefore emerges that, in screening programs, reference cannot be made to mammography alone but that a complex strategy is necessary that can lead to personalized diagnostic paths based on different factors and also using different diagnostic techniques. Hence, increasingly refined technologies in the study of the dense breast, such as mammography with 3D digital tomosynthesis and mammography with contrast medium. The need therefore emerges that, in screening programs, reference cannot be made to mammography alone but that a complex strategy is necessary that can lead to personalized diagnostic paths based on different factors and also using different diagnostic techniques [1, 2].

It therefore appears necessary to plan on the basis of various parameters which can lead to a rationalization and, at the same time, a reduction in healthcare spending from investigations that are not useful for diagnosis purposes [3].

For this reason, we have tried to identify a diagnostic path that is not only instrumental but which can be “intelligent”, that is, integrated with the most modern machine learning technologies, capable of helping healthcare workers in choosing a diagnostic path-personalized therapy for each individual woman.

The primary objective is to reduce spending on unnecessary investigations as well as optimize knowledge in the field of breast cancer diseases.

Nowadays, mammography and contrast-enhanced magnetic resonance imaging (DCE-MRI) are the gold standard for breast cancer screening. In particular, DCE-MRI, digital mammography with 3D tomosynthesis and contrast-enhanced mammography are second-level diagnostic tests, used above all for their ability to capture physiological aspects of the tissues under analysis. Unfortunately, compared to mammography, they are very expensive and, its long acquisition time (RNM), irradiation (3D tomosynthesis) and the level of expertise required (contrast mammography) place limits on the number of patients who, every day, can be analyzed and therefore to their applicability in the screening phase. For this reason, within the Synergy-Net project the objective is to create an AI system that can allow the use of further technology, i.e., a CAD developed with CNNs, for the analysis of echo elastosonographic images, a rapid ultrasound method, painless and relatively inexpensive. This technology integrated into an online platform and with the progress of increasingly advanced and economical ultrasound technologies, could in the future represent the elective procedure in the screening of patients aged < 40 years and integrate the main screening activities in patients aged between 40 and 50 years. Painless and relatively inexpensive. This technology integrated into an online platform and with the progress of increasingly advanced and economical ultrasound technologies, could in the future represent the elective procedure in the screening of patients aged < 40 years and integrate the main screening activities in patients aged between 40 and 50 years. Painless and relatively inexpensive. This technology integrated into

an online platform and with the progress of increasingly advanced and economical ultrasound technologies, could in the future represent the elective procedure in the screening of patients aged < 40 years and integrate the main screening activities in patients aged between 40 and 50 years [4–10].

2 Materials and Methods

In a first phase, we evaluated on the basis of personal experience and literature data what anamnestic and instrumental information can lead to a more or less complex diagnostic path. We took into consideration several easily identifiable factors such as clinical criteria and the calculation of risk factors (for example with the Gail method) with particular emphasis on age, family history, nulliparity, age of the first pregnancy, previous surgery for cancer breast or for potentially evolving anomalies, hormonal therapies (stimulation) (Table 1).

Below we have taken into consideration specific data from traditional mammography with findings such as density in relation to age the detection of newly appearing mammographic opacities also in relation to age the type of micro-calcifications detected and their variation in number and extension in two subsequent tests the type of opacity detected in relation to age, the volumetric increase in the opacities in subsequent checks and, both clinical and mammographic findings, the alterations of the profile and skin edema (Tables 2, 3, and 4).

Based on our experience, we have given a multiplicative value in order to help engineers standardize and quantify a good part of the information provided to make

Table 1 Clinical criteria

	40–50 years	51–60 years
Profile alterations	X 1.5	X 2.5
Unilateral nipple retraction as a new symptom	X 1.5	X 2.5
Unilateral mono- or oligo-porous blood or serous secretion	X 1.5	X 2.0
Megalic axillary lymph nodes	X 1.5	X 2.0

Table 2 Mammographic density-classes C and D

Age	Value
40–50 years	1.5
51–60 years	2.5

Table 3 Newly appearing mammographic opacity compared to previous exams

Age	Value
40–50 years	2
51–60 years	3

Table 4

Micro-calcifications:
typology and variation in
number and extent

	Value
Not suspicious	1
Suspicious	2.5
Suspicious inscribed in irregular opacity	4
Increase in number and extension	4

an evaluation that is as reliable as possible, independent of individual interpretative variations. Once the data to be recovered and integrated with the images had been defined, the current limits of the screening protocols in use were defined, thus highlighting the diagnostic limitations in juvenile dense breasts. The multiplicative values are relative to:

We have therefore focused our attention on diagnostic images acquired with a latest generation ultrasound system: Supersonic Mach 30, in particular, is an ultrasound supported by a processor that works with Ultra-Fast technology, which has 5 times more computing power. Intelligent signal processing uses algorithms in image reconstruction, based on neural networks, capable of acquiring up to 20,000 images per second, furthermore TriVu imaging combines real-time simultaneous imaging of B-mode, 1 ShearWave™ PLUS elastography and Angio PLUS imaging, allowing doctors to view anatomy, tissue stiffness and blood flow on the same image simultaneously. This mode provides access to vascularity-guided measurement and tissue stiffness. This real-time technique is able to measure tissue stiffness in a non-invasive way, but above all it is able to provide a quantitative evaluation in Kilopascals (kPa) of tissue hardness, a fundamental element for limiting subjective variables, a crucial methodological bias in training of an ultrasound cad. Once the diagnostic protocols and the type of data to be acquired were defined, we began to work on the protocol and machine learning algorithms to apply to our study for the creation of a CAD developed with convolutional neural networks. We enrolled a group of 315 female patients aged between 35 and 72 at the Department of Advanced Medical and Surgical Sciences and the Breast Unit of the University of Campania “Luigi Vanvitelli”, from January 2021 until December 2021. Years, average age 55.4 years, subjected to digital mammography with 3D tomosynthesis and advanced ultrasound investigation (Echo with quantitative elastosonography) and we compared the predictive capabilities of the two methods: The patients were enrolled and subjected to Micro-histological Biopsy, RNM, staging investigations and subsequent GOM protocol with related indications for neoadjuvant therapy and/or surgical therapy. In the sample under examination we had a dropout of 11.1% due to lack of acquisition standardization methodology (some of the most common echographic dataset weaknesses) and dataset incompleteness (some of the most common CNNs weaknesses); we analyzed two different subgroups made homogeneous by composition (35 with cancer and 100 with benign pathology 20.5%), in relation to the diagnostic predictive capabilities of 3D digital mammography and quantitative elastosonography [11].

- Age 40–50 years.
- Age 50–60 years.

The technological platform created for Synergy-Net has acquired a copious dataset of clinical-anamnestic information and over 280 ultrasound scans acquired in Dicom format, with a first image clean of comments, indications and/or measurements, a second with reference to the site, in at least two dimensions, with perimetric reconstruction of the Region of Interest (ROI) and volumetric analysis of the lesion, a third image with geometric area (rectangle) including the lesion, analysis of the vascular microcirculation and in the same image contemporary analysis of the elastographic index of the lesion in kPa, with chromatic representation of the lesions (represented in a chromatic scale that goes in order from the hardest tissue to the softest, from the chromatic shades of red to the color blue). The images were saved with a patient identification number and a classification code relating to the USBirads classification provided by the operator (e.g., K1USbirads5eco or ecoelastovasc, etc.). After selecting the ROI from the echo-elastosonographies, a CNN-based system is used for classification.

In particular, AlexNet was used to distinguish nodules between different classes. To handle the small size of the dataset, the network involved is pre-trained on the ImageNet dataset and fine tuning is used to adapt the network to the specific task to be solved. In order to increase the size of the dataset, 5 square patches are extracted from the ROI highlighted during the elastosonography. The final prediction is carried out considering a vote among the labels relating to the extracted patches, weighted based on the probability of prediction.

The CAD-CNNs Synergy-Breast-Net created was finally tested in the field during presentation days of the Project at a local level in a high incidence area, on 190 ultrasound images of approximately 100 pre-selected patients, of which 35 presented highly suspicious nodules, all of which were then confirmed by micro-histology as malignant.

This sample of 35 patients with malignant neoplasms was subjected to a complete analysis of the entire genome, approximately 46,000 variables/patient, which enriched the amount of data of the expert system developed and for which, at the moment, no data is yet available relevant for potential correlations.

- The first objective of the study is to verify in patients, based on the two different age classes, the predictive performance of echo-elastography compared to digital mammography.
- The second objective is therefore to confirm, at least preliminarily, the predictive performances of the CAD-CNNs Synergy-Breast-Net.

3 Results

Performance is evaluated in terms of Accuracy (ACC), Specificity (SPE) and Sensitivity (SENS). Positive Predictive Value (PPV) and Negative Predictive Value (NPV) and in relation to the first objective the results certainly appear interesting, demonstrating fair predictive performances of ultrasound elastosonography vs traditional mammography, especially in the first group (40–50 years), with statistically significant differences ($p < 0.05$) in favor of elastosonography regarding sensitivity and accuracy; performances that appear substantially comparable (statistically insignificant differences) in the second group (50–60 years) an interesting predictive element in elastosonography is the presence of nodular elastosonographic values above 100 kPa (Table 5).

Regarding the second objective, the result, although largely conditioned by the numerical limitation of the sample, appears promising.

The performances of the CAD Synergy-Net are all statistically significantly worse than both the mammographic and elastosonographic performances (if a statistically significant difference is excluded not significant in relation to sensitivity compared to digital mammography in group I) (Table 5). The non-exceptional performances largely depend on the small size of the available dataset, but can certainly be improved by integration with the clinical-anamnestic dataset, by numerical implementation and the consequent completion of the learning curve. Integration with gene profiling of cancer patients could also help reveal new development and research scenarios in the future (Table 6).

Table 5 Diagnostic performance between mammography and elastosonography based on age

Study	Mammography 40–50 years	Eco-Elasto 40–50 years	Mammography 51–60 years	Eco-Elasto 51–60 years
Sensitivity	28/35 (80.0%)	32/35 (91.4%)	30/35 (85.7%)	31/35 (88.5%)
Specificity	91/100 (91.0%)	95/100 (95.0%)	96/100 (96.0%)	97/100 (97.0%)
Accuracy	119/135 (88.1%)	127/135 (94.0%)	126/135 (93.3%)	128/135 (94.8%)
NPV	100/107 (93.4%)	100/102 (98.0%)	100/105 (95.2%)	100/103 (97.0%)
PPV	35/37 (94.5%)	35/36 (97.2%)	35/36 (96.9%)	35/36 (97.2%)

Table 6 Ultrasound diagnostic performance of the CAD Synergy-Net

SENS	SPE	ACC	NPV	PPV
28/35 (80.0%)	72/100 (72.0%)	100/135 (74.0%)	100/118 (84.7%)	35/42 (83.3%)

In compliance with the architectural document, a python script was also created that performs the segmentation of neoplastic lesions in elastosonography. This script represents the artificial intelligence module used by the Synergy-Net application for the analysis of images relating to breast lesions. When a more performing classification model is available, it will be sufficient to replace just one file, making the update operation totally transparent to the CAD application.

4 Discussion

Both quantitative elastosonographies with ShearWave and integrated diagnostics with machine learning systems based on CNNs represent some of the most innovative technologies in the diagnosis of breast lesions, as demonstrated by copious literature on the subject, thanks to the real-time mapping of the stiffness of the breasts. Breast lesions, an element that provides further diagnostic information useful for improving patient management. Elastographic imaging constitutes a complementary tool for the management of breast cancer patients for:

- Diagnosis and characterization of breast lesions [12].
- Biopsy planning and treatment [13].
- Therapy planning and monitoring [14].

ShearWave™ PLUS elastography can contribute to the diagnostic phase of breast lesions and therefore have a positive impact on patient management¹. It can also help locate lesions during ultrasound-guided biopsy [15].

It is clear that we are talking about high-end and latest generation ultrasound machines, certainly not ideal for a screening protocol, but the objective of our study seeks to acquire the greatest amount of ultrasound information, in order to imagine, in the future, an online, open source tool capable of interfacing with the largest number of ultrasound images and supporting diagnosis with a CAD based on convolutional neural networks. There are numerous and constantly evolving screening programs proposed with a view to an early diagnosis of breast cancer, however, over the years and with the development of new diagnostic techniques, the awareness that mammography, alone, is an insufficient tool. The integration of mammographic diagnostics with heavier or contrast imaging techniques or with ultrasound examination is now increasingly widespread, in particular by enhancing the interpretation of imaging with technologies based on Machine Learning systems [16–20].

However, not all diagnostic units have kept pace, even less with the most advanced diagnostic techniques such as magnetic resonance imaging and mammography with contrast medium.

It is clear that equipping every diagnostic unit in the area with such equipment is absolutely impractical both due to the costs of equipment and specialized human material.

The industry is developing complex screening mammogram analysis systems that can help in two ways:

- By analyzing mammograms and directing to a second level only those that the system deems worthy of human control and subsequent investigations.
- Electronically analyze mammographic images deemed uncertain or suspicious.

While waiting to have expert systems capable of detecting uncertain or suspicious mammographic images, the need to minimize mammographic errors which often entail unnecessary additional costs, reducing the possibility of error to a minimum, appears clear. From this perspective, the latest generation ultrasound investigations, integrated with increasingly high-performance machine learning systems, could represent a valid alternative in the future, especially in younger patients.

5 Conclusion

This work described the Synergy-Net, an ongoing project aimed at creating a technological platform to support early oncological diagnosis based on the integration of interoperable communication and also a clinical data management system that exploits AI. Due to its deeply interdisciplinary nature, the Synergy-Net system was designed as a modular CAD where each module cooperates with the others, under the guidance of an orchestrator, to provide the required computation. This interdisciplinary nature has allowed us to work in parallel on different organs, exploiting common architectures, solutions and ideas and designing specific solutions for each patient.

The downside is that the project’s progress is not aligned, mainly due to different needs (e.g., conditions that must be met to engage a patient). This has been further complicated by the ongoing COVID-19 emergency. The result is that, while the design of the artificial intelligence algorithm for some organs has been completed (or almost completed), for other organs we are still in the data collection phase. Despite this, while working to complete ongoing activities, future steps are already being planned. There are two main ideas that the project will work on in the near future.

- The first idea is to exploit modern ultrasound elastosonography as a low-cost methodology with no side effects for the early diagnosis of oncological diseases for all tumor types considered in the Synergy-Net project and in which it is applicable.
- The second idea is to provide integrated prevention through data via fusion techniques. The objective is to simultaneously analyze information from multiple sources, in order to provide an integrated tool built on the medical knowledge of each specialist.

This will be further improved by the use of DNA sequencing tests aimed at finding correlations between mutations and imaging abnormalities that lead to cancer.

Acknowledgements Prof Procaccini Eugenio, Director of the Breast Unit Vanvitelli until 202. Campania Regional Council to support the Project development.

Declaration

- Ethics approval and consent to participate

This study followed the ethical principles of the Declaration of Helsinki and every participant was signing and approving the consent by the Institutional Review Board at “Policlinico” Hospital of the University of Campania “Luigi Vanvitelli” in Naples, Italy. Participation in the study was voluntary. Before inclusion in the study, study staff explained the purpose of the study and informed consent form was secured from each participant.

- Consent for publication

We have had authorization for publication like our Hospital Privacy Policy already require.

- Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

- Competing interests

The authors declare that they have no competing interests.

Founding This article did not receive sponsorship for publication.

References

1. Łukasiewicz S, Czezelewski M, Forma A et al (2021) Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review. *Cancers (Basel)* 13(17):4287
2. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71:209–249
3. World Health Organization (2018) Global health estimates 2016: disease burden by cause, age, sex, by country and by region, 2000–2016. Geneva
4. Kim EY, Chang Y, Ahn J et al (2020) Mammographic breast density, its changes, and breast cancer risk in premenopausal and postmenopausal women. *Cancer* 126(21):4687–4696
5. Duffy SW, Morrish OW, Allgood PC et al (2017) Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. *Eur J Cancer* 88:48–56
6. Schacht DV, Yamaguchi K, Lai J et al (2014) Importance of a personal history of breast cancer as a risk factor for the development of subsequent breast cancer: results from screening breast MRI. *Am J Roentgenol* 202:289–292
7. Parmeggiani D, Avenia N, Sanguinetti A et al (2012) Artificial intelligence against breast cancer (ANNES-BC-Project). *Ann Ital Chir* 83(1):1–5
8. Piantadosi G, Bovenzi G, Argenziano G et al (2019) Skin lesions classification: a radiomics approach with Deep CNN. *Lect Notes Comput Sci* 11808:252–259
9. Bollino R, Bovenzi G, Cipolletta F et al (2022) Synergy-Net: artificial intelligence at the service of oncological prevention. *Intell Syst Ref Libr* 211:389–424

10. Gravina M, Marrone S, Docimo L et al (2022) Leveraging CycleGAN in Lung CT Sinogram-free Kernel conversion. *Lect Notes Comput Sci* 13231:100–110
11. Thomas R, Das SK, Balasubramanian G et al (2022) Correlation of mammography, ultrasound and sonoelastographic findings with histopathological diagnosis in breast lesions. *Cureus* 14(12):32318
12. Yu Y, Xiao Y, Cheng J et al (2018) Breast lesion classification based on supersonic shear-wave elastography and automated lesion segmentation from B-mode ultrasound images. *Comput Biol Med* 1(93):31–46
13. Kim EK, Ko KH, Oh KK et al (2008) Clinical application of the BI-RADS final assessment to breast sonography in conjunction with mammography. *AJR Am J Roentgenol* 190(5):1209–1215
14. Youk JH, Gweon HM, Son EJ (2017) Shear-wave elastography in breast ultrasonography: the state of the art. *Ultrasonography* 36(4):300–309
15. Song EJ, Sohn YM, Seo M (2018) Diagnostic performances of shear-wave elastography and B-mode ultrasound to differentiate benign and malignant third breast lesions: the emphasis on the cutoff value of qualitative and quantitative parameters. *Clin Imaging* 50:302–307
16. Bejnordi BE, Veta M, van Diest PJ et al (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318:2199–2210
17. Araújo T, Aresta G, Castro E et al (2017) Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE* 12:0177544
18. Liu J, Xu B, Zheng C et al (2018) An end-to-end deep learning histochemical scoring system for breast cancer tissue microarray. *IEEE Trans Med Imaging* 38(2):617–628
19. Kooi T, Litjens G, Van Ginneken B et al (2017) Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* 35:303–312
20. Fenton JJ, Taplin SH, Carney PA et al (2007) Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 356(14):1399–1409

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Artificial Intelligence in Prostate Cancer Diagnosis: “Synergy-Net” in Campania FESR-POR (European Fund of Regional Development—Regional Operative Program) Research Project



Domenico Parmeggiani, Marco De Sio, Giancarlo Moccia, Pasquale Luongo, Francesco Miele, Alfredo Allaria, Francesco Torelli, Stefano Marrone, Michela Gravina, Carlo Sansone, Ruggiero Bollino, Paola Bassi, Antonella Sciarra, Davide Arcaniolo, Maddalena Claudia Donnarumma, Chiara Colonnese, Lorenzo Romano, Federica Colapietra, Marina Di Domenico, Ludovico Docimo, and Massimo Agresti

Abstract Background: The diagnosis of prostate cancer can only be obtained following the analysis of the tissue taken by means of a biopsy. Given the position of the organ, the biopsy is typically assisted by ultrasound images and the procedure consists of taking different portions of tissue from different areas, according to a map well-defined by international standards. Given the invasiveness of the procedure, the

D. Parmeggiani (✉) · G. Moccia · P. Luongo · F. Miele · A. Allaria · F. Torelli · P. Bassi · A. Sciarra · M. C. Donnarumma · C. Colonnese · M. Agresti
Department of Advanced Medical and Surgical Sciences, University of Campania “Luigi Vanvitelli”, Naples, Italy
e-mail: domenico.parmeggiani@unicampania.it

M. De Sio · D. Arcaniolo · L. Docimo
Department of Woman, Child and General and Specialized Surgery, University of Campania “Luigi Vanvitelli”, Naples, Italy

S. Marrone · M. Gravina · C. Sansone
Department of Electrical Engineering and Information Technology (DIETI), University “Federico II”, Naples, Italy

C. Sansone
CINI, ITEM Laboratory “C.Savy”, Via Cintia 21, Naples, Italy

R. Bollino
Bollino.It SpA, Via Delle Ustrie 31, Naples, Italy

L. Romano
Department of Neurosciences, Reproductive Sciences and Odontostomatology “Federico II”, University of Naples, Naples, Italy

F. Colapietra · M. Di Domenico
Department of Precision Medicine, University of Campania “Luigi Vanvitelli”, Naples, Italy

© The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011, https://doi.org/10.1007/978-981-97-4581-4_23

313

objective set within the Synergy-Net project is to analyze biomedical images in order to guide the operator on identifying the most suspicious tissues. *Results:* The dataset acquired by the Synergy-Net Platform at the “Vanvitelli” Urology Operating Unit is made up of a total of 350 outpatient services from which the diagnosis emerged on ultrasound, elastosonography, RNM, and biopsy of 50 prostate carcinomas which were then operated on. In the context of the Synergy-Net project, a new convolutional architecture was therefore created based on the U-Net paradigm, designed to perform a slice-by-slice segmentation in DCE-MRI of the prostate. The data processing with CNNs was carried out on a dataset of 37 patients, selected from the initial 50 for completeness and uniformity of the data, all affected by k-prostatic disease, using a tenfold cross-validation in order to obtain a statistically more significant estimate of the goodness of the results obtained. The performance metric used was the DICE coefficient. *Conclusion:* The results present a low intra-subject variability and a high inter-subject variability, with DICE values ranging between a minimum of 5.8% and a maximum of 60.3%. On average, a value of 35% is reported, considering the arithmetic mean of the dice achieved on all folds (macro-average).

Keywords Prostate cancer · Elastography · RNM · Artificial intelligence (AI) · Convolutional neural networks (CNNs)

1 Introduction

This study is part of a ReS project (Sinergy-Net: Research and Digital Solutions in the fight against oncological diseases). Its objective is the development of a Synergy-Net technological platform aimed at improving and strengthening oncological cancer prevention. This study refers exclusively to prostate cancer.

Prostate cancer represents the most frequent neoplasm in terms of incidence in men. In Italy, according to the latest AIOM/AIRTUM report, more than 36,000 new cases of prostate cancer were recorded in 2020, and in Campania, the incidence reaches 3000 cases/year. However, 5-year survival is over 90% and this result has been achieved thanks to screening and early diagnosis, as well as the development of increasingly effective medical and surgical therapies. In the initial phase, prostate cancer is generally asymptomatic, but as the loco-regional disease progresses, urinary symptoms may appear. In the more advanced stages of the disease, the skeleton being the first site of metastasis, the development of bone pain or pathological fractures is characteristic. In patients with localized prostate cancer, there is the possibility of subjecting the patient to radical prostatectomy surgery, which can now be performed with minimally invasive methods such as laparoscopy and robotics, or to radiotherapy, which has also become less invasive and better tolerable for the patient. However, these therapies are not free from complications and sequelae which can significantly affect the quality of life of patients. Therefore, in low-risk cases, when the neoplasm does not present invasive characteristics or a high capacity for progression, the patient can be subjected to active surveillance, thus avoiding problems

secondary to active treatments. And it is precisely to recognize "indolent" neoplasms, not clinically significant from potentially aggressive ones which currently represents the most important diagnostic challenge for researchers in this field. In fact, in prostatic neoplasia, the risk of over diagnosis, i.e., the histological diagnosis of a tumor which, in the absence of screening, would not come to clinical attention, is extremely high. The costs, both in economic terms and in terms of patient morbidity, of diagnostics and therapy are extremely high, so reducing the rate of diagnosis of clinically insignificant tumors represents a fundamental objective to be pursued. Over diagnosis depends on two factors, namely the natural history of the neoplasm, which provides that some lesions have little or no probability of progressing. Currently, the diagnostic process for prostate cancer involves PSA blood testing, rectal examination, and the performance of a multiparametric magnetic resonance imaging of the prostate. The combination of these factors determines the indication to perform a prostate biopsy, which is essential for a definitive histological diagnosis. The role of ultrasound, including more advanced techniques such as elastosonography, CEUS, and micro-ultrasonography, and that of other markers still remains uncertain and to be defined [1–8].

Multiparametric magnetic resonance imaging is indicated by the guidelines as a mandatory investigation to be carried out before performing a prostate biopsy and a re-biopsy. This is a fundamental test for the indication of biopsy; however, it has a high cost and requires particular expertise for interpretation. One of the problems, for example, is the characterization of lesions with a PI-RADS score of 3, which often do not result in significant disease despite being an indication for performing a prostate biopsy. And the biopsy itself is far from free from complications, as it can lead to very severe hemorrhages and sepsis, with a high risk of hospitalization, as well as having a risk of false negatives that can reach 30% of the evolution of the prostate biopsy [9].

The aim, therefore, is twofold: on the one hand, we want to minimize the number of biopsy samples, on the other, we want the samples to maximize the probability of acquiring tissues containing the neoplastic cells sought (if present). The integrated procedure takes into account several types of data, including dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), ultrasound images, and information about the patient's age and family history of the disease. The objective of the project is to intervene on the diagnostic algorithm by integrating clinical data and instrumental data with the aim of reducing the risk of over diagnosis and optimizing resources. The Synergy-Net system is configured in the context of computer-aided diagnosis (CAD), i.e., computerized systems to support medical diagnosis. The data analyzed are purely images/videos and tabular data. The strength of the project is to have as much information as possible on the biological, clinical, and diagnostic nature of the tumor pathology. Therefore, ensure that there is no longer a solely diagnostic evaluation of patient data, but integrate this data with a series of information that allows the training of artificial intelligence systems [10–14].

2 Materials and Methods

The Synergy-Net system is configured in the context of computer-aided diagnosis (CAD), i.e., computerized systems to support medical diagnosis. In general, CAD systems can work on images, on tabular data (e.g., clinical, anamnestic, etc.) or on a combination of them. In the case of the Synergy-Net project, the data analyzed are mainly images/videos and tabular data. Among them, biomedical images are those that require a preliminary phase for their representation, in terms of assimilable characteristics by an expert system. As regards prostate cancer, the dataset acquired by the Platform from 01/01/2021 to 12/31/2021 at the Urology Operating Unit of the High Specialty Medical-Surgical Department of the University of Campania Hospital “Luigi Vanvitelli” is made up of a total of 350 outpatient services from which the diagnosis emerged on ultrasound, elastosonography, RNM, and biopsy of 50 prostate carcinomas which were then operated on.

The integrated procedure then took into account different types of data, including dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI); ultrasound images and information relating to the patient’s age and family history with the disease (maternal family history with breast cancer and paternal family history with prostate cancer). Following the study, in compliance with the architectural document, a python script will be created that carries out the detection and classification of the areas of interest in endoscopic images, using an artificial intelligence model trained for the purpose, to be inserted in a specific section of the Synergy-Net application.

- Identify factors predictive of clinically significant disease (single or in combination).
- Identify factors predictive of “indolent” disease, effectively reducing the number of prostate biopsies (reducing costs and morbidity).
- Selection of patients with PI-RADS 3 who may not undergo biopsy.
- Selection of patients who, based on risk factors, can be subjected to a random biopsy without the fusion technique (reducing costs and making it possible to perform the biopsy even in centers that do not have the fusion technique).

The data collected for each patient concerns:

Age, BMI, place of residence, profession, exposure to toxic substances/radiation, familiarity with prostate/breast cancer, alcohol consumption, smoking, consumption of dairy products, consumption of red meat, sedentary lifestyle, frequency of ejaculations (monthly), IIEF-5, IPSS, PSA (ng/ml), PSA velocity and PSA density, other markers (IXIP, PHI, etc.), rectal examination, prostate volume, presence of nodules on transrectal ultrasound, possible presence of positive lesions on elastosonography (value in kPascal), PI-RADS score on MRI-MP.

The analysis of the multiparametric MRI to be integrated with the data previously provided is performed by reporting, through a specific program for viewing and editing images in DICOM format, the tumor areas highlighted in the definitive histological examination.

Given the extremely innovative nature of the study, the design of the necessary artificial intelligence systems and the analysis of the results are constantly evolving. Furthermore, the limited availability of data and a fully evolving literature on the subject, if on the one hand confirm the innovativeness of the research, on the other make development in intermediate steps necessary [15–18].

Given the small size of the organ under examination, the most complex operation remains that of semantic segmentation, i.e., the identification of the precise contours of the suspicious lesions within the diagnostic images. In fact, having this information allows you to support the doctor in the analysis of suspicious areas, without distractions related to nearby organs and tissues (a situation that is particularly true for the prostate).

In the context of the Synergy-Net project, a new convolutional architecture based on the U-Net paradigm was therefore created, designed to perform slice-by-slice segmentation in DCE-MRI of the prostate. Furthermore, in order to allow more effective training of the model, even in the presence of limited amounts of data, the concept of transfer-learning was used in order to pre-train portions of the network, thus allowing faster and more effective training (see Fig. 1).

Data processing with CNNs was conducted on a dataset of 37 patients, selected from the initial 50 for completeness and uniformity of the data, all affected by k-prostatic disease, using a tenfold cross-validation in order to obtain a statistically more accurate estimate. Significant of the goodness of the results obtained. The performance metric used is the DICE coefficient, widely used by the scientific community in the context of semantic segmentation due to its high sensitivity even to small variations in the predicted segmentation masks.

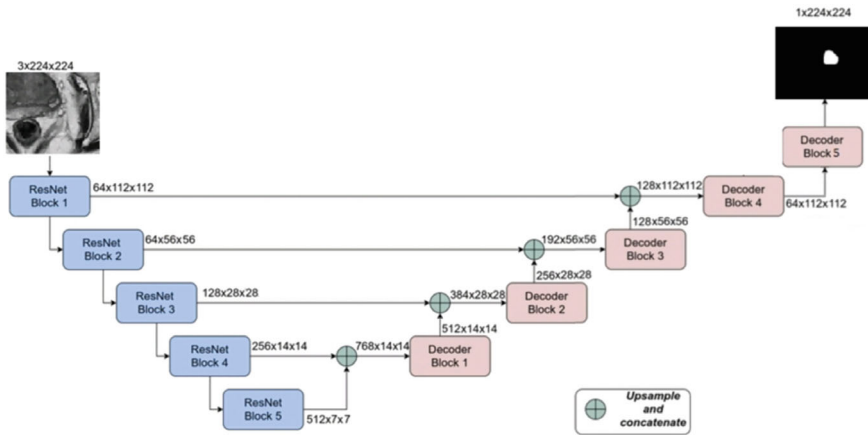


Fig. 1 U-Net convolutional architecture designed for K-prostate

3 Results

Despite the obvious limitations associated with the complexity of the task, the created system can be effectively used to support the identification of damaged areas in the prostatic tissue, in DCE-MRI all affected by k-prostatic disease, using a tenfold cross-validation in order to obtain a statistically more significant estimate of the goodness of the results obtained. The performance metric used was the DICE coefficient. The results present a low intra-subject variability and a high inter-subject variability, with DICE values ranging between a minimum of 5.8% and a maximum of 60.3%. On average, a value of 35% is reported, considering the arithmetic mean of the dice achieved on all folds (macro-average). This result is not unexpected, given the very small size of the organ (and therefore its possible lesions) and the sensitivity of the diagnostic instrument (approximately 1.5 mm³). An analysis of the predictive capabilities based on the percentage of damaged tissue compared to healthy tissue (indicated as $p\%$) tends to confirm the suspicion. It follows that, despite the obvious limitations associated with the complexity of the task, the created system can be effectively used to support the identification of damaged areas in the prostatic tissue (see Fig. 2).

In compliance with the architectural document, a python script was also created that performs the segmentation of neoplastic lesions, slice-by-slice, in DCE-MRI. This script represents the artificial intelligence module used by the Synergy-Net application for the analysis of images relating to prostatic lesions. When a more performing classification model is available, it will be sufficient to replace just one file, making the update operation totally transparent to the CAD application.

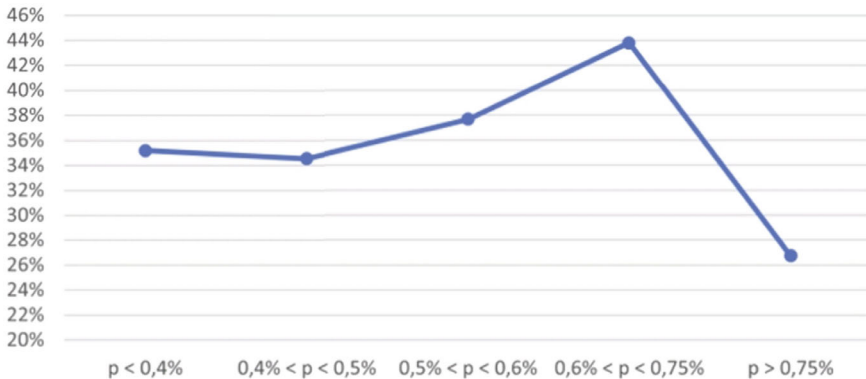


Fig. 2 Average DICE based on the portion of damaged tissue

4 Discussion and Conclusions

The work carried out consisted of collecting, first of all, a significant number of patient data and trying to use this data to "train" the system. Before starting the screening, the system was prepared, trained, to recognize certain oncological images. This might, of course, be intuitive for an expert clinician, but not for a still "inexperienced" machine. In the future, this system could make it possible to identify lesions that would not have been diagnosed with other investigations, improve the diagnostic accuracy of certain methods, select which patients to perform a certain test in, and provide support to less experienced doctors.

All this could allow for an increasingly precise, earlier diagnosis. In summary, the Synergy-Net Information System (IS) aims to be an AI-based CAD designed to support the doctor in a set of different pathologies of an oncological nature, in order to exploit the benefits that only an analysis integrated can fully enable.

As regards prostate cancer, the aim of the project was to evaluate the possibility, thanks to the most innovative machine learning methods applied both to prostatic elastosonography and to RNM with prostate contrast medium, to reduce the number of biopsy samples for diagnostic purposes and to be able to maximize the probability of acquiring anatomically–pathologically interesting tissues from the aforementioned samples. The results obtained with the creation of CAD-CNNs to be applied to RNM diagnostics are promising and certainly integrated with information on prostatic hardness, measured in kPa, obtained with quantitative elastosonography, which can make the CAD itself perform better.

Acknowledgements Campania Regional Council to support the project development.

Funding This article did not receive sponsorship for publication.

Declarations

Ethics Approval and Consent to Participate This study followed the ethical principles of the Declaration of Helsinki and every participant was signing and approving the consent by the Institutional Review Board at "Policlinico" Hospital of the University of Campania "Luigi Vanvitelli" in Naples, Italy. Participation in the study was voluntary. Before inclusion in the study, study staff explained the purpose of the study and informed consent form was secured from each participant.

Consent for Publication We have had authorization for publication like our Hospital Privacy Policy already requires

Availability of Data and Materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests The authors declare that they have no competing interests.

References

1. Coviello V, Buzzoni C, Fusco M et al (2017) Survival of cancer patients in Italy. *Epidemiol Prev* 41(2):1–244
2. Rassweiler J, Autorino R, Klein J et al (2017) Future of robotic surgery in urology. *BJU Int* 120:822–841
3. Davies BL, Hibber RD, Ng WS et al (1991) The development of a surgeon robot for prostatectomies. *Proc Inst Mech Eng* 205:35–38
4. Senevirathna P, Pires DEV, Capurro D (2023) Data-driven overdiagnosis definitions: a scoping review. *J Biomed Inform* 147:104506
5. Hogenhout R, Remmers S, van Slooten-Midderigh ME et al (2023) From screening to mortality reduction: an overview of empirical data on the patient journey in European randomized study of screening for prostate cancer Rotterdam after 21 years of follow-up and a reflection on quality of life. *Eur Urol Oncol* S2588-9311(23), 00172–4
6. Roehl KA, Antenor JA, Catalona WJ (2002) Serial biopsy results in prostate cancer screening study. *J Urol* 167(6):2435–2439
7. Shoji S (2019) Magnetic resonance imaging-transrectal ultrasound fusion image-guided prostate biopsy: current status of the cancer detection and the prospects of tailor-made medicine of the prostate cancer. *Investig Clin Urol* 60(1):4–13
8. Zhang J, Zhu A, Sun D et al (2020) Is targeted magnetic resonance imaging/transrectal ultrasound fusion prostate biopsy enough for the detection of prostate cancer in patients with PI-RADS ≥ 3 : results of a prospective, randomized clinical trial. *J Cancer Res Ther* 16(7):1698–1702
9. Rodríguez-Cabello MA, Mendez-Rubio S, Sanz-Miguelañez JL et al (2023) Prevalence and grade of malignancy differences with respect to the area of involvement in multiparametric resonance imaging of the prostate in the diagnosis of prostate cancer using the PI-RADS version 2 classification. *World J Urol* 41:2155–2163
10. Kroenig M, Schaal K, Benndorf M et al (2016) Diagnostic accuracy of robot-guided, software based transperineal MRI/TRUS fusion biopsy of the prostate in a high risk population of previously biopsy negative men. *Biomed Res Int* 2016:2384894
11. Yilmaz EC, Lin Y, Belue MJ et al (2023) PI-RADS version 2.0 versus version 2.1: comparison of prostate cancer Gleason grade upgrade and downgrade rates from MRI-targeted biopsy to radical prostatectomy. *AJR Am J Roentgenol* 37729551 (2023)
12. Bollino R, Bovenzi G, Cipolletta F et al (2022) Synergy-Net: artificial intelligence at the service of oncological prevention. *Intell Syst Ref Libr* 211:389–424
13. Gravina M, Marrone S, Docimo L et al (2022) Leveraging CycleGAN in Lung CT Sinogram-free kernel conversion. *LNCS* 13231:100–110
14. Piantadosi G, Bovenzi G, Argenziano G (2019) Skin lesions classification: a radiomics approach with deep CNN. *LNCS* 11808:252–259
15. Wang K, Xing Z, Kong Z et al (2023) Artificial intelligence as diagnostic aiding tool in cases of prostate imaging reporting and data system category 3: the results of retrospective multi-center cohort study. *Abdom Radiol (NY)* 37740046
16. Fron A, Semianuk A, Lazuk U et al (2023) Artificial intelligence in urooncology: what we have and what we expect. *Cancers (Basel)* 15(17):4282
17. Hong S, Kim SH, Yoo B et al (2023) Deep learning algorithm for tumor segmentation and discrimination of clinically significant cancer in patients with prostate cancer. *Curr Oncol* 30(8):7275–7285
18. Thomas M, Murali S, Simpson BSS et al (2023) Use of artificial intelligence in the detection of primary prostate cancer in multiparametric MRI with its clinical outcomes: a protocol for a systematic review and meta-analysis. *BMJ Open* 13(8):e074009

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Designing Reactive Route Change Rules with Human Factors in Mind: A UATM System Perspective



Jeongseok Kim and Kangjin Kim

Abstract This paper investigates the dynamic rerouting of electric vertical takeoff and landing (eVTOL) aircraft in the context of urban air traffic management (UATM). Focusing on the interaction between human managers and the UATM network, we present a novel approach to reactive rerouting based on step-oriented simulation and condition-action rules. Our framework enables human intervention in response to congestion observed at vertiport corridors, allowing managers to request detours for approaching eVTOLs. We formulate the problem within the knowledge representation and reasoning (KR&R) paradigm and employ a multi-shot approach within an Answer Set Programming (ASP) solver to drive the step-oriented simulation. The structure of the paper follows a logical progression, including related work, preliminaries, a problem statement, proposed solutions, and a discussion, followed by a concluding section.

Keywords UAM · UATM · KRR · Answer Set programming · Articulating agent

1 Introduction

Urban air mobility (UAM) is a fast-growing aviation industry paradigm that requires low-altitude aircraft to be integrated into dense urban environments, high-density air traffic management, and autonomous operations [20]. Hence, an air traffic management system that can adapt to a varied environment and extend to meet rising data volume and complexity is needed. This study uses UAM air traffic management (UATM) methods to solve these challenges.

J. Kim
SK Telecom, Seoul 04539, Republic of Korea
e-mail: jeongseok.kim@sk.com

K. Kim (✉)
Department of Drone Systems, Chodang University, Jeollanam-do 58530, Republic of Korea
e-mail: kangjinkim@cdu.ac.kr

© The Author(s) 2024
X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
https://doi.org/10.1007/978-981-97-4581-4_24

The paper offers a graph model of the UAM corridors, with each node representing a vertical airport (vertiport) and a human traffic manager for landing and takeoff. The autonomous operating span of every UATM system is determined by its proximity to a vertiport, requiring UATM network instructions for particular agents [16].

A more robust, adaptive, and scalable future urban air traffic management system that can accommodate a wide range of stakeholders and satisfy their data transmission requirements would improve UAM development and dependability.

1.1 Contributions

The contributions of this paper are as follows:

1. Theoretical development for reactive rules,
2. Step-oriented simulation, and
3. Validity evaluation mechanisms for explainability.

1.2 Outlines

The rest of the paper is organized as follows: We review related work. We study background knowledge. We then examine the problem. We propose a solution to the problem. We study how to utilize reactive rules for human factors. Next, we address possible discussions. Finally, we conclude our work.

2 Related Works

The papers [9, 20, 23, 27] cover UAM technology, regulatory framework, benefits, and drawbacks. The article Administration [1] offers a tiered system to arrange UAM airspace by vehicle type, autonomy, and altitude. German Aerospace Center (DLR) programs discuss ATM initiatives in Schuchardt et al. [29]. The authors in [15] suggest a vehicle-obstacle crash assessment model. In article [26], the authors present cutting-edge ATM deep learning algorithms.

Solution in [17] for collision avoidances has been limited due to the complexity of UATM systems, stakeholder diversity, and unexpected aerial accidents. The papers [19, 28, 30] uses non-monotonic reasoning to characterize route detours involving multiple UATMs, rejecting previous conclusions based on new information. The paper [31] presents an example scenario for changing the destination while the paper [18] shows how a round route can be added for clearing the corridor.

The research paper [4] is great at explaining complex systems, often known as explainable AI. To validate the complicated system, their epistemological approach

uses knowledge representation and reasoning. The article [5] develops automatic classifiers that may provide ontology-based explanations. This study, [25], covers five methods for building descriptive logic (DL) ontologies: association rule mining, formal concept analysis, inductive logic programming, computational learning theory, and neural networks. Since their works are to provide a theoretical foundation, practically adapting their concepts has been limited.

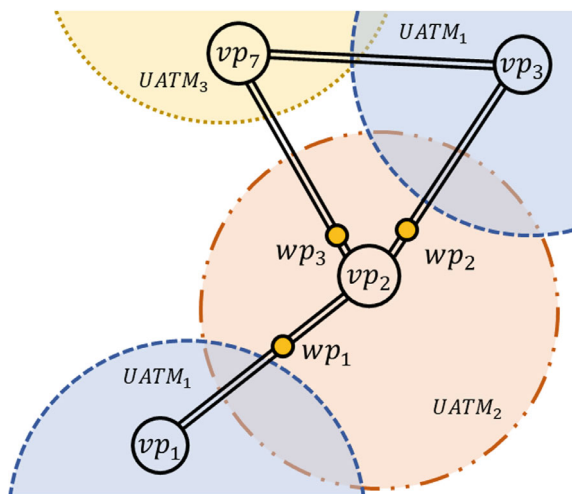
In contrast, the paper [12] proposes an ASP-based intra-logistics paradigm. Papers [13, 24] advocate using ASP for AGV work assignments and routing. They demonstrate how to compute their route and supply subtasks, while we focus on the logic utilized in human manager-system and system-system interaction. The uncertainty representation and reasoning evaluation framework analyzes space object tracking uncertainty in [3]. This provides a common uncertainty vocabulary and reasoning ontology. This shows how URREF can model tracking uncertainty and assess tracking findings' validity, precision, and recall.

3 Preliminaries

3.1 An Example Scenario

Figure 1 depicts a UATM network connecting four cities as a graph of nodes and edges. There are four nodes, which represent the VTOL airports (called vertical airports, or vertiports in short) in each city. The edges connecting two nodes represent bidirectional corridors that connect the vertiports in each city. The UATM network is organized so that three UATM systems can be nested on top of each other for communication and control.

Fig. 1 UATM Network comprises four vertiports— vp_1 , vp_2 , vp_3 , and vp_7 —connected by bidirectional corridors. There are three UATMs ($UATM_1$, $UATM_2$, and $UATM_3$), each represented by an outer circle indicating their coverage area



There are three main parties in the UATM network described above. The first is the airplane traveling to and from the city, which is referred to as the agent in this paper. The second is the three UATM systems, which are responsible for route determination for the agents. The third party is the air traffic manager (called the human manager), who monitors and controls the takeoffs and landings of airplanes at each vertiport.

3.2 Assumptions

Agents migrate between vertiports through these corridors, and UATMs can communicate directly with agents in their region. The “UATM Network” serves as a communication relaying mechanism for agents.

3.3 Roles for Each Party

The UATM system can operate autonomously. It decides and computes all plans for the agents without confirmation from the human manager. The human manager monitors each vertiport and reports its status to the UAM system. He or she also interacts with legacy traffic systems. The agents execute their plan, which is determined by the UATM system. They ask permission from the human manager of the vertiports where they take off or land.

3.4 Environment Specification

Each corridor restricts agents to entering one at a time. It also prevents agents from passing each other in a corridor. Hence, delays can occur when agents in front of them slow down. Given the speed at which agents travel and the length of each corridor, each corridor has the capacity to hold agents for some period of time.

Since this is a simplified route change problem, an agent enters one vertiport at a time and leaves another vertiport later. Thus, we assume that once it reaches its target vertiport, it is removed from the vertiport.

3.5 Pathfinding and Route Changing

Before considering the route change, we have to consider that there is a given path for each agent from the beginning. Hence, the concept of pathfinding should be addressed first. For any agent, its path is a sequence of vertiports from its start vertiport to its

target vertiport. For example, in Fig. 1, consider that agent 1 starts its path at vp_1 and finishes at vp_3 . In this case, its path can be vp_1 , vp_2 , and vp_3 . Note that in a path, a vertiport and its following vertiport are directly connected to each other. Thus, pathfinding is a procedure to compute the sequence of vertiports when its start and target vertiports are given.

A route change is another pathfinding procedure, considering an agent's initial path instead of its start and destination vertiports. Given an initial path, the route-changing procedure returns its alternative path.

3.6 Possible Conflicts and Ways to Anticipate Them

Before addressing possible conflicts, the concept of waypoints should be introduced. To be more specific, we discretize each corridor with sampled points (or waypoints, in short). We then consider each agent as moving from one waypoint to the next on a corridor.

For example, there is a path consisting of vp_1 , vp_2 , and vp_3 . Let us denote the corridor from vp_1 to vp_2 $\mathcal{C}_{1,2}$ and the corridor from vp_2 to vp_3 $\mathcal{C}_{2,3}$. Then, assume $\mathcal{C}_{1,2}$ has 20 waypoints in it and $\mathcal{C}_{2,3}$ has 30 waypoints in it. Then, the path can be represented as the following detailed path: $vp_1 = (\mathcal{C}_{1,2}).wp_1, wp_2, \dots, wp_{20} = vp_2 = (\mathcal{C}_{2,3}).wp_1, wp_2, \dots, wp_{30} = vp_3$.

Possible Conflicts Consider there are two agents— r_i, r_j —heading to the same target vertiport, say vp_3 . Assume that they are on a same corridor, say $\mathcal{C}_{2,3}$, and r_j is ahead of r_i . Given that the time is synchronized and each agent moves one step at a time, the location of r_j at time t , $\mathcal{L}_t^{r_j}$, can be one step ahead of the r_i 's one. Then, $\mathcal{L}_{t-1}^{r_j} = \mathcal{L}_t^{r_i}$ and $\mathcal{L}_t^{r_j} = \mathcal{L}_{t+1}^{r_i}$.

If r_j unexpectedly stops moving at $t + 1$ and r_i couldn't sense that, the location for these agents at $t + 1$ would be $\mathcal{L}_{t+1}^{r_i} = \mathcal{L}_{t+1}^{r_j}$. This will cause a collision.

Consider r_i is heading to vp_3 through $\mathcal{C}_{7,3}$ and r_j is heading to vp_3 through $\mathcal{C}_{2,3}$. If these two agents visit vp_3 at the time t simultaneously, then $\mathcal{L}_t^{r_i} = \mathcal{L}_t^{r_j}$. This will also cause a collision.

We note that each corridor has a dedicated lane for each direction so that a collision wouldn't happen by switching over their locations.

How to anticipate the conflicts The capacity of each corridor is calculated taking into account delay and possible collisions. Thus, the evenly distributed agents will mitigate collisions due to densely coupled traffic in a corridor. In addition, keeping the number of agents in a corridor below its capacity will prevent collisions due to uncontrolled entry of agents, which provides an anticipation of further conflicts.

3.7 Possible Resolutions

Rerouting before entering the congested corridor If agents can detect the traffic sufficiently earlier than approaching the center of the congested space, they can

Table 1 Predicates for the environment

Predicates	Description	Examples
uatm/1	It represents the Urban Air Traffic Management	uatm(1), uatm(2), uatm(3)
vp/1	It represents vertiports	vp(1), vp(2)
edge/2	It represents corridors	edge(vp(1), vp(2))
cover/2	It represents the covered uatm of a vertiport	cover(uatm(1), vp(3))
edge_range/3	It represents waypoints within a corridor	edge_range(vp(1), vp(2), 1..20)
covered_wp/4	It represents covered waypoints in a corridor by a UATM	covered_wp(vp(1), vp(2), uatm(1), P) :- edge_range(vp(1), vp(2), P), P < 16.
capacity/3	It represents the safe limit each corridor can accommodate agents	capacity(5, vp(1), vp(2))

avoid the traffic by altering their route. However, in order to achieve this, frequent traffic monitoring and communication with the UATM are necessary.

For example, suppose some agents in $C_{1,2}$ need to pass through $C_{2,3}$, but it is reported that $C_{2,3}$ is congested. In this case, they can reroute to $C_{2,7}$ and $C_{7,3}$ instead. *Changing the target due to the significant delays* If there are significant delays to a destination, agents heading to that destination may need more time to reach the destination due to the expected delays. In such a case, their destination can be changed to another one, taking into account the fuel issue and the total travel time required. *Rerouting after entering the congested corridor: Clearing the corridor to prevent possible collisions* Consider that some agents enter a corridor where the unexpected delay is dynamically increased. In simple terms, we can think of an agent entering a corridor where the capacity is already full. In such cases, to resolve potential collisions, agents ahead can clear the corridor by changing their destination or adding more routes.

In this paper, we note that we do not consider the case when the collision has already happened. We try to prevent future conflicts so that collisions can be avoided.

3.8 All Used Predicates, Queries, and Expected Results

Before formulating the problem we want to solve, we need to specify the predicates used. For the environment, the following predicates are used as shown in Table 1. For the agents and the plans, the following predicates are used as shown in Table 2. Then, Table 3 shows the queries and the expected results.

Table 2 Predicates for the agent

Predicates	Description	Examples
agent/1	represents eVTOL	agent(1)
loc/4	It represents the location (or the waypoint WP) where an agent is in a corridor at the time T	loc(agent(1), T, vp(1), vp(2), WP)
plan/4	It represents the plan where an agent has at the time T	plan(agent(1), T, vp(1), vp(2))
source/3	It represents the start vertiport of an agent at the time T	source(agent(1), T, vp(1))
target/3	It represents the destination vertiport of an agent at the time T	target(agent(1), T, vp(3))

Table 3 Queries and expected results

Queries	Expected Results
[Q1] How many agents are in the related corridor?	[A1] The agent count where in the corridor
[Q2] Is the corridor capable for more agents?	[A2] The number of capable agents in that corridor
[Q3] Is the corridor congested?	[A3] True or false, depending on the fact
[Q4] Change the route of vp_3 heading agents	[A4] Update the current plan with their backup plan

3.9 Time-Wise Computations

Most queries, including the ones we covered above, consider the current status of agents’ movements and plans, as well as where they are assigned. Such information changes depending on traffic and/or some other decision, such as a detour. Therefore, in some way, we need to track agents’ movements over time. In particular, it should be possible to track any piece of information that changes over time.

3.10 Considering Human Factors

When interacting with human managers in this setting, it is important to keep in mind that humans are not a single entity. This is because humans are prone to making mistakes, they require time to react, they take time to recognize situations, even when they are being monitored, and they frequently become confused when they are not provided with any additional explanation.

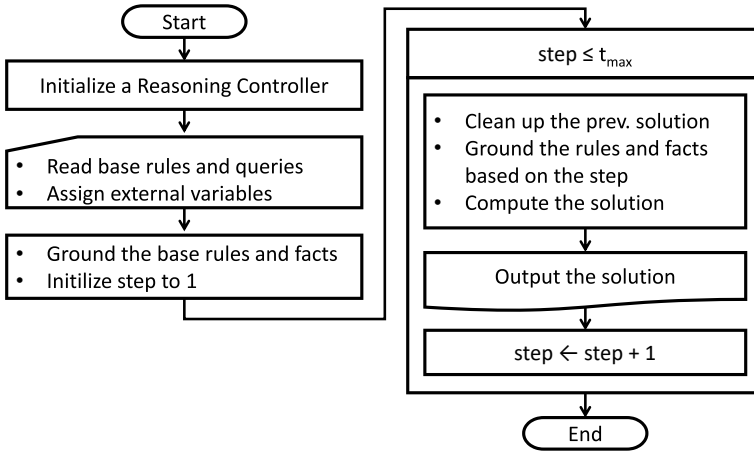


Fig. 2 Flowchart illustrating the step-by-step process for calculating the solution

4 Problem Formulation

Problem Statement: Given a group of agents that operate based on their scheduled tasks, a UATM network, and a human manager, provide an answer to him or her when a query is requested. In order for the human manager to comprehend the answer, provide a way that each procedure can be traceable.

Solution Overview: We have constructed a time-based simulator as shown in Fig. 2. According to the plans provided to the agents, we simulate their movements as the time step is increased. After a request is made, the necessary rules are computed for each time step. The result is then generated. All procedures are tracked throughout this process.

5 Proposed Solution

5.1 Suggested Results of the Queries

To analyze agent movement, we translated each predicate in Tables 1, 2 3 at different time steps into dedicated queries within Answer Set Programming (ASP) [2, 21, 22]. ASP defines relationships between objects and their properties, leading to “answer sets” representing solutions [14]. We used an ASP solver for each query, effectively retrieving the agents’ actions and the results of queries while incrementing the time step [11].

This “multi-shot” approach (detailed in [11]) recalculates earlier decisions at higher time steps but offers valuable insights. We can clearly observe individual decision-making processes and trace the overall movement flow (see Table 4).

To deepen understanding, we’ll first analyze individual query results. Then, we’ll examine the step-wise procedure, revealing the reasoning behind each decision.

[A1] *Predicate **occupied/4** is produced.* Given time step t and the associated corridor connecting vertiport u to vertiport v , the result of Q1 is the predicate `occupied(C, t, u, v)`, where C is the number of agents in that corridor at time t .

[A2] *Predicate **capable/4** is produced.* Once the predicate `occupied(C, t, u, v)` is given, the result of Q2 is the predicate `capable($C_{max} - C, t, u, v$)`, which can be computed by referring to the `capacity(C_{max}, u, v)` of the corridor, where C_{max} is the maximum capacity of the corridor that the agents can occupy.

[A3] *Predicate **congestion/3** is produced.* Once the predicate `capable(C_2, t, u, v)` is given, the result of Q3 is the predicate `congestion(t, u, v)`, which can be easily computed by checking if $C_2 \leq 0$. This means that the corresponding corridor cannot accommodate any more agents.

[A4] *Predicate **change_route/2** is produced.* Once the detour request has been made, the result of Q4 is the predicate `change_route(A, t)`, where A are the agents to be detoured at time t .

The result of each query is shown in Table 4. While the time step goes on, each agent’s location is updated. Through the `congestion(3, 2, 3)iii`, we can see that at step 4, the detour request can be triggered. Then, through the `change_route/2`, the human manager can be informed of the route change. It is worth noting that at step 5, agent 2 took the corridor from vp_2 to vp_7 (as indicated in wp_3 in Fig. 1), which demonstrates its route has been changed.

5.2 Step-Oriented Simulation

After initializing a reasoning controller, it reads base rules for environment, agents, and plans in Table 1~2. Then it reads queries in Table 3. Next, it grounds the rules and facts, and it prepares the while loop, setting *step* to 1. The while loop will continue grounding the rules and facts as long as the *step* remains less than or equal to t_{max} . The *step* will be incremented by 1 throughout each iteration. The solution will be computed and returned after the grounding process.

Table 4 Result of the queries at each step. The superscripts i to iv indicate that they are obtained from the results of query [1] to [4].

Step	Results of Query [1] to [4]
1	loc(1,1,1,2,15) loc(2,1,1,2,17) loc(3,1,1,2,19) loc(4,1,2,3,4) loc(5,1,2,3,7) occupied(0,0,2,3) ⁱ capable(3,0,2,3) ⁱⁱ
2	loc(5,2,2,3,8) loc(4,2,2,3,5) loc(3,2,1,2,20) loc(2,2,1,2,18) loc(1,2,1,2,16) occupied(2,1,2,3) ⁱ capable(1,1,2,3) ⁱⁱ
3	loc(1,3,1,2,17) loc(2,3,1,2,19) loc(4,3,2,3,6) loc(5,3,2,3,9) loc(3,3,2,3,1) occupied(2,2,2,3) ⁱ capable(1,2,2,3) ⁱⁱ
4	loc(3,4,2,3,2) loc(5,4,2,3,10) loc(4,4,2,3,7) loc(2,4,1,2,20) loc(1,4,1,2,18) occupied(3,3,2,3) ⁱ capable(0,3,2,3) ⁱⁱ congestion(3,2,3) ⁱⁱⁱ detour_request(3,4) ^{iv} detour_request(2,4) ^{iv} detour_request(1,4) ^{iv} change_route(3,4) ^{iv} change_route(2,4) ^{iv} change_route(1,4) ^{iv}
5	loc(1,5,1,2,19) loc(4,5,2,3,8) loc(5,5,2,3,11) loc(3,5,2,3,3) loc(2,5,2,7,1) occupied(3,4,2,3) ⁱ capable(0,4,2,3) ⁱⁱ congestion(4,2,3) ⁱⁱⁱ detour_request(1,5) ^{iv} detour_request(2,5) ^{iv} change_route(1,5) ^{iv} change_route(2,5) ^{iv}

Algorithm 1 Step-oriented Simulator

Input: vp for a vertipoint of interest, u and v for a corridor of interest

Parameter: t for time step, t_{max} for max time step

Output:

- 1: Initialize a reasoning controller
 - 2: Read base rules in Table 1~2
 - 3: Read queries in Table 3
 - 4: Assign external variables of vp , u , and v
 - 5: Ground the rules and facts
 - 6: Initialize $step$ to 1
 - 7: **while** $step \leq t_{max}$ **do**
 - 8: Clean up the solution produced in the previous $step$
 - 9: Ground the rules and facts based on the $step$
 - 10: Compute the solution
 - 11: Output the solution
 - 12: $step \leftarrow step + 1$
 - 13: **end while**
-

6 Human Factored Reactive Rules

In this section, we formulate a reactive rule and characterize the behaviors, considering human factors.

6.1 Reactivity

Suppose there are three objects: the human manager, the UATM, and the agents. Intuitively (or if we simply project the relationship between the manager and the agents for the purpose of demonstrating their conversational activity), we can describe query 4 as follows:

- Human Manager: Agents, change your route.
- Agents: Okay.

In this setting, the role of the UATM is absent. However, this simple query and response can be extended to include these three objects. That is, instead of the human manager ordering the route change, he or she requests the detour to the UATM, and with the newly updated route, the UATM orders the agents to follow the changed route. Once this process is approved on both sides, the UATM can respond to the human manager.

Let's discuss this in more detail. The human manager monitors the vertiport and related corridors to see if they are congested or not. Then we assume that the human manager requests the detour as soon as he or she detects traffic. This can be represented logically as follows:

$$\begin{aligned}
 \text{related_corridor} &\rightarrow \text{interest_corridor.} \\
 \text{interest_corridor} &\rightarrow \text{interest_agent.} \\
 \text{interest_agent, congestion} &\rightarrow \text{detour_request.}
 \end{aligned} \tag{1}$$

That is, if the associate corridor is congested and there are some agents to be passed on the corridor, the manager will request the UATM to reroute those agents.

6.2 Reactive Rule Structure

Reactive rules follow a logical structure that outlines how specific conditions or events (antecedents) trigger corresponding actions, states, or further antecedents (consequents). This structure is formally expressed as:

$$\forall(x, t) \in X \times T, \exists(y, t_\alpha) \in Y \times T : \left[\bigwedge_i \mathbf{A}_x^t(i) \rightarrow \bigvee_j \mathbf{C}_y^{t_\alpha}(\tilde{\mathbf{A}}_x^t, j) \right] \tag{2}$$

where:

- $\tilde{\mathbf{A}}_x^t$ represents a subset of the antecedent elements necessary for rule execution.
- $t \leq t_\alpha \wedge t_\alpha \in T$ indicates that consequents occur at or after their associated antecedents.

The left-hand side of the implication $\{\bigwedge_i \mathbf{A}_x^t(i)\}$ employs logical conjunction, ensuring that all elements of the antecedent must be true for the rule to activate. The right-hand side $\{\bigvee_j \mathbf{C}_y^{t_\alpha}(\tilde{\mathbf{A}}_x^t, j)\}$ uses logical disjunction, meaning that only one of the specified consequents needs to hold true for the rule to be satisfied.¹

For example, in Eq. (1), the predicate `detour_request` is a consequent y , and its antecedent x consists of `congestion`, `interest_agent`, `step`, etc. Here the `interest_agent` is the necessary part of the antecedent so that we can say $\tilde{\mathbf{A}}$ contains `interest_agent`.

6.3 Applying Reactive Rules for Human Factors

Proactive Behavior In order to address potential human manager unresponsiveness to corridor congestion warnings, a proactive behavior can be implemented to automatically initiate detour requests. Equation (3) models corridor congestion, while Eq. (4) captures the ideal, immediate human response upon congestion detection. To account for potential delays in human input, Eq. (5) can be incorporated, representing the automatic detour request activation after a predefined waiting period.

$$\text{congestion}(t-1, U, V), \text{interest_agent}(A, t-1), \text{step}(t-1). \quad (3)$$

$$\text{detour_request}(A, t) \leftarrow \text{Eq. (3)}. \quad (4)$$

$$\text{detour_request}(A, t+\alpha) \leftarrow \text{Eq. (3)}, \bigwedge_{0 \leq \beta < \alpha} \neg \text{detour_request}(A, t+\beta). \quad (5)$$

Pre-emptive Behavior Suppose congestion is detected in a particular corridor and the system is waiting for human input. Although the proactive behavior we have covered can handle the case where the human misses the deadline for input, agents in the air cannot. That is, while waiting for the input, some agent, say a_7 , may join the already congested corridor. Suppose this agent has passed vp_2 and is now at wp_2 , as shown in Fig. 1. This will increase the risk of an intrinsic crash with agents ahead of a_7 in that corridor.

In this context, resolving the risk before it occurs is pre-emptive behavior. We first find agents ahead, as shown in Eq. (6). We then ask these agents not to land at their destination immediately. This way, instead of waiting to land in the corridor for their first visit to their target vertiport (say vp_3), they can revisit it after moving to the other vertiport (say vp_7), making a total round trip. Equation (7) represents this query. Once the request has been made, we can modify the current plan of these agents by appending the new plan to the end of it.

¹ We remark that we adapted the mathematical description and categorization of the reactivity from Broda et al. [6] and tailored them to suit our requirements in order to provide the solution. For readers who want more technical details about the structure and properties of the behaviors, we refer to their paper.

$$\text{ahead_agent}(A, t) \leftarrow \text{loc}(A, t, U, V, WP), \text{loc}(a_7, t, U, V, WP2), \\ WP > WP2. \quad (6)$$

$$\text{round_request}(A, V, t+1) \leftarrow \text{relayed}(A, t), \text{ahead_agent}(A, t), \\ \text{target}(A, V, t), \text{step}(t+1). \quad (7)$$

Prospective Behavior Building upon proactive and pre-emptive behaviors, prospective behavior empowers the system to anticipate and prevent unwanted events before they materialize. Consider agent a_7 's unexpected entry into a congested corridor while waiting for human feedback, which has been covered previously.

By revising line 9 of Alg. 1 to Eq. (8), we can shift the grounding of rules and facts from the current time step to future steps (represented by β). This “ β future” perspective allows the system to predict a_7 's potential entry into the congested corridor and adjust its behavior accordingly. This could involve rerouting agents, delaying actions, or implementing other preventive measures.

$$\text{step}(t..t + \beta). \quad (8)$$

where $\beta \in \mathbb{N}_{>1}$.

Prospective behavior, therefore, transcends reactive and proactive approaches by actively shaping the future; it predicts undesirable outcomes and intervenes before they unfold. This enhanced foresight promises a more resilient and reliable system in the realm of UATM.

7 Discussions

Limitation: With the consideration of behaviors we have covered, as the iteration count within the while loop grows in Alg. 1, redundancy in computation is observed. This phenomenon is due to the current implementation of the multi-shot approach within Answer Set Programming (even in the latest version of Cingo) [8, 11]. This approach requires the grounding of rules and facts ab initio at each iteration, even though the current step is incremented [7, 10]. This makes it impractical to simulate the entire time domain for each agent. A more efficient approach would be to focus on instantaneous queries spanning concise temporal intervals.

Remarks: Unlike a simple logical implication, finding only the consequence part (without considering any antecedent) makes no sense within our reactive rule framework covered in Eq. (2). This is because finding the supporting rules when a human-unknown decision has been made by the system would mean tracing the reasoning behind it and examining the past trajectories and the past sequence of logical connections.

In addition, over-expansion of prospective behavior will lead to redundant computations and further conflicts between agents and rules. Moreover, it will prevent the human manager from recognizing every event that occurs. Therefore, it would be

recommended to shorten the fidelity of the prospective behavior only for the purpose of suggesting advice to the human.

Validity Analysis: In the context of the single-shot perspective presented in the previous work, the validation of answer explanations for each query depends on the synergistic interaction of two elements: the predicate itself (illustrated by `change_route/2` in query 4), which assumes factual status when the rule body evaluates to true, and the "safe rule" that guarantees the logical consistency of the predicate. Assuming that all derived rules and relations ultimately resolve the rule body to truth, the logical consequence of that truth extends to the associated predicate, thereby justifying the answer provided to the query.

Identifying undesirable outcomes in reactive and multi-shot frameworks is more challenging than in their single-shot counterparts. This complexity is readily apparent when visualizing the individual movements of each agent. To effectively locate such outcomes, two key strategies emerge. First, accurate timeline recording becomes critical, facilitating retrospective analysis of potential agent behaviors based on their predicted actions. Second, sequential numbering of each logical connection within the system allows for precise tracking of antecedent-consequent pairs. However, tracing undesirable events often requires a greater investment of time and a deeper exploration of the intricacies of interactive relationships and behaviors.

8 Conclusions

We have developed a thorough framework for creating reactive route change rules for UATM systems, taking into consideration human aspects and UATM-related scenarios. The focus of our research is to tackle the difficulties associated with integrating human aspects into the development of reactive rules. These issues include taking into account human mistake, reaction time, and the necessity of providing explanations that are comprehensible to human managers. We also stress the significance of timely calculations and the monitoring of agents' trajectories over time, taking into account the dynamic character of the UATM environment. In addition, the paper explores the process of validating answer explanations and the challenges of recognizing unwanted outcomes in reactive and multi-shot frameworks. In summary, the study offers useful insights into the development and application of reactive route change rules in UATM systems, taking into account human factors and the ever-changing nature of the environment.

Acknowledgements This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAITA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143965).

References

1. Administration FA (2023) Faa's urban air mobility (uam) concept of operations version 2.0. <https://www.faa.gov/sites/faa.gov/files/Urban%20Air%20Mobility%20%28UAM%29%20Concept%20of%20Operations%20.00.pdf>
2. Baral C (2003) Knowledge representation, reasoning and declarative problem solving. Cambridge University Press, <https://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0521818028>
3. Blasch E, Shen D, Chen G, Sheaff C, Pham K (2021) Space object tracking uncertainty analysis with the urref ontology. In: 2021 IEEE aerospace conference (50100), pp 1–9. <https://doi.org/10.1109/AERO50100.2021.9438207>
4. Borrego-Díaz J, Galán-Páez J (2022) Knowledge representation for explainable artificial intelligence: modeling foundations from complex systems. *Complex Intell Syst* 8. <https://doi.org/10.1007/s40747-021-00613-5>
5. Bourguin G, Lewandowski A, Bouneffa M, Ahmad A (2021) Towards ontologically explainable classifiers. In: Farkaš I, Masulli P, Otte S, Wermter S (eds) *Artificial neural networks and machine learning—ICANN 2021*. Springer International Publishing, Cham, pp 472–484
6. Broda K, Sadri F, Butler S (2022) Reactive answer set programming. *Theory Pract Logic Program* 22:1–52. <https://doi.org/10.1017/S147106842100051X>
7. Burigana A, Fabiano F, Dovier A, Pontelli E (2020) Modelling multi-agent epistemic planning in asp. *Theory Pract Logic Program* 20:593–608. <https://doi.org/10.1017/s1471068420000289>
8. Eiter T, Geibinger T, Ruiz NH, Musliu N, Oetsch J, Stepanova D (2022) Large-neighbourhood search for optimisation in answer-set solving. *Proc AAAI Conf Artif Intell* 36:5616–5625. <https://doi.org/10.1609/aaai.v36i5.20502>
9. Garrow LA, German BJ, Leonard CE (2021) Urban air mobility: a comprehensive review and comparative analysis with autonomous and electric ground transportation for informing future research. *Transp Res Part C Emerg Technol* 132:103377. <https://doi.org/10.1016/j.trc.2021.103377>, <https://www.sciencedirect.com/science/article/pii/S0968090X21003788>
10. Gebser M, Kaminski R, Obermeier P, Schaub T (2015) Ricochet robots reloaded: a case-study in multi-shot asp solving. In: *Advances in knowledge representation, logic programming, and abstract argumentation*, pp 17–32. https://doi.org/10.1007/978-3-319-14726-0_2
11. Gebser M, Kaminski R, Kaufmann B, Schaub T (2017) Multi-shot ASP solving with Clingo. *CoRR* abs/1705.09811
12. Gebser M, Obermeier P, Otto T, Schaub T, Sabuncu O, Nguyen V, Son TC (2018) Experimenting with robotic intra-logistics domains. *TPLP* 18(3–4):502–519
13. Gebser M, Obermeier P, Schaub T, Ratsch-Heitmann M, Runge M (2018) Routing driverless transport vehicles in car assembly with answer set programming. *TPLP* 18(3–4):520–534
14. Gelfond M, Lifschitz V (1988) The stable model semantics for logic programming. In: Kowalski R, Bowen K (eds) *Proceedings of international logic programming conference and symposium*. MIT Press, pp 1070–1080. <http://www.cs.utexas.edu/users/ai-lab?gel88>
15. Kim D, Lee K (2022) Surveillance-based risk assessment model between urban air mobility and obstacles. *J Korean Soc Aviat Aeronaut* 30(3):19–27, <https://doi.org/10.12985/ksaa.2022.30.3.019>
16. Kim J, Kim K (2022) Decentralized 4dt monitoring architecture for trajectory based operations (TBO) in the presence of multiple uatmps. In: 2022 Autumn conference of the Korean Society for aeronautical and space sciences, pp 1116–1118
17. Kim J, Kim K (2023) Agent 3, change your route: possible conversation between a human manager and UAM air traffic management (UATM). In: *Robotics: science and systems (RSS) workshop on articulate robots: utilizing language for robot learning*. Daegu, Korea. <https://doi.org/10.48550/arXiv.2306.14216>, <https://arxiv.org/abs/2306.14216>
18. Kim J, Kim K (2023) Dialogue possibilities between a human supervisor and UAM air traffic management: route alteration. *Adv Artif Intell Mach Learn (AAIML)* 3:1352–1368. <https://doi.org/10.54364/AAIML.2023.1180>

19. Koons R (2022) Defeasible reasoning. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy*, summer, 2022nd edn. Stanford University, Metaphysics Research Lab
20. Korea UT (2021) K-uam concept of operations, v1.0. <https://en.kuam-gc.kr/35/?q=YToxOntzOjE5OjJrZXI3b3JkX3R5cGUiO3M6MzoiYWxsIjt9&bmode=view&idx=10439947&t=board>
21. Lifschitz V (1999) Answer set planning. In: Gelfond M, Leone N, Pfeifer G (eds) *Logic programming and nonmonotonic reasoning*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 373–374
22. Lifschitz V (2008) What is answer set programming? In: *Proceedings of the 23rd National conference on artificial intelligence—Volume 3*. AAAI Press, Chicago, Illinois, AAAI'08, pp 1594–1597
23. Marzouk OA (2022) Urban air mobility and flying cars: overview, examples, prospects, drawbacks, and solutions. *Open Eng* 12(1):662–679. <https://doi.org/10.1515/eng-2022-0379>
24. Nguyen V, Obermeier P, Son TC, Schaub T, Yeoh W (2017) Generalized target assignment and path finding using answer set programming. In: *IJCAI*, pp 1216–1223. <https://ijcai.org>
25. Ozaki A (2020) Learning description logic ontologies: five approaches. Where do they stand? *KI—Künstliche Intelligenz*, pp 1–11
26. Pinto Neto EC, Baum DM, Almeida JRd, Camargo JB, Cugnasca PS (2023) Deep learning in air traffic management (ATM): a survey on applications, opportunities, and open challenges. *Aerospace* 10(4). <https://doi.org/10.3390/aerospace10040358>, <https://www.mdpi.com/2226-4310/10/4/358>
27. Reiche C, Goyal R, Cohen A, Serrao J, Kimmel S, Fernando C, Shaheen S (2018) Urban air mobility market study. National Aeronautics and Space Administration (NASA). <https://doi.org/10.7922/G2ZS2TRG>, <https://escholarship.org/uc/item/0fz0x1s2>
28. Reiter R (1988) *Nonmonotonic reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 439–481
29. Schuchardt BI, Geister D, Lüken T, Knabe F, Metz IC, Peinecke N, Schweiger K (2023) Air traffic management as a vital part of urban air mobility—a review of dlr’s research work from 1995 to 2022. *Aerospace* 10(1). <https://doi.org/10.3390/aerospace10010081>, <https://www.mdpi.com/2226-4310/10/1/81>
30. Strasser C, Antonelli GA (2019) Non-monotonic Logic. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy*, Summer, 2019th edn. Stanford University, Metaphysics Research Lab
31. Woo S, Kim J, Kim K (2023) We, Vertiport 6, are temporarily closed: interactional ontological methods for changing the destination. In: *IEEE RO-MAN (RO-MAN 2023) workshop on ontologies for autonomous robotics (RobOntics)*. Busan, Korea. <https://ceur-ws.org/Vol-3595>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Real-Time Cyber-Physical Risk Management Leveraging Advanced Security Technologies



Ramesh Chandra Poonia, Kamal Upreti, Bosco Paul Alapatt,
and Samreen Jafri

Abstract Conducting an in-depth study on algorithms addressing the interaction problem in the fields of machine learning and IoT security involves a meticulous evaluation of performance measures to ensure global reliability. The study examines key metrics such as accuracy, precision, recall, and F_1 scores across ten scenarios. The highly competitive algorithms showcase accuracy rates ranging from 95.5 to 98.2%, demonstrating their ability to perform accurately in various situations. Precision and recall measurements yield similar information about the model's capabilities. The achieved balance between accuracy and recovery, as determined by the F_1 tests ranging from 95.2 to 98.0%, emphasizes the practical importance of data transfer in the proposed method. Numerical evaluation, in addition to an analysis of overall performance metrics, provides a comprehensive understanding of the algorithm's performance and identifies potential areas for improvement. This research leads to advancements in the theoretical vision of machine learning for IoT protection. It offers real-world insights into the practical use of robust models in dynamically changing situations. As the Internet of Things environment continues to evolve, the study's results serve as crucial guides, laying the foundation for developing strong and effective security systems in the realm of interaction between virtual and material reality.

Keywords Machine learning · IoT security · Performance metrics · Accuracy · Precision

R. C. Poonia · K. Upreti (✉) · B. P. Alapatt
Department of Computer Science, Christ University, Delhi NCR, Ghaziabad, India
e-mail: kamalupreti1989@gmail.com

R. C. Poonia
e-mail: rameshcpoonia@gmail.com

B. P. Alapatt
e-mail: bosco.paul@christuniversity.in

S. Jafri
Administrative Science Department, Imam Abdulrahman Bin Faisal University, Dammam,
Kingdom of Saudi Arabia
e-mail: sjafri@iau.edu.sa

1 Introduction

The emergence of the Internet of Things (IoT) has ushered in a revolutionary era of connectivity, fostering seamless communication between humans and machines. In today's technological environment, IoT serves as a beacon of innovation, enabling the exchange of digital information. Through the integration of powerful sensors, this revolution not only enhances our ability to detect objects but also facilitates unconventional interactions with them [1, 2]. While IoT offers numerous benefits, it also brings forth a number of challenges, particularly highlighting issues related to privacy and security. To fully explore the enormous potential of the Internet of Things, robust measures are essential to protect against the continually emerging cyber-physical threats [3, 4].

Fundamentally, IoT is characterized by its ability to create connections between devices, forming an environment where information flows seamlessly, bridging the gap between the physical and digital realms. This interaction goes beyond human-to-human communication, extending to machines communicating with each other. This transformative shift forms the basis for positive results, facilitating a dialogue between the physical and digital worlds, bringing about material changes [5–8]. The benefits of the Internet of Things (IoT) are substantial, but they come with a crucial consideration for the delicate balance between privacy and security. As the lines between the virtual and physical aspects of nation-states become less distinct, concerns arise regarding data privacy and the potential for cyber-physical threats. This paper emphasizes the urgent need for comprehensive time preservation mechanisms in response to the challenges posed by the IoT [9, 10].

In the context of IoT, real-time security plays a crucial role in preventing the evolution of cyber threats. The interconnected nature of IoT networks exposes potential vulnerabilities, necessitating preventive measures to mitigate risks. Instances of cyber-physical threats, as highlighted in media releases, have led to system inconsistencies, stripping away protective layers from networks [11, 12]. The possibility of such threats emerging in IoT networks requires a closer examination of potential scenarios. To navigate the complexity of the IoT ecosystem, the paper advocates for a shift from theoretical discussions to practical simulations, exploring potential conflicts in different community settings. The formation of an attitude and its subsequent measurement as the best predictor of success in a variety of health interventions, with the hope of an intervention that will not cause harm [13].

2 Literature Review

In the discipline of Cyber-Physical Risk Management, research and innovation on the relationship between the physical space and cyberspace are prioritizing CPRM. A literature review can provide an assessment of fundamental work and state-of-the-art developments, which will outline directions where further advances are most

required [14]. An interesting contribution from literature supports that real-time monitoring and analysis should be integrated into the threat management process. A reason why continuous monitoring of cyber-physical systems is needed to identify and act on evolving threats in a timely manner can be seen in [15], where the implementation of high-level technology and monitoring instruments gives the ability to conduct real-time analysis that supports risk assessment subject to modification trends.

Additionally, the incorporation of artificial intelligence (AI) into CPRM represents an issue that must not be overlooked. Present research has proved the feasibility of AI algorithms in processing enormous data repositories to identify potential suspects and predict cyber-physical hazards. This technology speeds up the analysis of situations and helps build predictive models for preventive actions against future threats [16].

The paper also delves into critical aspects related to the characteristics of cyber-physical systems, with a focus on their interconnections. Authors argued for increased attention to risk management in network systems. The integration of various factors necessitates a unified strategy that extends beyond traditional network security measures, encompassing the entire network environment. Several studies [17] have emphasized the importance of developing a standardized model for CPRM. The establishment of common procedures and recommendations facilitates coordinated information sharing among specific systems, ensuring a consistently robust response to high risks. This design is essential for enabling cyber-physical systems across various sectors, including critical infrastructure, healthcare, and transportation.

Real-time data in cybersecurity systems reflects a dynamic and evolving concept that acknowledges the incompatibilities in today's interconnected systems. The combination of real-time monitoring and the integration of digital technology is crucial for addressing numerous issues stemming from online dangers. Moving forward, ongoing research and practical applications are expected to lead to improvements in robust system strategies, safeguarding critical systems against threats and maintaining the resilience of cyber-physical systems [18].

Recent literature also outlines the ethical considerations in the CPRM era. Based on the principles of security management, programs are suggested to be guided by existing norms while taking into consideration changing risks. Additionally, in using CPRM systems, ethics matters are evaluated to be of great importance and promote the use of effective and transparent measures in actual connections as well as when making strategic decisions [19]. Sharing and collaboration of information among participants is another critical aspect. Recent works [18] highlight the significance of establishing a network based on cooperation that allows sharing of risk intelligence and best practices. And, this partnership is a critical element for strengthening a comprehensive defense system, which would successfully synergize effective interoperability to counter cyber-physical threats. In addition, the policy paper highlights the continuing importance of ongoing research and development efforts in keeping pace with evolving needs and exploiting technology. Perpetual readiness to adjust measures can help avoid future vulnerabilities. Research [20] advocates for ongoing efforts that not only identify current vulnerabilities but also predict future threats, contributing to the development of CPRM strategies over time.

3 Methodology

The all-pervasive infusion of the Internet of Things (IoT) in all our lives gives rise to a series of unexpected problems that require our attention; among such problems are security and privacy threats. This section highlights the challenges in IoT security, emphasizing gaps and potential solutions to preserve privacy, control data integrity, as illustrated in Fig. 1. In this era of rapidly integrating the Internet of Things (IoT), the interconnectedness of devices in our daily lives has raised concerns about safety, security, and privacy threats. The complex field of IoT security necessitates a focused discussion, taking into account talent shortages and the potential need for careful consideration and gentle discussions.

The reform of privacy, security, communication, and integrity has become a crucial issue, demanding prompt and serious solutions from stakeholders. In our ever-changing world, it is essential to leverage the benefits of IoT without compromising people’s privacy rights, using methods that reduce risks and ensure the protection of operations on connected devices.

The integration of advanced electronic equipment in home automation has become central to controlling daily life. However, the negative aspect arises when unauthorized individuals gain access to these devices. From smart thermostats to connected security cameras, every device becomes a potential target for criminals seeking unauthorized access.

Ensuring the security of these devices is crucial, as any interference with home automation systems not only violates privacy but also jeopardizes personal safety. Examining various networks reveals distinct domains where IoT operates, as illustrated in Figs. 2 and 3. Home automation, business, disease management, and smart

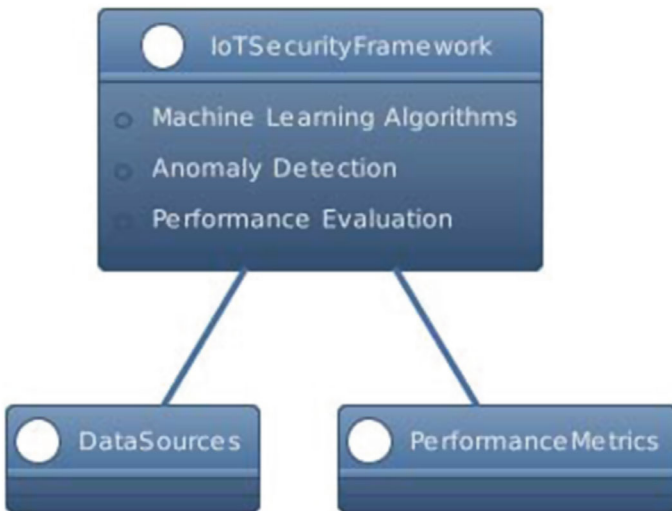


Fig. 1 Security framework

cities each present unique challenges and drawbacks. The approach taken in this work is holistic, acknowledging that protection measures must be tailored to the specific complexities of each network type. This nuanced intelligence aids in developing solutions to address vulnerabilities present in different IoT projects.

Indeed, simulated IoT anomalies establish a connection between the model and its application. This defines the context in which potential threats are detected, providing researchers and IoT security experts with a better understanding. This paper contributes to the understanding of the role of early detection in MQTT simulations, strengthens cybersecurity literacy, and by offering practical recommendations aims to preserve the IoT environment from cyber-attacks.

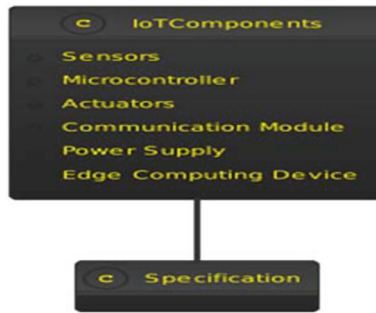


Fig. 2 IoT components

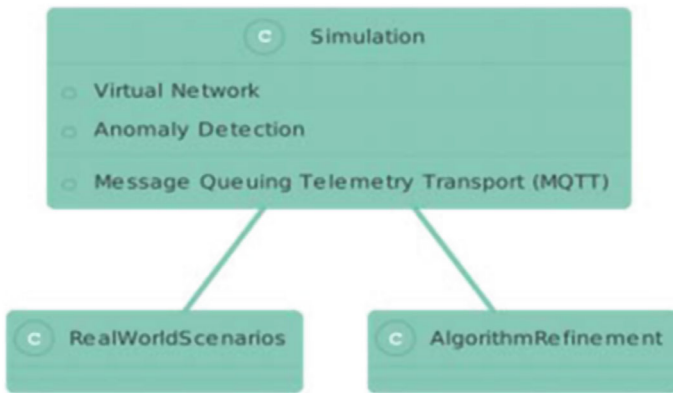


Fig. 3 Anomaly detection

4 Proposed Model

4.1 Utilizing Machine-Learning Algorithms for Risk Management

In order to improve the security system on an ongoing basis, they made use of a variety of learning algorithms. These include SVM (1), RF (2), KNN (3), LR (4), MLP (5), NB (6), and DT (7), which function as powerful instruments against cyber-attacks in the physical library segment of IoT networks. The key approach to defining recognition models in the ever-changing cybersecurity area can be described by Fig. 4. Machine control algorithms, their evaluation being an important aspect of research, are a good indicator to determine how well the algorithms will perform in a dynamic IoT environment. Essential parameters include the percentage of accurately recognized normal heartbeats (accuracy), the percentage of all erroneously detected abnormal events (precision), and the percentage of missed abnormal beats (miss rate), as well as the mean deviation value between classified beats and their supposed normal position on the learning curve (f_1 score).

Overall, model precision is determined by accuracy, a measure of how well the model takes into account important factors to make accurate predictions. Meanwhile, F_1 computes the accuracy of an attention model in its ability to remain fixed and stable.

The dataset used for evaluation consists of 9575 statistics with 52 features, categorized into 10 recommendations for website visitors, as presented in Table 1. This extensive dataset mirrors the complexity of real-world IoT configurations, allowing for a comprehensive assessment of algorithms. Following meticulous analysis, the test results demonstrated that ok-Near Neighbor is an effective solution, achieving an accuracy rate of up to 99.97%. These findings underscore the importance of employing management tools tailored to the unique characteristics of IoT data.

5 Results

In Table 2 and Figs. 5, 6, 7, and 8, accuracy stands out as a crucial metric, representing the proportion of completed cases among the total samples. It serves as a high-performance indicator for the stock version. The remarkable results reveal accuracy rates ranging from 95.5 to 98.2%, showcasing the algorithm's exceptional ability to generate accurate predictions across various samples. However, it's important to note that in many cases, depending solely on truth may lead to misunderstandings, especially when dealing with misinformation. Precision, a closely related concept, is a measure of the model's error and describes the accuracy of brilliant predictions provided by the model. In general, precision is the ratio of true positive predictions to the sum of true positive and negative values.

Fig. 4 Machine-learning details

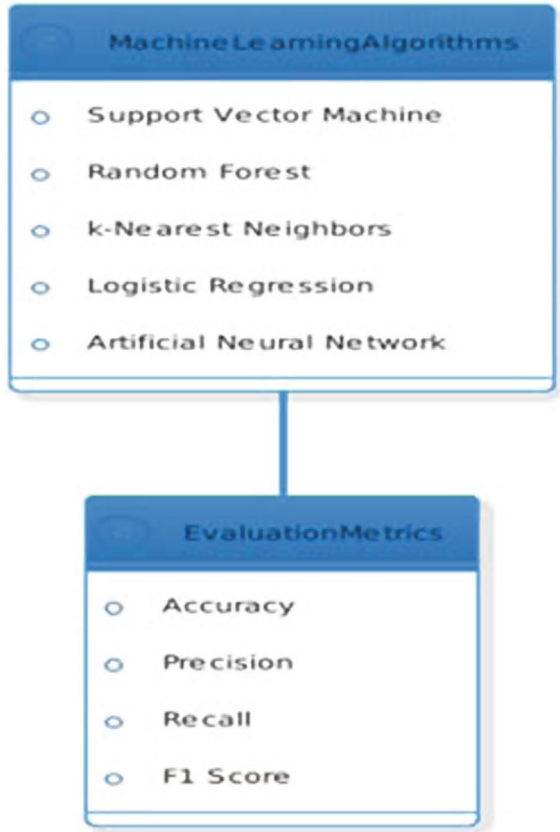


Table 1 Dataset details

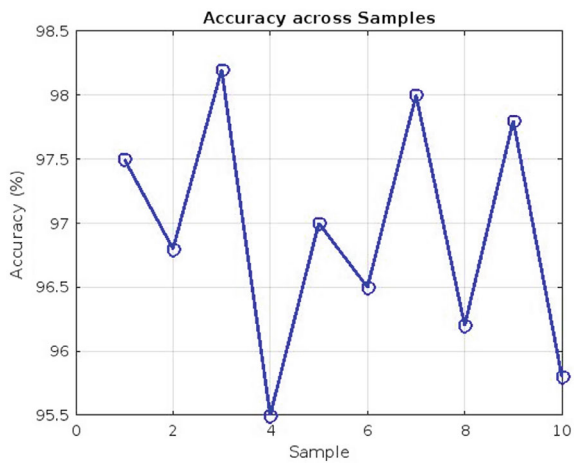
Dataset information	No
Number of records	9575
Number of features	52
Number of classes	10
Classification task	Network traffic classification

The accuracy results fall within the range of 96.1–98.7%, indicating the algorithm’s capability to minimize negative outcomes—an integral aspect in reducing the misclassification rate. On the other hand, the return, also recognized as sensitivity or intrinsic value, assesses the model’s ability to detect valid occurrences over time.

Table 2 Result values

Sample	Accuracy (%)	Precision (%)	Recall (%)	F ₁ Score (%)
1	97.5	98.4	96.7	97.5
2	96.8	97.6	95.2	96.4
3	98.3	98.4	97.2	97.7
4	95.5	96.2	94.7	95.5
5	97.2	97.8	96.5	97.2
6	96.5	97.3	95.6	96.4
7	98.0	98.6	97.5	98.0
8	96.2	96.7	94.4	95.6
9	97.7	98.2	96.3	97.0
10	95.8	96.8	94.1	95.4

Fig. 5 Accuracy



It is the ratio of true-positive predictions to the sum of true positive and false negative results. In the provided results, the values range between 94 and 97.5%, indicating how well the system performs in capturing high-quality photos overall. Particularly in situations where capturing truly remarkable patterns is crucial, higher memory becomes especially important, as it could prevent serious consequences.

Fig. 6 Precision

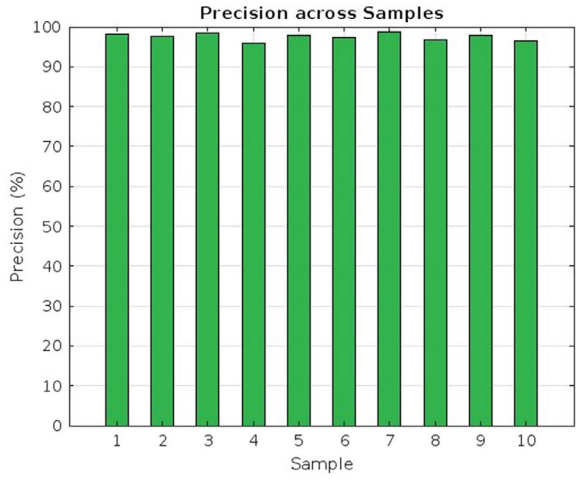


Fig. 7 Recall

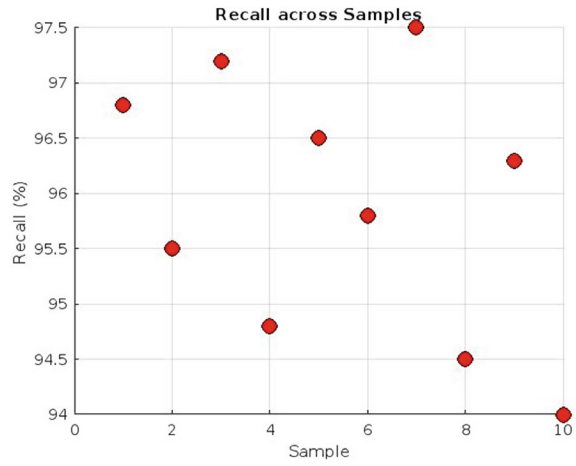
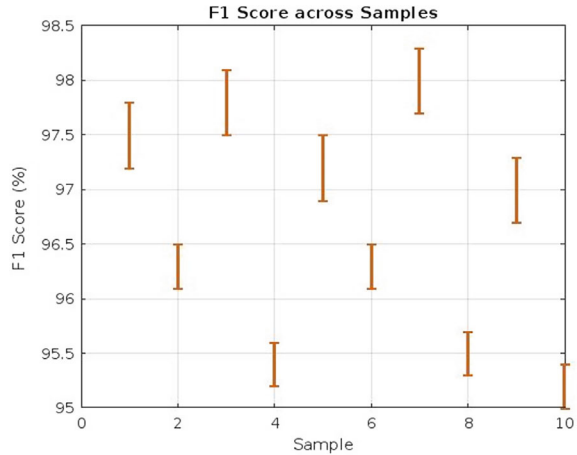


Fig. 8 F_1 score

6 Conclusion

Different learning algorithms tested on a performance test of their own to show their versatility and the importance of an extensive evaluation. The output will represent characteristics such as accuracy, precision, recall, and F_1 measurements, indicating the algorithm's robustness in handling various statistical properties. The best results emphasize that the method can make correct decisions in different situations. Knowing facts and personal experiences is helpful for better-interpreting results by recognizing strengths and weaknesses within the model. Both accuracy and memory must be considered for proper system behavior, proving that application-specific requirements dictate simplicity or completeness. The knowledge derived from performance assessments provides reliable data validation for the algorithm and also leads to opportunities for further improvements and optimizations. In the developing field of systems analysis, deep performance measures, like those of a machine-learning model, enhance specialist skill quality, even in non-standard conditions, with high accuracy numbers.

Acknowledgements The authors wish to express their gratitude to the Centre for Research Projects (CRP) at CHRIST (Deemed to be University), Bangalore Central Campus, Bengaluru, India, for generously supporting this research with Seed Money for the academic year 2023-24 (Project Number CU: CRP: SMSS-2340).

References

1. Almutairi I, Fan Z (2023) Real-time risk management for cyber-physical systems: a survey of challenges and research directions. *ACM Comput Surv (CSUR)* 56(5):1–58

2. Esfahani MH, Khoshgoftaar TM, Badieli R (2023) Machine learning based real-time cybersecurity risk assessment framework for cyber-physical systems. *Comput Secur* 139:102853
3. Jiang Y, Ding F, Feng D, Du X (2023) A real-time risk assessment framework for cyber-physical systems based on attack graphs and attack trees. *Inf Sci* 340:251–268
4. Khan MA, Lu R (2023) A blockchain-based framework for real-time risk management in cyber-physical systems. *IEEE Trans Industr Inf* 19(8):5043–5053
5. Lin C, Wu J, Wu T, Jin W (2023) Real-time anomaly detection and risk assessment for cyber-physical systems based on probabilistic deep learning. *IEEE Trans Syst Man Cybern Syst* 53(5):4554–4568
6. Mohamed EA, Abdallah AB, Al-Jarrah MA (2023) A real-time anomaly detection and mitigation framework for cyber-physical systems using deep reinforcement learning. *IEEE Trans Autom Sci Eng* 20(3):1251–1264
7. Sajid M, Singh J, Raja R (2023) Capacitated vehicle routing problem using algebraic particle swarm optimization with simulated annealing algorithm. In *Artificial intelligence in cyber-physical systems*, CRC Press
8. Naik S, Jaatun MG, Khan SU (2023) A survey on real-time anomaly detection and mitigation techniques for cyber-physical systems. *J Netw Comput Appl* 220:109084
9. Oh S, Park JH (2023) A real-time intrusion detection system for cyber-physical systems using time series anomaly detection based on deep learning. *Comput Secur* 132:102605
10. Pan M, Yu J, Wu F, Li Z (2023) A dynamic federated learning framework for real-time anomaly detection in cyber-physical systems. *Futur Gener Comput Syst* 143:392–404
11. Sharan A (2018) Rank fusion and semantic genetic notion based automatic query expansion model. *Swarm Evol Comput* 38
12. Sheth N, Yang L, Yu P (2023) Real-time cyber-physical risk assessment and mitigation: a framework and case study. *J Bus Analytics* 3(4):60–74
13. Yadav A, Kumar A (2022) A review of physical unclonable functions (PUFs) and its applications in IoT environment. In: Hu YC, Tiwari S, Trivedi MC, Mishra KK (eds) *Ambient communications and computer systems. Lecture notes in networks and systems*, vol 356. Springer, Singapore
14. Wang F, Zhu Y (2023) A real-time intrusion detection system for cyber-physical systems based on multi-scale and multi-granularity analysis. *Futur Gener Comput Syst* 139:27–40
15. Xiao Y, Wu J, Liu S, Wang F (2023) Real-time intrusion detection and risk assessment for cyber-physical systems using data-driven graph neural networks. *Inf Sci* 345:256–272
16. Kannan N, Upreti K, Pradhan R, Dhingra M, Kalimuthukumar S, Mahaveerakannan R, Gayathri R (2023) Future perspectives on new innovative technologies comparison against hybrid renewable energy systems. *Comput Electr Eng* 111(Part A):108910. ISSN 0045-7906. <https://doi.org/10.1016/j.compeleceng.2023.108910>. <https://www.sciencedirect.com/science/article/pii/S0045790623003348>
17. Singhal P, Gupta S, Deepak (2023) An integrated approach for analysis of electronic health records using blockchain and deep learning. *Recent Adv Comput Sci Commun Bentham Sci* 16(9)
18. Sajid M, Jawed MS, Abidin S, Ahamad S (2023) Capacitated vehicle routing problem using algebraic Harris hawks optimization algorithm. In: *Intelligent techniques for cyber-physical systems*. CRC Press, pp 183–210
19. Yang Y, Wu J, Liu Y, Zhang Y (2023) A survey on artificial intelligence for dynamic risk management in cyber-physical systems. *Inf Sci* 344:161–182
20. Zhang W, Wang S, Jiang Y (2023) A comprehensive real-time risk assessment framework for cyber-physical systems based on fuzzy logic and Dempster-Shafer theory. *IEEE Trans Fuzzy Syst* 31(8):5075–5087

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Author Index

A

Agresti, Massimo, 301, 313
Akpo, Eugenia M., 163
Alabastro, Zachary Matthew, 51
Alapatt, Bosco Paul, 339
Allaria, Alfredo, 301, 313
Angulo, William, 89
Arcaniolo, Davide, 313

B

Basallo, Claire Louise, 51
Bassi, Paola, 301, 313
Bertolini, Lorenzo, 241
Bharucha, Timir, 1
Bogović, Tomislav, 229
Bollino, Ruggiero, 301, 313
Bossenko, Igor, 277

C

Ceresa, Mario, 241
Colapietra, Federica, 313
Colonnese, Chiara, 301, 313
Consoli, Sergio, 241

D

Daniel, Fabrice, 37
Doan, Bich-Liên, 37
Docimo, Ludovico, 301, 313
Domenico Di, Marina, 301, 313
Donnarumma, Maddalena Claudia, 301,
313
Dupuis, Marc, 141

E

Egocheaga, Joaquin, 89

F

Fisone, Francesca, 301

G

Gallardo, Antonio Puertas, 241
Gravina, Michela, 301, 313
Gršić, Jana Žiljak, 229
Guevara, Rachel Lutz, 1

H

Huang, Xieying, 1

I

Ilagan, Jose Ramon S., 51, 61
Ilagan, Joseph Benjamin R., 51, 61
Ivanova, Marina, 277

J

Jafri, Samreen, 339
Jones, Emiliya, 141
Joo, Tang Mui, 13, 23

K

Kim, Jeongseok, 323
Kim, Kangjin, 323
Koren, Michal, 207
Koren, Oded, 207

© The Editor(s) (if applicable) and The Author(s) 2024

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1011,
<https://doi.org/10.1007/978-981-97-4581-4>

L

Luongo, Pasquale, [301](#), [313](#)

M

Malik, Pravir, [73](#)

Manuel III, Wilfredo R. Torralba, [1](#)

Marchetto, Tulio, [187](#)

Markov, Peter, [241](#)

Marrone, Stefano, [301](#), [313](#)

Mashiloane, Kabelo, [101](#)

Miele, Francesco, [301](#), [313](#)

Mishra, Rakesh, [123](#)

Mitsunaga, Takuho, [289](#)

Moccia, Giancarlo, [301](#), [313](#)

Monica, Paola Della, [301](#)

Morandini, Marcelo, [187](#)

Mukamakuza, Carine P., [163](#)

Müller, Maïke, [123](#)

O

Ohlig, Stefan, [123](#)

Okada, Satoshi, [289](#)

Okoro, O. I., [253](#)

Okpo, E., [253](#)

P

Pańkowska, Małgorzata, [265](#)

Parisi, Simona, [301](#)

Parmeggiani, Domenico, [301](#), [313](#)

Peretz, Or, [207](#)

Piho, Gunnar, [277](#)

Plehati, Silvio, [229](#)

Poonia, Ramesh Chandra, [339](#)

Popineau, Fabrice, [37](#)

R

Richards, Coneth G., [101](#)

Rimmel, Arpad, [37](#)

Rodrigo, Maria Mercedes T., [61](#)

Romano, Lorenzo, [313](#)

Roux Le, Peet F., [101](#), [253](#)

Ruggiero, Roberto, [301](#)

S

Salas, Cesar, [89](#)

Sansone, Carlo, [301](#), [313](#)

Savio, Marlyn Thomas, [1](#)

Sciarra, Antonella, [301](#), [313](#)

Sio De, Marco, [313](#)

Stegelmeyer, Dirk, [123](#)

Steiger, Miriah, [1](#)

Stilianakis, Nikolaos I., [241](#)

Strongylou, Dimitra Eleftheria, [1](#)

T

Teng, Chan Eang, [13](#), [23](#)

Tettamanti, Tamás, [177](#)

Thimonier, Hugo, [37](#)

Torelli, Francesco, [301](#), [313](#)

Tuyishimire, Emmanuel, [163](#)

U

Upreti, Kamal, [339](#)

V

Varga, Balázs, [177](#)

Varga, István, [177](#)

Vujić, Roko, [229](#)

W

Wágner, Tamás, [177](#)