FAU Studien aus der Informatik 8

Dominik Schuldhaus

Human Activity Recognition in Daily Life and Sports Using Inertial Sensors

FAU UNIVERSITY PRESS

Dominik Schuldhaus

Human Activity Recognition in Daily Life and Sports Using Inertial Sensors

FAU Studien aus der Informatik

Band 8

Dominik Schuldhaus

Human Activity Recognition in Daily Life and Sports Using Inertial Sensors

Erlangen
FAU University Press
2019

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Der vollständige Inhalt des Buchs ist als PDF über den OPUS Server der Friedrich-Alexander-Universität Erlangen-Nürnberg abrufbar:
https://opus4.kobv.de/opus4-fau/home

Bitte zitieren als
Schuldhaus, Dominik. 2019. *Human Activity Recognition in Daily Life and Sports Using Inertial Sensors.* FAU Studien aus der Informatik Band 8. Erlangen: FAU University Press.
DOI: 10.25593/978-3-96147-226-0

Verlag und Auslieferung:
FAU University Press, Universitätsstraße 4, 91054 Erlangen

Druck: docupoint GmbH

ISBN: 978-3-96147-225-3 (Druckausgabe)
eISBN: 978-3-96147-226-0 (Online-Ausgabe)
ISSN: 2509-9981
DOI: 10.25593/978-3-96147-226-0

Human Activity Recognition in Daily Life and Sports Using Inertial Sensors

Erkennung Menschlicher Aktivität im Alltag und Sport
unter Verwendung von Inertialsensoren

Der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr.-Ing.

vorgelegt von

Dominik Schuldhaus

aus Nürnberg

Als Dissertation genehmigt
von der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	12. Februar 2019
Vorsitzender des Promotionsorgans:	Prof. Dr.-Ing. Reinhard Lerch
Gutachter:	Prof. Dr. Bjoern M. Eskofier
	Prof. Dr. Kristof van Laerhoven

Abstract

Human Activity Recognition (HAR) deals with the automatic recognition of physical activities and plays a major role in the health and sports sector. Knowledge about the performed activities can be used to monitor compliance regarding physical activity recommendations, investigate the causes of physical activity behavior, implement sport-specific training programs, and replicate the physical demands during sport competition.

Currently available tools for HAR can be divided into subjective and objective techniques. Subjective techniques often rely on questionnaires which involve problems in the reliability when recalling activities. In contrast, wearable computing systems such as activity trackers provide an objective solution for HAR. Small, lightweight, and low-power sensors, e.g. Inertial Measurement Units (IMUs), measure human movement. Machine learning algorithms are further applied to the acquired sensor data in order to enable an automatic recognition of physical activities.

In this thesis, algorithms for IMU-based HAR are introduced and evaluated considering daily as well as sport-specific activities. Daily activities are taken from the 'Compendium of Physical Activities' and include e.g. sitting, washing dishes, and climbing stairs. Soccer is chosen as an example sport due to the high number of actively involved people. The considered soccer-specific activities mainly include full-instep and side-foot kicks. Besides the development and implementation of algorithms, mandatory extensions regarding the design of HAR systems are further identified and future research directions are provided. In the following paragraphs, the contributions of this thesis are further explained.

Various algorithms are described which are able to infer the daily and soccer-specific activities regarding activity type and intensity. Three HAR algorithms classify daily activities using a hierarchical architecture or decision level fusion. The proposed approaches achieved accuracy values above 85 % and reduce the retraining effort in case of environmental changes such as new activities or sensors. Compared to current research work, the algorithms provide more generic components which can automatically be adapted to different application settings. The concept of decision level fusion is further applied to estimate the expended energy during treadmill walking. The system achieved a Root Mean Square Error (RMSE) of 0.42 MET.

Two HAR algorithms are introduced to classify full-instep as well as side-foot soccer kicks using a hierarchical architecture or expert

knowledge about kick phases. The proposed approaches achieved accuracy values above 85 %. One HAR algorithm is further proposed which estimates the ball speed after a full-instep kick. The system achieved a Mean Absolute Error (MAE) of 5.5 km/h. The soccer-specific approaches provide a better applicability to classify kicks and estimate ball speed in more realistic conditions, e.g. during games, compared to current approaches in literature.

Besides algorithms for activity inference, two novel techniques in HAR research are introduced and assessed. First, a low-cost database fusion strategy is described and investigated which aims at increasing the amount of instances which could be used to train and test HAR systems. The technique does not require a time-consuming collection of additional data. Second, a common evaluation framework is presented which could be used to provide a fair comparison of various activity inference methods. The framework is used to compare proposed HAR algorithms to state-of-the-art algorithms from literature based on the same benchmark dataset and experimental setup. The framework enables a better identification of the best algorithmic approach for a certain application compared to current approaches in literature.

The soccer-specific HAR algorithms are further integrated in a novel application for sensor-driven video summary generation. Soccer players are provided with individual highlight reels. The selection of the corresponding highlight scenes are driven by the decision of IMU-based HAR algorithms. Video and audio effects are further added to the highlight reels in specific automatically detected kick phases.

The previously mentioned algorithms are evaluated on small data. The huge amount of data collected by Internet-connected devices coincide with the trend toward the generation of Big Data. Although Big Data offers new possibilities in HAR research, e.g. automated detection of deviant behavior, the current design of HAR systems does not consider Big Data concepts and tools. In this thesis, Big Data is introduced to the HAR research field. A HAR-related definition of Big Data as well as Big Data specific extensions of the HAR system design are proposed and open issues as well as future research directions are further provided.

The contributions, which are presented in this thesis, address current limitations in IMU-based HAR research and enable a better assessment of humans' behavior in health and sports. Future multi-sensor HAR systems, which integrate the proposed algorithms and further consider the Big Data specific components, will provide a robust, holistic, and long-term analysis of humans' physical state.

Zusammenfassung

Die automatische Erkennung menschlicher Aktivität spielt eine große Rolle im Gesundheits- und Sportsektor. Das Wissen über die Aktivität, die durchgeführt wird, kann verwendet werden, um die Einhaltung von Empfehlungen hinsichtlich körperlicher Aktivität zu überprüfen, die Ursachen des Aktivitätsverhaltens zu untersuchen, für einen Sport spezifische Trainingsprogramme zu implementieren und die körperlichen Anforderungen, die in einem sportlichen Wettkampf vorkommen, nachzubilden.

Derzeitig zur Verfügung stehende Systeme zur Aktivitätserkennung lassen sich in subjektive und objektive Techniken einteilen. Subjektive Techniken beruhen oftmals auf Fragebögen. Fragebögen zu benutzen beinhaltet das Problem, dass man sich nicht zuverlässig an die Aktivität, die durchgeführt wurde, erinnert.

Im Gegensatz dazu stellen tragbare Computersysteme wie Aktivitätstracker eine objektive Lösung zur Aktivitätserkennung dar. Kleine, leichte und energiesparende Sensoren, z.B. Inertialsensoren, erfassen dabei die Bewegung des Menschen. Die Sensordaten, die aufgenommen werden, werden mithilfe von Algorithmen des Maschinellen Lernens weiterverarbeitet, um eine automatische Erkennung menschlicher Aktivität zu ermöglichen.

In dieser Doktorarbeit werden Algorithmen zur sensorbasierten Aktivitätserkennung vorgestellt und evaluiert. Die automatische Erkennung von Alltags- und Sportaktivitäten bildet dabei den Fokus. Beispiele für Alltagsaktivitäten werden aus dem 'Compendium of Physical Activities' entnommen und beinhalten Sitzen, Geschirrspülen und Treppen steigen. Fußball wird aufgrund der hohen Anzahl aktiver, involvierter Personen als Sportart ausgewählt. Die fußballspezifischen Aktivitäten beinhalten hauptsächlich den Vollspann- und Innenseitstoß. Neben der Entwicklung und Implementierung von Algorithmen werden des Weiteren notwendige Erweiterungen hinsichtlich des Designs von Aktivitätserkennungssystemen identifiziert und zukünftige Forschungsschwerpunkte aufgezeigt. Im Folgenden werden die wissenschaftlichen Beiträge, die in dieser Arbeit vorgestellt werden, näher erläutert.

Verschiedene Algorithmen werden beschrieben, die in der Lage sind, die genannten Alltagsaktivitäten und fußballspezifischen Aktivitäten hinsichtlich ihres Typs und ihrer Intensität abzuleiten. Drei Aktivitätserkennungsalgorithmen klassifizieren Alltagsaktivitäten. Dabei wird

eine hierarchische Architektur oder eine Fusion auf Entscheidungsebene verwendet. Die Ansätze erzielten eine Genauigkeit von über 85 % und reduzieren bei einer Umgebungsänderung, z.B. neue Aktivitäten oder Sensoren, den Aufwand eines erneuten Trainings. Im Vergleich zu aktuellen Forschungsarbeiten stellen diese Algorithmen einen höheren Anteil an generischen Komponenten bereit, die automatisch an verschiedene Anwendungsbedingungen angepasst werden können. Das Konzept der Fusion auf Entscheidungsebene wird weiterhin verwendet, um den Energieverbrauch beim Laufen auf einem Laufband zu schätzen. Das entsprechende System erzielte einen Root Mean Square Error von 0.42 MET.

Zwei Aktivitätserkennungsalgorithmen werden vorgestellt, die einen Vollspann- und einen Innenseitstoß im Fußball klassifizieren unter der Verwendung einer hierarchischen Architektur oder unter Ausnutzung von Expertenwissen hinsichtlich Schussphasen. Die vorgeschlagenen Ansätze erzielten eine Genauigkeit von über 85 %. Ein Aktivitätserkennungsalgorithmus wird weiterhin beschrieben, der die Ballgeschwindigkeit eines Vollspannstoßes schätzt. Das System erzielte einen Mean Absolute Error (MAE) von 5.5 km/h. Die fußballspezifischen Algorithmen stellen, verglichen mit aktuellen Ansätzen aus der Literatur, eine bessere Anwendbarkeit der Schussklassifikation und der Ballgeschwindigkeitsschätzung unter realistischeren Bedingungen bereit, wie etwa einem Fußballspiel.

Neben Aktivitätsinferenzalgorithmen werden zwei neuartige Techniken in der Aktivitätserkennungsforschung vorgestellt und ausgewertet. Zum einen wird eine kostengünstige Strategie zur Datenbankfusion beschrieben und untersucht, die darauf abzielt, die Anzahl an Instanzen zu erhöhen, die für das Trainieren und Testen von Aktivitätserkennungssystemen verwendet werden kann. Die Technik benötigt keine zeitaufwändige Aufnahme zusätzlicher Daten. Zum anderen wird ein allgemeines Auswertungsframework vorgestellt, das verwendet werden kann, um verschiedene Aktivitätsinferenzmethoden fair zu bewerten. Das Framework wird angewandt, um die vorgeschlagenen Algorithmen zur Aktivitätserkennung mit Standard-Algorithmen aus der Literatur zu vergleichen, basierend auf dem gleichen Benchmarkdatensatz und Experimentsetup. Das Framework erlaubt eine bessere Identifikation des besten Algorithmus für eine bestimmte Anwendung, verglichen mit aktuellen Ansätzen aus der Literatur.

Die Algorithmen zur Aktivitätserkennung im Fußball werden zudem in eine neuartige Anwendung zur sensorgetriebenen Erstellung von individuellen Videozusammenfassungen integriert. Die Auswahl der zugehörigen Highlightszenen hängt von der Entscheidung sensorbasierter Aktivitätserkennungsalgorithmen ab. Den Highlightvideos werden weiterhin Video- und Audioeffekte an spezifischen Schussphasen hinzugefügt, die automatisch detektiert werden.

Die bisher vorgestellten Algorithmen werden auf kleinen Datenmengen getestet. Geräte, die mit dem Internet verbunden sind, erzeugen eine große Menge an Daten, was mit dem Trend hinsichtlich der Generierung von Big Data einhergeht. Obwohl Big Data neue Möglichkeiten für die Aktivitätserkennungsforschung bietet, z.B. die automatische Erkennung von abnormalem Verhalten, berücksichtigt das momentane Design von Aktivitätserkennungssystemen keine Big Data Konzepte und Werkzeuge. In dieser Doktorarbeit wird Big Data in das Forschungsfeld der Aktivitätserkennung eingegliedert. Eine Definition von Big Data, die sich auf die Aktivitätserkennung bezieht, und Ergänzungen des Designs von Aktivitätserkennungssystemen, die spezifisch Big Data adressieren, werden vorgeschlagen. Des Weiteren werden offene Punkte und zukünftige Forschungsschwerpunkte dargelegt.

Die wissenschaftlichen Beiträge, die in dieser Doktorarbeit präsentiert werden, adressieren aktuelle Einschränkungen im Forschungsfeld der sensorbasierten Aktivitätserkennung und ermöglichen eine bessere Bewertung von menschlichem Verhalten im Gesundheitsumfeld und im Sport. Zukünftige Multisensor-Aktivitätserkennungssysteme, die die vorgeschlagenen Algorithmen integrieren und zusätzlich Big Data spezifische Komponenten berücksichtigen, ermöglichen eine robuste, ganzheitliche und langfristige Analyse des körperlichen Zustandes des Menschen.

Acknowledgment

First of all, I would like to thank Björn Eskofier for the great support during the last years. He gave very good feedback and hints regarding paper writing, presentation skills, and personal behavior.

I would like to acknowledge all colleagues from the Machine Learning and Data Analytics Lab, former Digital Sports Group of the Pattern Recognition Lab. I will miss the scientific discussions, private conversations, and jokes (in the lunch break, of course).

I would also like to take this opportunity to thank all the students who did their thesis or student assistant work under my supervision.

I would like to acknowledge all colleagues from adidas, who accompanied me. I really enjoyed the time with you. You offered great projects and very interesting topics. The access to your research lab, the player pool, and your expert knowledge were important puzzle pieces in my thesis. The atmosphere in your group was further amazing. I enjoyed every minute on the adidas campus in Herzogenaurach and I was proud to wear apparel which had the three stripes on it.

I would like to thank Kristof van Laerhoven to assess this thesis as the external reviewer. It was also an honor to publish a paper with you. You provided me with excellent feedback.

Lastly, I would like to thank my family. They heavily supported me during my whole university time. From childhood on, my parents further taught me what is important in life and they were great role models.

The work presented in this thesis was supported by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology and the European Fund for Regional Development.

Dominik Schuldhaus

Contents

Chapter 1

Introduction

HAR deals with the automatic recognition of physical activities and plays a major role in the health and sports sector [Bull 14]. In this thesis, algorithms were developed and implemented, which were able to infer daily and sports activities from IMU data using machine learning. In addition, concepts and directions for future HAR systems were given.

This chapter is structured as follows. First, a motivation is given why knowledge about performed activities is important in the health and sports domain (section 1.1). Second, a brief introduction of methods is provided how these activities can be monitored (section 1.2). Third, the scope of this thesis in the field of HAR is described (section 1.3). Fourth, technical challenges of IMU-based HAR systems are comprised (section 1.4). Fifth, a literature review about HAR is given (section 1.5). Sixth, HAR related products are introduced (section 1.6). Seventh, the main contributions regarding HAR in daily life and sports are given (section 1.7). Eighth, the outline of the thesis is provided (section 1.8).

1.1 Motivation

Physical inactivity is the fourth leading risk factor for global mortality [Worl 09] and about 3.2 million deaths per year are because of physical inactivity [Worl 11]. Researchers showed that an inverse relationship exists between being physically active and suffering from cardiovascular diseases [Bare 04, Schn 06], cancers [Teha 06, Meye 06], pulmonary diseases [Garc 06], and type 2 diabetes [Hu 01].

In this context, physical activity monitors were recommended [Hill 14] in order to provide

- Information about meeting physical activity recommendations [Warb 06, Kohl 12]
- Investigate the causes of physical activity behavior [Baum 12, Stra 13]
- Assess the effect of intervention programmes [Ogil 07, Hall 12].

Knowledge about performed activities is further beneficial in sports. Understanding the activity patterns of sports types enables coaches to prescribe and implement sport-specific training programs that replicate the physical demands during competition [High 12, Lyle 02] as well as to develop recovery strategies [Duff 10]. Quantities such as heart rate or distances are evaluated regarding activities like jogging, sprinting, and jumping [McIn 95, Barb 08, MacL 09]. In team sports, the identification of talents at an early age is necessary so that clubs remain competitive [Will 00]. Scouts implement test batteries which determine the performance during various activities such as shooting, dribbling, and sprinting [Reil 00, Lido 05]. Information e.g. about the time spent in certain activities such as dribbling gives an indication about sport-specific skills.

1.2 Physical Activity Monitoring

According to [Casp 85, Shep 99, Warr 10], monitoring humans' physical activities should ideally consider four dimensions:

1. **Frequency:** the number of activity instances (e.g. number of steps, number of shots on goals)
2. **Duration:** time of participation in an activity (e.g. time spent in low, medium, and high intensity level during the day or in a game)
3. **Intensity:** physiological effort associated with participating in an activity (e.g. effort associated in household activities or in game actions in sports)
4. **Type:** kind of activity (e.g. climbing stairs, walking, kicking a ball, tackling an opponent).

Currently available human activity monitoring tools in the health and sports domains address these four dimensions. The tools can be divided into subjective and objective techniques [Lamo 01, Vala 06, Fitz 15].

Subjective techniques often rely on questionnaires [Crai 03, Bull 09, Kohl 12] and previous experience, e.g. of coaches [Hugh 15]. Drawbacks include problems in the reliability and validity when recalling activities [Fran 91, Warr 10]. Objective techniques often rely on direct observation [McKe 00, Hugh 04, McKe 06] or sensor-based systems, such as camera [Di S 06, Popp 10, Agga 11, Ke 13] and positioning systems [Fren 10, Grü 11, Kaut 15a, Seid 16]. Drawbacks include among others the burden of many required operators and high costs [Hill 14, Cast 14, Fitz 15]. Furthermore, equipment setup and application may be challenging and is restricted by the capture scenarios [Carl 05, Carl 08, Barr 08].

Nowadays, Body Sensor Networks are more and more used for objective activity monitoring and assessment [Avci 10, Chan 12, Mova 14]. Body Sensor Networks consist of small, low-power, miniaturized, lightweight, and intelligent sensor nodes. Examples of commonly used sensors include temperature, humidity, blood sugar, blood pressure, carbon dioxide, respiration, pulse oximetry, electrocardiography, electromyography, electroencephalography, and inertial sensors [Cao 09]. These sensors are often part of today's wearable computing systems including smart accessories (earwear, wristwear, eyewear, shoewear, and beltwear), smart clothing, and smart on-skin patches [Amft 17]. In particular, IMUs in combination with machine learning techniques are heavily applied to determine the four previously described dimensions frequency, duration, intensity, and type of activity [Crou 06a, Plas 07, Chen 12b, Schu 13b, Schu 16, Pric 16].

1.3 Scope of This Thesis

This thesis deals with machine learning techniques based on IMU data, which address the automatic recognition of the activity type, commonly referred to as HAR [Lara 13, Bull 14]. The scope of the thesis is twofold. First, IMU-based machine learning algorithms are developed, implemented, and evaluated on two different sets of activities:

- **Daily activities** taken from the 'Compendium of Physical Activities' [Ains 93, Ains 00, Ains 11] such as sitting, washing dishes, climbing stairs, walking, and bicycling. The proposed approaches classified the aforementioned activities and were compared to state-of-the-art techniques based on the same benchmark dataset. In particular, dynamic physical activities, e.g. walking can further be classified into intensity levels, such as light, moderate, and vigorous [Bono 09a, Motl 09]. The intensity level is usually defined by the expended energy. Thus, an algorithm for IMU-based energy expenditure estimation was further developed. The aforementioned algorithms can be used to monitor compliance regarding physical activity recommendations as well as effects of intervention programs.

- **Sport-specific activities**. The focus in this thesis is on soccer activities. HAR is especially needed in soccer, the world's number one sport [Reil 03, Kunz 07] due to 270 million actively involved people (4 % of the world's population) [Kunz 07]. Algorithms were developed, which recognized the two main soccer-specific activities, namely full-instep and side-foot kicks [Leva 98]. Since the selected daily and soccer-specific activities varied in their characteristics, the algorithmic components for the recognition of both activity types differed. Full-instep kicks can further be classified due to the achieved ball speed which is the main indicator of kicking success in soccer [Lees 98, Dör 02, Kell 07]. Thus, an algorithm for IMU-based ball speed estimation was further implemented and evaluated. The aforementioned algorithms can be used to provide coaches and scouts with important feedback about the physical state of soccer players.

The second scope of this thesis is to provide concepts and directions for future HAR systems in the context of Big Data [Ragh 14, Chen 14a, Sing 14]. The availability of Big Data for HAR would enable important findings in different domains, e.g. health-enabling technologies for the elderly. Monitoring over a long period of time supports e.g. the automatic detection of cardiac emergencies, automated detection of deviant behavior, and the recognition of unknown diseases [Ludw 12].

1.4 Technical Challenges

Various technical challenges have to be considered if machine learning techniques are applied to IMU data for the purpose of HAR.

The user of interest, e.g. the athlete, should not be hindered by the sensor system. An unobtrusive **sensor location** has to be considered [Gemp 98, Yang 10]. Nevertheless, the sensor data of the chosen sensor location should provide suitable information to assess the performed activities. In certain sports types, collisions occur and might damage the sensors. Thus, the sensor should be located such that the probability of damage is minimal. In order to reduce the computational complexity and the energy consumption, the **number of sensors** should further be minimized [Lara 13]. Apart from the raw sensor data, machine learning techniques usually require **ground truth annotation** for determining approach-specific parameters in the training phase [Duda 00].

Walking, running, and cycling are rather periodic activities, whereas kicking a ball or performing a tackle are rather sporadic activities. Machine learning techniques should provide solutions for various types of activities. The activities differ in **occurrence characteristics** [Bull 14]. Humans often perform activities in different ways influenced by factors such as stress or fatigue [Bull 14]. Machine learning techniques should be robust regarding **intraclass variability**. The ratio of target activities that should be monitored and assessed to activities that should be neglected, often denoted as **NULL class**, is usually small [Bull 14]. Machine learning techniques should filter out undesired activities of the NULL class. Depending on the application, **expert knowledge** about the problem domain is often required or beneficial, e.g. in the feature extraction step [Bull 14, Jone 14]. Factors that influence the performance of developed machine learning techniques include the considered activity set, applied feature extraction techniques, and applied learning algorithms [Lara 13]. Challenges include the **comparison** of various proposed algorithms and defining common **metrics** for performance assessment.

The described technical challenges were addressed in this thesis and possible solutions were provided for the recognition of daily as well as soccer-specific activities.

1.5 Literature Review

In this section, a literature review was conducted and grouped into four parts. First, algorithms for the recognition of common daily activities are described (section 1.5.1). The information was used to gather knowledge about the typical design of HAR systems and to identify open issues. Second, research about energy expenditure estimation is presented (section 1.5.2). The findings were used to develop a method, which can further classify daily activities according to the intensity level. Third, algorithms for the recognition of soccer-specific activities are described (section 1.5.3). The information was used to design the kick classification techniques. Fourth, research about ball speed estimation is presented (section 1.5.4). The findings were used to develop a method, which can further classify a full-instep kick according to the intensity. Big Data in HAR is a rather new research topic and corresponding research work is limited. Therefore, a literature review about Big Data in HAR could not be provided in this section. The introduction of the abstract concept of Big Data for HAR was based on the literature review comprised in section 1.5.1.

1.5.1 Recognition of Daily Activities

Research work about the recognition of daily activities was categorized regarding data, algorithm, and application perspective. The data perspective describes which activities and which sensors are usually considered. The algorithm perspective introduces algorithmic components and concepts of HAR systems such as Activity Recognition Chain (ARC), fusion strategies, hierarchical classification, user-specific methods, and learning techniques. The application perspective mentions various examples of deployed HAR systems. The section concludes with general findings, limitations, and open issues. The following survey was part of a journal article which was submitted [Schu 18].

Data Perspective

The **acquired activities** were mainly listed in the 'Compendium of Physical Activities' [Ains 93], which included 19 activity groups with activities of daily living or self care, leisure and recreation, occupation, and rest. The groups explained the reason a human is engaging in a certain activity. The energy cost was further determined for each activity by available published and unpublished data. The goal was to enhance the comparability of results across studies which used self-reports for physical activity assessment. The compendium was updated in 2000 [Ains 00] and 2011 [Ains 11]. The different intensity levels included sedentary behavior (1.0-1.5 METs), light-intensity (1.6-2.9 METs), moderate-intensity (3-5.9 METs), and vigorous-intensity ($\geq$6 METs). Current HAR algorithms mainly classified single activities performed by one human. Nevertheless, limited research also exists dealing with other types of activities such as concurrent, interleaved, and composite activities performed by one human [Wu 07b, Hu 08, Blan 10], multi-user group activities [Gord 11], and abnormal activities [Hu 09].

Various **sensor sources** were used to infer daily activities including environmental sensors, GPS sensors, and biosensors [Chen 12b, Lara 13]. Publicly available databases, which include sensor data of various daily activities are comprised in Table 1.1. Accelerometers were mainly used for HAR due to their low cost, small size, and light weight [Barb 01]. Accelerometers were further preferred for HAR since acceleration is proportional to external force and therefore can reflect intensity as well as frequency of human movement [Schu 01, Math 04b, Godf 08, Yang 10]. The applicability of accelerometers to HAR was investigated in [Velt 93]. Due to their findings, the discrimination of the three activities standing, sitting, and lying was possible with one uniaxial sensor on the trunk and one on the upper leg.

In [Clel 13], the optimal sensor position for activity recognition was investigated. The best performance was achieved by an accelerometer placed at the hip. This coincides with the findings in [Bout 97a]. According to [Bout 97a], a sensor placement close to the center of mass of the body is preferred, if the movement of the whole body should be analyzed. Nevertheless, the accelerometer is usually placed on the body part whose movement is analyzed [Math 04b]. Accelerometers were attached to ankle and shin, if leg movement during walking should be investigated [Godf 08]. In [Bhat 80, Capp 82, Bout 97a, Khan 16], the optimal sampling rates and sensor ranges for activity recognition were investigated. According to [Bhat 80, Capp 82], frequencies and amplitudes often increase from cranial to caudal body parts and reach maximum values e.g. during running [Bout 97a]. According to [Bout 97a], accelerometer sensors should be able to capture an amplitude range of $\pm$ 12 g and frequencies up to 20 Hz. For accelerometers placed on waist level, an amplitude range of $\pm$ 6 g was suggested. In [Khan 16], the optimal sampling rate was estimated based on the original sampling rate. A two-sample Kolmogorov-Smirnov test was proposed to compute the similarity of a downsampled dataset and the original dataset. For an example dataset, an optimal sampling rate of 30 Hz instead of originally 100 Hz was proposed.

Accelerometers were often used in combination with gyroscopes, and magnetometers to recognize daily activities [Schw 12, Ordo 16, Ahn 16].

Table 1.1.: List of publicly available datasets for HAR (adapted from [Schu 18]). Table contains name of dataset, number of subjects (# Sub), number of activities (# Act), provided sensor types (ACC: accelerometer, IMU: inertial measurement unit, IMMU: inertial-magnetic measurement unit, HR: heart rate), and reference (Ref). The study that was conducted for this thesis is marked in bold.

Name	# Sub	# Act	Sensors	Ref
ALKAN	> 200	NA	ACC	[Hatt 10]
HAR Using Smartph.	30	6	IMU	[Angu 13]
UniMiB SHAR	30	9	ACC	[Micu 17]
WISDM	29	6	ACC	[Kwap 10]
DaLiAc	**19**	**13**	**4 x IMU**	**[Schu 13b]**
USC-HAD	14	12	1 x IMU	[Zhan 12]
REALDISP	17	33	9 x IMMU	[Bano 14]
Berkeley MHAD	12	11	ACC	[Ofli 13]
PAMAP2	9	12	3 x IMMU, HR	[Reis 12]
OPPORTUNITY	4	9	e.g. 19 x IMU/ACC	[Chav 13]
PLCouple1	2	43	> e.g. 2 x ACC	[Loga 07]
PLIA1	1	89	e.g. 2 x ACC	[Inti 06]

Nevertheless, using gyroscopes is rather new compared to accelerometers [Amin 04, Dobk 11]. Advantages are a lower noise level compared to accelerometers and the possibility to estimate rotation angles of human movement by integration [Amin 04].

In this work, thirteen single activities were taken from the 'Compendium of Physical Activities' from the following groups: bicycling, conditioning exercises, home activities, inactivity, running, and walking. The details are given later in this thesis. The conducted studies used inertial sensors consisting of accelerometer and gyroscope. The sensors were attached to different body positions in order to capture various daily activities, which were performed with different body parts. The set of sensor positions included e.g. the hip as proposed in literature [Bout 97a].

Algorithm Perspective

The core of all algorithms for HAR is the **Activity Recognition Chain** (ARC), which is a generic description of the design and performance assessment of HAR systems [Bull 14]. Current state-of-the-art approaches for HAR implemented such an ARC and are comprised in Table 1.2. The ARC consists of a sequence of machine learning techniques such as preprocessing, segmentation, feature extraction, and classification (Figure 1.1).

Preprocessing steps aim at reducing noise in the signal and preparing the signals for feature extraction [Bull 14]. The authors in [Kara 06] applied median filtering to remove abnormal noise peaks produced by the accelerometer sensor. In [Alle 06], acceleration data was divided into body acceleration and gravity component. An elliptic low-pass filter with a cut-off frequency at 0.25 Hz was applied to determine the gravity component. The body acceleration was computed by subtracting the gravity component from the original acceleration signal. In [Mant 01], Principle Component Analysis was applied to the accelerometer data followed by Wavelet transformation.

The segmentation step of the ARC aims at identifying segments of the pre-processed sensor data that include information about the desired activities. Examples for window lengths included 6.7 s [Bao 04], 5.12 s [Ravi 05], 5 s [Altu 10], and 30 s [Liu 12]. A window overlap of 50 % was often preferred [Bao 04, Pree 09].

Table 1.2.: State-of-the-art algorithms for HAR, which are based on the ARC (adapted from [Schu 18]). Table contains used sensor types (ACC: accelerometer, IMU: inertial measurement unit, IMMU: inertial-magnetic measurement unit, IR: infrared, GPS: Global Positioning Systems, MAG: magnetometer), number of acquired subjects (# Sub), number of classified activities (# Act), applied algorithm (GMM: Gaussian Mixture Model, NCC: Nearest Class Center, DT: Decision Trees, PCA: Principle Component Analysis, MLP: Multilayer Perceptron, SVM: Support Vector Machine, kNN: k-Nearest Neighbors, CNN: Convolutional Neural Network, DBN: Deep Belief Networks, HMM: Hidden Markov Model, EKF: Extended Kalman Filter, DCM: Direction Cosine Matrix, LSTM: Long-Short-Term Memory, CRF: Conditional Random Fields, NB: Naive Bayes, QDA: Quadratic Discriminant Analysis, HSMM: Hidden Semi-Markov Model, DLF: Decision Level Fusion, RF: Random Forest), accuracy (Acc), and reference (Ref). The algorithms developed and implemented in this thesis are marked in bold.

Sensors	#Sub	#Act	Algorithm	Acc[%]	Ref
6 x ACC	5	NA	Threshold	NA	[Velt 93]
3 x ACC	10	7	Threshold	NA	[Velt 96]
4 x ACC	26	8	Threshold	98.1	[Fahr 97]
4 x ACC	24	9	Threshold	66.7	[Foer 99]
2 x ACC	5	4	Threshold	87.0	[Amin 99]
7 x ACC	1	11	EKF, HMM	96.5	[Wu 07a]
2 x IMMU	10	4	DCM, NB	95.0	[Ahn 16]
2 x ACC	6	4	PCA + MLP	84.8	[Mant 01]
22 signals	16	7	DT	86.0	[Park 06]
19 x ACC	8	10	DLF	98.0	[Zapp 07]
e.g. 2 x ACC	50	13	SVM	88.1	[Liu 12]
6 x IMMU	9	10	Manifolds	89.0	[Schw 12]
4 x **IMU**	**19**	**7**	**DLF**	**93.9**	**[Schu 13a]**
5 x ACC	20	9	DLF	95.0	[Bano 13]
5 x ACC	20	20	C4.5	84.3	[Bao 04]

Table 1.2.: State-of-the-art algorithms for HAR (continued).

Sensors	#Sub	#Act	Algorithm	Acc[%]	Ref
9 x ACC	10	8	kNN	97.6	[Full 17]
4 x IMU	**42**	**8**	**DLF**	**85.8**	**[Schu 14b]**
3 x ACC	9	8	DLF	NA	[Chow 18]
6 x ACC	8	7	SVM	97.8	[Clel 13]
ACC	26	9	DT	97.7	[Math 04b]
4 x IMU	**19**	**13**	**Generic**	**89.6**	**[Schu 13b]**
1 x ACC	14	10	SVM, NB	95.6	[Zhen 15]
2 x ACC	NA	7	Probabilistic	75.8	[Van 00]
ACC	6	8	GMMs	92.2	[Alle 06]
10 x ACC	1	6	NCC	74.4	[Rogg 13]
ACC	2	8	Boosted SVM	73.3	[Ravi 05]
e.g. 3 x ACC	1	9	Joint boosting	64.2	[Stik 08]
ACC, GPS	32	3	C4.5	91.5	[Long 10]
ACC	1	3	Mixt.-of-exp.	91.8	[Lee 14]
5 x IMMU	6	6	Joint boosting	NA	[Blan 10]
8 x ACC	1	4	NCC	72.7	[Cala 11]
ACC	NA	3	Probabilistic	91.0	[Hach 12]
ACC	6	12	Rule-based	90.8	[Kara 06]
3 x ACC	20	8	k-NN	96.0	[Pree 09]
ACC	20	20	DT	93.0	[Bono 09a]
5 x IMMU	8	19	SVM	87.6	[Altu 10]
e.g. ACC	8	4	NB	97.4	[Sapo 08]

Table 1.2.: State-of-the-art algorithms for HAR (continued).

Sensors	**#Sub**	**#Act**	**Algorithm**	**Acc[%]**	**Ref**
ACC	8	4	DT	78.9	[Milu 08]
ACC	4	6	DT	90.6	[Yang 09]
2 x ACC	20	10	Fuzzy classif.	86.0	[Berc 10]
ACC, GPS	16	5	DT + HMM	94.0	[Redd 10]
ACC	9	4	SVM	94.2	[Park 12]
ACC	7	5	QDA	95.8	[Siir 12]
IMMU	20	5	kNN	96.8	[Uste 13]
ACC	29	6	MLP	91.7	[Kwap 10]
e.g. IR	2	12	HSMM	71.6	[Kast 10]
e.g. ACC/MAG	13	7	SVM	86.2	[Fleu 10]
e.g. motion	1	15	SVM	93.4	[Fati 13]
e.g. IR	10	8	RF	68.5	[Nef 15]
ACC	7	4	CNN	93.4	[Zhen 14]
ACC	29	6	DBN, HMM	98.2	[Alsh 15]
19 x IMMU/ACC	4	18	CNN, LSTM	91.8	[Ordo 16]
IMU	30	6	Autoencoder	97.5	[Alma 17]

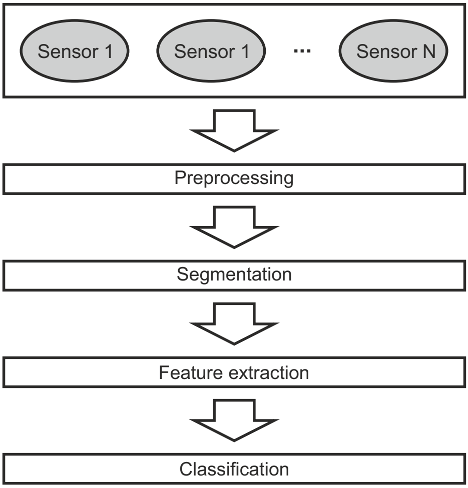


Figure 1.1.: Activity recognition chain (adapted from [Bull 14]). N sensors are used as input for the preprocessing step followed by segmentation, feature extraction, and classification.

Feature extraction is applied to reduce the dimension of the input data for classifiers [Bull 14]. Features were computed which were discriminative for the desired activities and can be extracted in the time and frequency domain [Lara 13]. State-of-the-art approaches in the literature extracted features such as mean, standard deviation, percentiles, energy, Signal Magnitude Area (SMA), entropy, correlation between axes, peak frequency, and Fast Fourier Transform coefficients [Bao 04, Park 06, Kara 06, Pree 09, Liu 12]. Depending on the number of extracted features, feature selection routines are applied in order to reduce the amount of features. Feature selection routines included e.g. a visual analysis [Park 06, Zhen 15], Principle Component Analysis [Altu 10], and correlation-based feature selection [Chow 18].

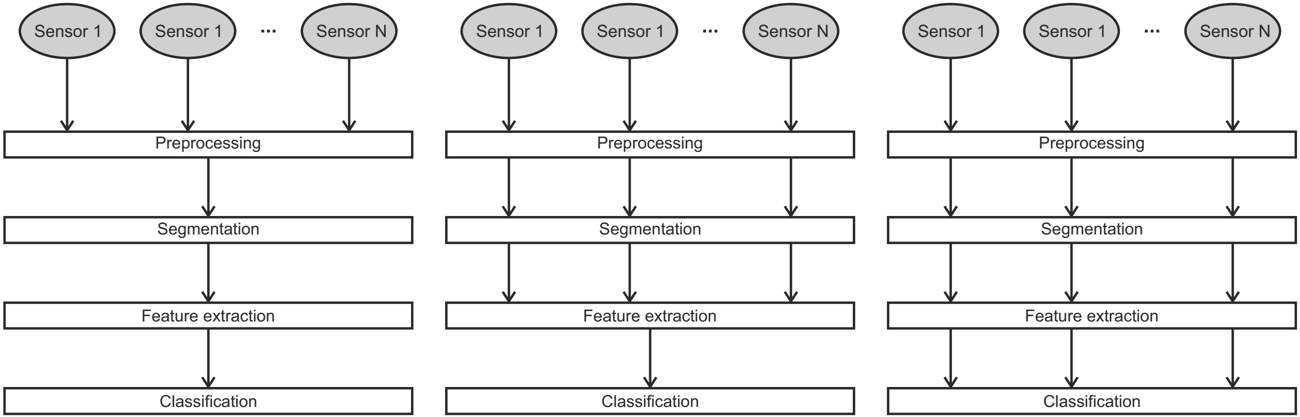
Various classifiers were used for HAR, e.g threshold-based approaches [Foer 99, Amin 99], Hidden Markov Models [Redd 10], Hidden Semi-Markov Models [Kast 10], Naive Bayes (NB) [Sapo 08, Ahn 16], Multilayer Perceptron [Mant 01], decision trees [Bao 04, Math 04b, Park 06], Support Vector Machine (SVM) [Liu 12, Zhen 15], k - Nearest - Neighbor (k-NN) [Pree 09, Uste 13, Full 17], and Random Forest (RF) [Nef 15].

In order to combine the information of different data sources, various **fusion strategies** were applied. Fusing multiple sensor observations can be usually applied at data level, feature level, and decision level [Hall 97]. Figure 1.2 illustrates the three fusion types considering the ARC shown in Figure 1.1.

Data level fusion was applied to combine the sensor information in the preprocessing step of an ARC [Kara 06, Wu 07a, Ahn 16]. The Signal Magnitude Vector (SMV) was computed in [Kara 06], which combined all three accelerometer axes to one signal. The SMV was used to eliminate information about sensor orientation, which was not needed to detect falls. Falls were assumed to occur, if at least two subsequent peaks in the SMV vector above a threshold were present. Accelerometer axes were combined in [Wu 07a] by an Extended Kalman Filter in order to estimate the flexion angles of body segments and to compute the angular velocity. In [Ahn 16], a Direction Cosine Matrix was applied to accelerometer, gyroscope, and magnetometer data in order to remove noise and estimate the correct sensor position.

Feature level fusion was applied to combine information of sensor sources at feature level. In [Ravi 05, Kwap 10, Uste 13], features such as mean and standard deviation were extracted from each of the three accelerometer axes and were comprised in one feature vector. The authors in [Park 06, Bao 04, Full 17] extracted features from accelerometer sensors, which were attached to various sensor positions: chest and wrist [Park 06]; hip, wrist, arm, ankle, and thigh [Bao 04]; both ankles, hips, wrists, and arms [Full 17]. It was shown that combining multiple sensors at feature level increased the overall classification accuracy [Bao 04]. In [Altu 10, Liu 12], features were extracted and fused based on multiple sensor types including accelerometer, gyroscope, magnetometer, and ventilation sensor. The authors in [Liu 12] computed 33 features from the accelerometer and ventilation sensor. An accuracy of 88.1 % was achieved, which was 12.3 % higher than using a single hip accelerometer. A further reduction of the subject-to-subject variability was observed by using the ventilation sensor. One main advantage of feature level fusion was that only one classifier had to be trained [Bao 04, Park 06, Full 17].

In contrast to feature level fusion, decision level fusion was performed in [Zapp 07, Bano 13, Chow 18]. The authors in [Zapp 07] compared majority voting and a naive Bayesian fusion of classifier decisions.



(a) Data level fusion. Sensor information is fused in the preprocessing step.

(b) Feature level fusion. Preprocessing and segmentation are performed separately for each sensor source. Sensor information is fused at feature level.

(c) Decision level fusion. Preprocessing, segmentation, feature extraction, and classification are performed separately for each sensor source. Sensor information is fused at decision level.

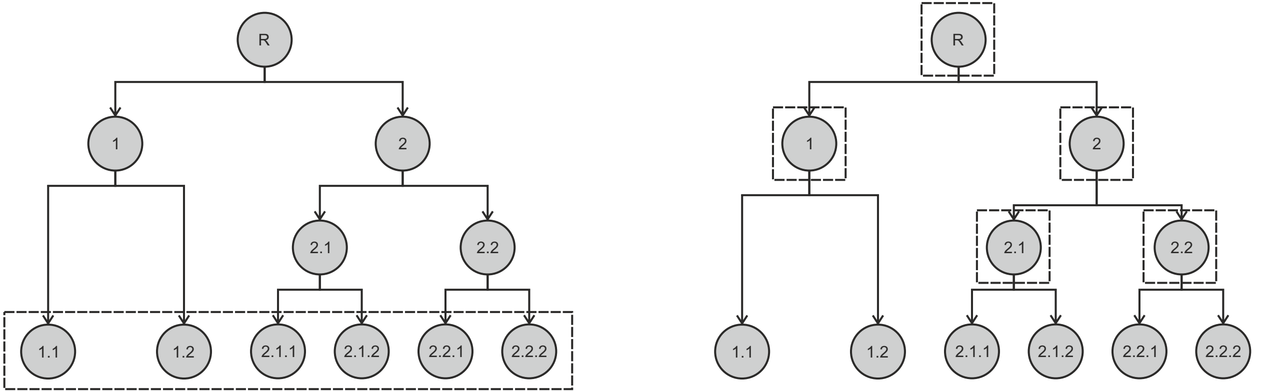
Figure 1.2.: Fusion strategies of HAR systems (adapted from [Hall 97]). (a), (b), and (c) show examples for data level, feature level, and decision level fusion, respectively.

The naive Bayesian fusion outperformed majority voting and achieved an accuracy of 98 %. In [Bano 13], a hierarchical-weighted decision level fusion was proposed. Binary one-vs-rest classifiers were trained in the first level and the classifier decisions were weighted according to the achieved accuracy. The system achieved an accuracy of above 95 % using SVM.

In [Chow 18], three accelerometer sensors were used to distinguish between eight activities. Three decision level fusion approaches were compared including model-based, class-based, and posterior-adapted class-based fusion. The latter fusion technique achieved the best F1-score of 92.3 % using all three sensors. According to [Zapp 07, Bano 13], decision level fusion reduced the complexity of lower levels of the ARC, increased the robustness as well as the adaptability of the recognition system, and provided an efficient scaling to a huge amount of sensors. Nevertheless, decision level fusion often required a large amount of sensors [Zapp 07] and a training of multiple classifiers [Bano 13].

According to [Sill 11], many real-world classification problems can be seen as **hierarchical classification problems**, in which the predicted classes are organized in a pre-defined class hierarchy, e.g. a tree. An example of a class hierarchy is given in Figure 1.3. The recognition of daily activities were also often treated as a hierarchical classification problem [Velt 96, Fahr 97, Amin 99, Math 04a, Zhen 15]. The authors in [Velt 96, Fahr 97, Amin 99, Math 04a] first defined activities as either static or dynamic. Further levels often included a clustering of walking activities (flat, upstairs, downstairs) [Fahr 97] or a grouping of postures (sitting, standing). In [Zhen 15], several walking activities and static activities as well as single activities such as jumping and running were considered in the first level. The second level clustered the walking activities in 2-D movements (walking forward, walking left, walking right) and 3-D movements (walking upstairs, walking downstairs). In addition, the second level contained the single static activities. The last level contained all single walking activities.

Although activities can be organized into a class hierarchy, state-of-the-art algorithms for HAR mainly developed and applied flat classification systems [Liu 12, Bao 04, Pree 09]. The flat classification systems relied on a single classifier decision including all desired classes [Babb 13] and ignored the class hierarchy [Sill 11]. In Figure 1.3a, the flat classification system is put into the context of the class hierarchy.



(a) Flat classification system considering leaf nodes.

(b) Hierarchical classification system with one local classifier per parent node.

Figure 1.3.: Class hierarchy with flat (a) and hierarchical (b) classification systems. Circles and rectangles denote classes and classification systems, respectively (adapted from [Sill 11]). R denotes root node. Leaf nodes 1.1, 1.2, 2.1.1, 2.1.2, 2.2.1, and 2.2.2 represent final desired classes.

In contrast to flat classification systems, research about hierarchical classification systems for HAR is limited. Hierarchical classification systems provided an inherent hierarchical structure consisting of multiple levels, which merged certain classes in single hierarchy levels [Sill 11]. One local classifier was often developed per parent node. In Figure 1.3b, the hierarchical classification system is put into the context of the class hierarchy. Single local classifiers are indicated by the rectangles. In [Velt 96, Fahr 97, Amin 99, Math 04a, Zhen 15], hierarchical classification systems were applied to recognize various daily activities. Uniaxial accelerometers were used to classify specific static and dynamic activities [Velt 96, Fahr 97, Amin 99]. Hierarchical threshold-based and distance-based systems were implemented which first discriminated between merged static and dynamic activities followed by a further classification of single activities for each sub group. A generic, hierarchical framework for activity recognition was described in [Math 04a, Zhen 15], which was independent regarding the activity set. In [Math 04a], a binary decision tree was implemented for classification dividing the movements into classes and subclasses at different hierarchical levels. Thresholding approaches were mainly used, e.g. for the discrimination of activity and rest. If sensitivity and specificity measures above 95 % were not achieved, more sophisticated, pattern-based techniques were implemented, e.g. rule-based systems. In [Zhen 15], a hierarchical HAR system was evaluated on a dataset containing ten activities. One activity estimator was implemented in each level consisting of a SVM followed by a NB classifier. Six features were computed as input for the SVM based on a single accelerometer. The proposed system achieved an accuracy of 95.6 %.

The authors in [Van 00, Alle 06, Rogg 13] developed **user-specific** HAR systems [Bull 14], which were optimized for a specific human. In [Van 00], a Kohonen Self-Organizing Map clustering algorithm was applied to assign map-units to pre-computed feature vectors. For each new incoming feature vector, one unit of the already existing map was selected and adapted itself a bit more towards the new incoming feature vector. In [Alle 06], Gaussian Mixture Models were adapted to an intended subject using Bayesian adaptation. By using the adaptation method, an accuracy of 92.2 % was achieved compared to 76.6 % without using the adaptation. In [Rogg 13], an adaptive ARC was introduced including self-monitoring, adaptation strategies, and external feedback components. The accuracy of the implemented system was improved by 13.4 % in the case of on-body displacement compared to applying no adaptation. User-dependent systems often showed a higher performance than user-independent. Nevertheless, the generalization of user-dependent system to other users is worse.

Algorithms for HAR can further be distinguished by the applied **learning techniques**. State-of-the-art approaches in activity recognition mainly applied supervised learning [Chap 10]. A mapping from features to an activity class was learned given a set of training instances [Bao 04, Ravi 05, Pree 09]. According to [Stik 08], the main disadvantage of supervised learning is that a huge amount of labeled data is required in order to learn activity models.

In contrast, semi-supervised learning methods integrate unlabeled data in the learning process [Chap 10]. Techniques for semi-supervised methods include e.g. co-learning [Blum 98, Chap 10, Long 10]. In [Lee 14], a mixture-of-experts model was implemented in order to perform activity recognition on a mobile device. A global-local co-learning technique was applied with labeled and unlabeled data. Investigations showed a maximum performance improvement of 10 % compared to using no co-training. Semi-supervised learning techniques such as co-learning mainly use predictions as labels [Long 10].

In contrast, active learning approaches select samples of interest and the user is asked to label the chosen instances [Cohn 94]. In [Stik 08], two active learning strategies were tested on the PLCouple1 dataset (Table 1.1). The techniques were based on a pool-based setting, in which a small set of labeled instances and a large set of unlabeled instances were taken as input for several classifiers. The proposed learning system outperformed comparable supervised learning techniques.

Machine learning approaches for HAR mainly assumed that training and test data are drawn from the same distribution. According to [Pan 10], statistical models often require a retraining of the complete system, when the data distribution changes. New data have to be collected in order to update the proposed system. In order to reduce the need and effort for a new data collection session, transfer learning can be applied [Cook 13]. Examples for transfer learning applied to HAR can be found in [Blan 10, Cala 11, Hach 12]. In [Blan 10], a layered approach was implemented including both activity events and composite activities. Trained activity models based on one single dataset was reused to test the system based on a second dataset. The proposed approach reduced the requirement of retraining new composite activities. In [Cala 11], a system-supervised learning approach was implemented. The labels were provided by a ground-truth trained teacher sensor node and were used by a learner sensor node. The learner sensor node applied the labels to perform batch training. In [Hach 12], an importance-weighted least-squares probabilistic classifier was implemented considering a covariance shift, i.e. the distributions of training and test phases differed.

Most state-of-the-art algorithms implemented an ARC, which consisted of a rather shallow architecture [Beng 09]. Nevertheless, activity recognition systems can further be based on deep learning techniques [LeCu 98, Hint 06, Beng 09]. Deep learning architectures combine the preprocessing, feature extraction, and classification step of the ARC. A time-consuming handcrafted feature engineering is often not needed. In [Ordo 16, Alsh 15, Zhen 14, Alma 17], deep learning such as Convolutional Neural Networks, Long-Short-Term Memory units, Deep Belief Networks, and Autoencoder were applied to activity recognition and showed better performance than state-of-the-art shallow machine learning methods. The disadvantage of deep learning techniques is mainly the requirement of a large amount of training data [Beng 09].

The algorithms developed in this thesis were based on the described ARC (Figure 1.1). Three HAR algorithms were developed and implemented containing at least one of the three data fusion types. The SMV was computed in order to fuse sensitivity axes of inertial sensors, to remove the direction of movement, and to reduce the computational complexity. Feature level and decision level fusion were applied to combine the information from different sensor sources. Flat and hierarchical classification systems were developed and compared. In contrast to [Math 04a, Zhen 15], the proposed hierarchical HAR algorithm contained more generic components like an automatic selection of features and different sophisticated classifiers in each level of the hierarchy. The proposed approach was further evaluated on a dataset containing data of various daily activities, sensor positions, and sensor types.

Application Perspective

Algorithms for HAR were often part of various applications. In the **UbiFit Garden** project, levels of physical activity were acquired. The information was compared to pre-defined health goals [Cons 08]. An on-body worn fitness device recognized various types of activities such as walking, running, and cycling. The information was communicated via Bluetooth to a smartphone. An application running on the mobile device compared the performed activities to the goals. In [Milu 08], a novel people-centric sensing application was proposed (**CenceMe**). Smartphone data was used to automatically infer human's sensing presence. The information was shared through social network portals. Activities such as sitting, standing, walking, and running were determined by machine learning techniques based on the accelerometer data. The classification result was sent to a backend server for further processing. A backend classifier provided higher level forms of classification based on the single activities. The authors in [Sapo 08] developed the **iLearn** platform consisting of three modules. First, a smartphone tool was used to gather accelerometer data labeled by the user. Second, the acquired data was imported into a desktop tool for learning and testing a classifier. Third, the trained classifier was stored on the smartphone and could be used by iPhone applications. The platform was used in an interactive video game application.

Findings

Inertial sensors were heavily used in HAR. State-of-the-art algorithms for HAR implemented the ARC introduced in [Bull 14]. Various data sources were fused at data level, feature level, and decision level in order to recognize mainly single activities. The overall architectures varied from flat to hierarchical structures. User-specific systems often increased the system performance. Different learning strategies were applied to include unlabeled data, reduce the effort of a new data collection session, or reduce the effort for handcrafted feature engineering. HAR algorithms were further included in various deployed systems.

Limitations and Open Issues

Although current state-of-the-art approaches for the recognition of daily activities tackled various research questions, limitations and open issues still remain. Most currently available HAR algorithms provided a **flat classification** algorithm distinguishing various activities with one single system. Two limitations were identified. First, flat classification algorithms relied on a single decision including all final activity classes without exploring the information about parent-child class relationships. Thus, the single HAR system might be complex, since the different characteristics of the activities should be captured by one system. The complexity of the HAR system might further increase, if a large number of classes should be considered [Sill 11, Babb 13]. Second, the capability of flat classification systems to react to environmental changes was limited. Integrating new activities or sensors might result in a retraining of the whole system. In order to be flexible in adding new elements to the HAR system and to provide a final activity prediction based on multiple single decisions, hierarchical architectures are more suitable than flat structures [Math 04a]. Nevertheless, only limited research work about hierarchical classification systems was available. Four limitations of current hierarchical algorithms for HAR were identified. First, the single local classifiers per node were mainly threshold-based [Math 04a], which might not be sufficient for more complex activities. Second, the configuration of single components of the ARC was often performed manually, e.g. a visual feature inspection and selection [Zhen 15]. Adjusting an existing trained HAR system to a new environment would be cumbersome and time-consuming, since a manual re-design is required. Third, most of the provided systems used the same ARC design in each local classifier per parent node [Math 04a, Zhen 15], e.g. in terms of selected features or applied classification techniques. Different types of local classifiers might improve the performance of HAR systems. Fourth, current available hierarchical classification systems were mainly evaluated on static activities and cyclic, dynamic activities [Velt 96, Fahr 97, Amin 99, Math 04a, Zhen 15]. The applicability of hierarchical systems to more daily activities remains unknown.

An accelerometer at waist level was often preferred in current state-of-the-art approaches for HAR [Bout 97a]. Nevertheless, additional sensor positions were often needed for more complex activities. Multiple sensors were mainly **fused** at **feature level**. Three limitations were identified. First, changing the number of sensors results in a retraining of the whole system [Bano 13]. Second, the performance of activity recognition systems during run-time is affected by degradation, interconnection failures, and jitter in sensor placement as well as orientation of single sensors [Zapp 07]. Third, a high number of sensors often resulted in an increased computational complexity in certain steps in the machine learning before the actual classification [Altu 10, Bano 13]. Decision level fusion provides an alternative solution to combine information of different sensors, but the corresponding research work is rather limited [Zapp 07, Bano 13, Chow 18]. Algorithms were evaluated on car manufacturing [Zapp 07], a small set of partially merged activities [Bano 13], were restricted in the types of possible applicable classifiers [Chow 18], and were mainly based on accelerometer sensors. An open issue includes the evaluation of a larger set of daily activities, which might need the integration of the gyroscope and more classifier types.

Identifying the best algorithmic approach for a certain application is currently challenging and requires a comparison of various state-of-the-art algorithms. Two limitations of currently applied comparison approaches were identified. First, existing algorithms were mainly compared regarding the performance, which was achieved in research papers. Nevertheless, the achieved performance of the techniques were based on different datasets, which varied regarding the number of sensor axes, number of sensors as well as sensor placements, sampling rates, and number of subjects. Second, various different performance measures were often computed across research papers including F1-score, accuracy, or balanced accuracy [Soko 09]. Thus, the set of algorithms which can be compared was limited depending on the provided performance measure. An open issue is to develop a framework which provides a fair comparison of various state-of-the-art algorithms considering a common benchmark dataset and the same set of performance measures.

The previously described research work mainly used one or more **isolated small datasets** for evaluation, which often included a low number of subjects, number and type of activities, number of attached sensors, and short activity duration (Table 1.1). Two limitations were identified. First, the explanatory power of findings is reduced, when only a limited amount of data is used in the evaluation. Second, the performance of HAR systems dropped due to low amount of training instances [Wu 07b]. In order to provide machine learning algorithms with more training data, the collection of additional data would be required. Since gathering data is often cumbersome and time-consuming, an open issue is to develop an alternative procedure for increasing training data which can be performed without much effort.

Wearable sensors are one of the biggest drivers of the Internet of Things (IoT) [Swan 12, Agga 13, Chen 15]. The huge amount of data collected by Internet-connected devices coincide with the trend toward the generation of **Big Data** [Ragh 14, Chen 14a, Sing 14, Zhan 16]. Big Data solutions enable the analysis of a huge amount of raw structured as well as semi-structured and unstructured data from multiple sources, the analysis of most of the data rather than considering only a sampling of data, and exploratory analysis when the final measures on data are not predetermined [Ziko 11]. The availability of Big Data for HAR would enable important findings in different domains. Monitoring over a long period of time supports e.g. the automated detection of deviant behavior and the recognition of unknown diseases [Ludw 12]. Although Big Data could have a positive impact on HAR research, the current version of the ARC did not consider Big Data components (Figure 1.1). Thus, open issues are the integration of Big Data concepts and tools in the ARC as well as a general introduction of Big Data to the HAR community.

1.5.2 Energy Expenditure Estimation

Algorithms for energy expenditure estimation mainly applied regression analysis. Thus, research work was grouped according to the following regression-specific topics: using acceleration counts as predictors, regression models considering data fusion, applying activity-specific regression, and comparison of regression algorithms. The literature review concludes with general findings, limitations, and open issues. The papers are summarized in Table 1.3.

Acceleration Count

One commonly used approach for energy expenditure estimation was to apply regression techniques to acceleration counts [Bout 97a]. The acceleration count was defined as the sum of absolute values of a triaxial accelerometer for a given epoch, e.g. 30 s.

In [Bout 94, Bout 97a, Bout 97b, Leen 03, Chen 05, Motl 09, Hend 00], simple **linear regression** was applied based on the previously mentioned acceleration count. In [Bout 94], the study protocol included sedentary activities and walking at five different speed levels (3, 4, 5, 6, 7 km/h) on a motor-driven treadmill. Energy expenditure was estimated from a 30 s interval at the end of each activity. The proposed system achieved a correlation coefficient of 0.95. In [Bout 97b], the acceleration count was computed only based on vertical and antero-posterior direction. The proposed system achieved a correlation of 0.96. In [Bout 94, Bout 97b], the accelerometer was placed at the lower back. The authors in [Motl 09], performed regression analysis on uni-axial accelerometer data acquired during treadmill running with three different speed levels (3.2, 4.8, 6.4 km/h). A correlation of 0.89 was achieved. In [Leen 03], the ability of various physical activity monitors to estimate the expended energy during treadmill walking was investigated. The Tritrac-R3-D (Professional Products, A Division of Reining Int., Madison, WI) and a Computer Science & Applications (CSA) Inc. accelerometer (Model

7164, Shalimar, FL) achieved a RMSE of 0.46 and 0.53 MET, respectively. The authors in [Hend 00] examined the validity of accelerometry in assessing moderate intensity activities in the field. Relationships between acceleration counts and ground truth Metabolic Equivalent of Task (MET) determined by a portable metabolic system were stronger for walking than for other activities such as indoor or outdoor household tasks. It was concluded that the count vs MET relationship was dependent on the type of activity. Accelerometers seemed not to be able to capture the increased energy costs from certain activities.

In [Su 05], a triaxial accelerometer was attached to the lower back. The acceleration count was computed and used as input for **Support Vector Regression**. The proposed system achieved a mean square error of 0.37 W/kg compared to 0.39 W/kg using simple linear regression.

Data Fusion

In [Chen 97, Plas 05], Multiple Linear Regression (MLR) was performed based on **acceleration counts, additional body composition and anthropometric data**. In [Chen 97], the vertical and horizontal components were extracted from a triaxial accelerometer worn on the right hip. The model further considered weight, height, age, and gender. The proposed system achieved a correlation of 0.93. In [Plas 05], two sets of independent variables were compared. The first set consisted of age, weight, height, and acceleration count. The second set consisted of age, fat free mass, fat mass, and acceleration count. By using the second set, a higher correlation (0.93 vs. 0.90) and lower Standard Error of Estimate (SEE) (0.59 MJ/d vs. 0.70 MJ/d) were achieved. The accelerometer was attached to the lower back. In [Chen 97], nonlinear regression was further performed by applying two power parameters to the vertical and horizontal components of the triaxial accelerometer. The model achieved a correlation of 0.94.

In [Zake 10, Vath 10], the output of different **sensor modalities** were fused in order to predict the expended energy. The authors in [Zake 10] applied multivariate adaptive regression spline models to heart rate and accelerometer counts. The proposed approach achieved a RMSE of 154.3 kcal for 24-h TEE. In [Vath 10], the average integral of the mean-subtracted data for triaxial accelerometer and triaxial gyroscope were computed. Bayesian Linear Regression was applied to the features. The proposed system achieved a RMSE of 35 ml/min.

In [Roth 07, Stau 09, Mack 16, Mont 16, Mont 17], **Artificial Neural Networks** were used for regression mainly based on features from the accelerometer sensor. In [Roth 07], a biaxial accelerometer was placed on the hip. The proposed system achieved a MAE of 0.29 $\pm$ 0.10 kcal/min and outperformed approaches based on MLR. In [Stau 09], features were computed based on uniaxial accelerometer data. The features were used as input for an Artificial Neural Network. The proposed approach achieved a RMSE of 1.22 METs. In [Mack 16], features were extracted from nine single accelerometers worn on various body parts. Different sensor locations were compared. Using all accelerometers resulted in a correlation of 0.69. Using only one sensor resulted in correlations 0.77-0.81. Using the best two, three, and four sensor setting resulted in correlations of 0.80, 0.81, and 0.82, respectively. In [Mont 16], three accelerometer sensors were attached to right wrist, thigh, and ankle. Artificial Neural Networks were applied for energy expenditure estimation. The proposed system achieved a correlation of 0.79 and a RMSE of 2.16 METs. The three-sensor setup was further compared to a hip mounted accelerometer and showed better performance.

Activity-Specific Regression

In [Crou 06b, Pobe 06, Bono 09b, Crou 10, Sazo 11, Chen 13b, Elli 14], activity-specific regression was applied. In [Crou 06b], the coefficient of variation was used to classify walking/running and other activities. Depending on the classifier decision, one of two regression models were applied. A one-way repeated-measures ANOVA was used for performance assessment. A pairwise comparison with Bonferroni adjustments were applied to locate significant differences when needed. The proposed system achieved a correlation of 0.96 and a SEE of 0.73 ml/min. The proposed system achieved a significant improvement over single regression models. A refined version of the previously described approach was proposed in [Crou 10]. The goal was to eliminate the misclassification of walking or running when starting the activity in the center of a minute on the ActiGraph time axis. In addition to the method mentioned in [Crou 06b], the refined method considered surrounding epochs for the determination of the coefficient of variation. In [Pobe 06], a Hidden Markov Model was used to classify four activities. The corresponding MET value of the activity was taken from the 'Compendium of Physical Activities' [Ains 93]. The performance assessment included the fraction of time spent in light, moderate and vigorous activities. In [Bono 09b], six activities were recognized by a classification tree classifier. Class-dependent activity counts and metabolic equivalent values taken from the 'Compendium of Physical Activities' [Ains 93] were used to predict total energy expenditure, activity-related energy expenditure, and physical activity level. The proposed system achieved a better performance compared to using acceleration counts. In [Sazo 11], a footwear-based branched algorithm was proposed for energy expenditure estimation. The algorithm was based on integrated accelerometer and pressure sensor data as well as activity type information. First, the activities were categorized in sitting, standing, walking, and cycling resulting in four branches. Second, branch-dependent Passing-Bablok regressions were performed. The proposed approach achieved a RMSE of 0.69 METs. In [Chen 13b], k-means clustering was applied on features computed from ECG and accelerometer sensor as a preprocessing step. MLR was performed for each cluster separately. The proposed system achieved a RMSE of 0.96 kcal/min compared to 1.09 kcal/min using only MLR without clustering. In [Elli 14], a RF classifier was used to classify certain activities. RF regression was used to predict MET. The proposed system achieved a classification accuracy of 80.2 % and a RMSE of 1.00 METs using a wrist accelerometer.

Comparison of Algorithms

In [Twom 10, Alti 15, Mont 17], various algorithms for energy expenditure estimation were compared. In [Twom 10], five subjects were asked to perform walking on a motor-driven treadmill. The speed levels, which were proposed in [Bout 94], were used. Five accelerometers were placed on the subject's body, namely on ankle, knee, waist, wrist, and arm. The ground truth system provided the expended energy on a breath-by-breath basis. In order to synchronize reference and accelerometer data, the energy expenditure levels inside a certain epoch was averaged and considered as energy expenditure level of the whole epoch. Algorithms described in [Bout 97b, Chen 97, Crou 06b] were implemented and compared to each other. The non-linear model described in [Chen 97] achieved the best performance.

In [Alti 15], three methods for energy expenditure estimation and five body locations were compared.

Table 1.3.: State-of-the-art in energy expenditure estimation. Table contains topic, used sensor types (IMU: Inertial Measurement Unit, ACC: accelerometer, ECG: Electrocardiogram), number of acquired subjects (# Sub), number of acquired activities (# Act), applied algorithm (SLR: Simple Linear Regression, MLR: Multiple Linear Regression, NLR: Nonlinear Regression, DLF: Decision Level Fusion, ANN: Artificial Neural Network, HMM: Hidden Markov Model, BLR: Bayesian Linear Regression, SVR: Support Vector Regression, RF: Random Forest), performance (Root Mean Square Error (RMSE)), study environment (Env; L: laboratory, F: field) and reference (Ref). The algorithm developed and implemented in this thesis is marked in bold.

Topic	Sensors	# Sub	# Act	Algor.	RMSE	Env.	Ref.
Counts	ACC	11	9	SLR	NA	L	[Bout 94]
Counts	ACC	2	5	SLR	NA	L	[Bout 97b]
Counts	2 x ACC	25	7	SLR	NA	F	[Hend 00]
Counts	2 x ACC	28	5	SLR	0.53 MET	L	[Leen 03]
Counts	ACC	6	5	SVR	NA	L	[Su 05]
Counts	ACC	24	3	SLR	NA	L	[Motl 09]
Fusion	ACC	125	9	MLR/NLR	NA	L	[Chen 97]
Fusion	ACC	29	NA	MLR	NA	F	[Plas 05]
Fusion	**2 x IMU**	**10**	**6**	**DLF**	**NA**	**L**	**[Schu 14a]**
Fusion	ACC	102	NA	ANN	0.48 kcal/min	L	[Roth 07]
Fusion	ACC	48	18	ANN	1.22 MET	L	[Stau 09]
Fusion	e.g ACC	170	NA	Splines	154.3 kcal/24-h	L	[Zake 10]
Fusion	IMU	8	5	BLR	35 ml/min	L	[Vath 10]
Fusion	3 x ACC	25	14	ANN	2.16 MET	L	[Mont 16]
Fusion	9 x ACC	27	NA	ANN	1.46 MET	L	[Mack 16]
Activity	ACC	48	18	NLR	NA	L	[Crou 06b]
Activity	ACC	6	4	HMM	NA	L	[Pobe 06]
Activity	ACC	15	6	MLR	NA	F	[Bono 09b]

Table 1.3.: State-of-the-art in energy expenditure estimation (continued).

Topic	Sensors	# Sub	# Act	Algor.	RMSE	Env.	Ref.
Activity	ACC	48	18	NLR	NA	L	[Crou 10]
Activity	e.g. ACC	16	4	Branches	0.69 MET	L	[Sazo 11]
Activity	e.g. ACC	10	42	Clustering	0.96 kcal/min	L	[Chen 13b]
Activity	3 x ACC	40	8	RF	1.0 MET	L	[Elli 14]

The three methods included counts-based estimation methods, activity-specific estimation methods using METs lookup, and activity-specific estimation methods using accelerometer features. Activity-specific estimation methods using accelerometer features outperformed the two remaining methods by at least 23 %. It was further concluded that a five sensor system does not increase the accuracy compared to the best performing single sensor.

In [Mont 17], Artificial Neural Network models, linear regression, and linear mixed models were compared. The approaches were evaluated based on accelerometer data. The accelerometer sensors were attached to right hip, right thigh, and both wrists. The results showed that Artificial Neural Networks offered a significant improvement of the performance over linear models in case of a wrist-worn accelerometer. Nevertheless, linear models achieved similar performance like Artificial Neural Networks for hip- and thigh-worn accelerometers.

Findings

The literature review showed that mainly accelerometers were used for energy expenditure estimation (Table 1.3). According to [Bout 94, Bout 97a, Bout 97b, Chen 05, Hill 14], accelerometer counts were highly correlated to expended energy. Compared to accelerometer count-based regression models, improvements were achieved by adding anthropometric data [Plas 05], fusing accelerometer and gyroscope data [Vath 10], applying advanced machine learning methods such as Artificial Neural Networks [Roth 07] or Support Vector Regression [Su 05], and using activity-specific regression models [Crou 06b, Bono 09b]. The accelerometer devices were often worn at the lower back [Bout 94, Su 05, Plas 05] or hip [Chen 97, Roth 07, Mont 16]. The proposed systems were often evaluated on data from a laboratory setting, e.g. during walking on a motor-driven treadmill with different speed levels [Bout 94, Twom 10]. Usually, the IMU data and the ground truth energy expenditure were evaluated at the end of activity stages because the oxygen consumption reached a steady state [Bout 94, Vath 10, Alti 15].

Limitations and Open Issues

Although various different approaches for energy expenditure estimation were developed, limitations and open issues still remain. An accelerometer at waist level was often preferred in current state-of-the-art approaches

Table 1.4.: State-of-the-art in recognition of soccer-specific activities. Table contains number of acquired subjects (# Sub), number of classified activities (# Act), applied algorithm (NB: Naive Bayes, RF: Random Forest, RBM: Restricted Boltzman Machines, SVM: Support Vector Machine), accuracy (Acc), and reference (Ref). The algorithms developed and implemented in this thesis are marked in bold.

# Sub	# Act	Algorithm	Acc [%]	Ref
15	7	NB	80.0	[Mitc 13]
10	6	RF	83.9	[Ahma 15]
12	**2**	**Hierarchical**	**84.2**	**[Schu 15]**
11	**2**	**SVM**	**96.8**	**[Schu 16]**
6	6	RBM	86.5	[Hoss 17]

for energy expenditure estimation [Bout 94, Chen 97]. Nevertheless, estimating the energy expenditure during more complex activities such as sweeping, biking, and jumping required data fusion of multiple sensor types and sensor positions [Vath 10, Mont 16]. Current energy expenditure estimation approaches mainly **fused** various sensor sources at **feature level** [Hall 97, Yang 14]. Two limitations were identified. First, changing the number of sensors resulted in a retraining of the whole system [Mont 16]. Integrating new activities, which might require additional sensor sources, is cumbersome. Second, the performance of the sensor-based systems during run-time is affected by degradation, interconnection failures, and jitter in placement of single sensors [Zapp 07]. An open issue is to develop a multi-sensor system, which is flexible in adding new sensor types and positions.

1.5.3 Recognition of Soccer Activities

IMUs were often used to monitor athletes in different sports including Australian football [Boyd 13, Walk 16], netball [Corm 14], rugby [Kell 12, Gabb 15, Kaut 15b], basketball [Mont 10, Bai 16], tennis [Conn 11], table tennis [Blan 15], golf [Nam 14], cricket [McNa 15], ski jumping [Groh 14], snowboarding [Groh 16], ice hockey [Hard 15], swimming [Fult 09, Jens 13, Jens 16], and cross-country skiing [Mars 12]. Nevertheless, only a few algorithms were available dealing with HAR in soccer using IMU data [Mitc 13, Ahma 15, Hoss 17]. The algorithms are described in the next paragraphs and are comprised in Table 1.4. The different algorithms were investigated regarding the considered data collection, algorithmic design, and performed evaluation. The literature review concludes with general findings, limitations, and open issues.

Data Collection

Accelerometer data were acquired with one sensor on the upper area of the back [Mitc 13], on the shank [Ahma 15], or on the wrist [Hoss 17]. In [Mitc 13] and [Hoss 17], 5-a-side soccer matches were recorded with each lasting one hour. The authors in [Ahma 15] collected data using a pre-defined protocol consisting of various exercises.

In this thesis, two sensors were used in the data collection. One sensor was placed in a cavity of each soccer boot. The sensor included an accelerometer and a gyroscope. In the mentioned research work, data of either exercises or games were collected. In this work, data of both exercises and a game were acquired.

Algorithmic Design

The proposed algorithms mainly developed flat classification systems based on the ARC distinguishing various activities with one single classifier (Figure 1.3a and 1.1). Discrete Wavelet Transform features were extracted and used as input for a NB and RF classifier in [Mitc 13, Ahma 15], respectively. Compared to the shallow architectures in [Mitc 13, Ahma 15], Restricted Boltzman Machines were used as a deep learning technique for classifying soccer-specific activities. The proposed architectures in [Mitc 13, Ahma 15, Hoss 17] did not consider the modeling of a NULL class containing activities, which are unknown to the classifiers.

In this work, shallow architectures were applied, which were optimized for kicks. The developed methods were based on peak detection and considered instances of a NULL class containing various activities such as dribbling, running, and tackling.

Evaluation

The proposed algorithms in [Mitc 13, Ahma 15, Hoss 17] were evaluated considering activities like walking, jogging, sprinting, attempting a tackle, jumping, dribbling, kicking, and passing. The data acuired in [Mitc 13, Ahma 15] were mainly segmented manually. In [Mitc 13], a window of 3 s centered around the activity was considered for further processing. The authors in [Ahma 15] annotated sensor data considering a window length of 3 s for training and testing the proposed algorithm. The authors in [Hoss 17] applied an automatic change point detection algorithm to pre-segment the data. Each segment had a different length depending on the performed activity. In [Mitc 13, Hoss 17], instances of the same player were used in the training and test set. In contrast, the authors in [Ahma 15] applied a Leave-One-Subject-Out Cross-Validation (LOSO-CV) to compute the accuracy. The three algorithms achieved accuracy values of above 80.0 % (Table 1.4).

In this work, two evaluation strategies were considered. First, training and testing were based on sensor data acquired during exercises. Second, a trained algorithm based on exercise data was applied to game data. In order to provide training instances for the exercises, a peak detection approach was used to automatically segment kicks. Compared to [Mitc 13, Ahma 15], the complete game data was considered in the testing phase without pre-segmenting different parts.

Findings

The literature review showed that a single accelerometer can be used to recognize soccer-specific activities with accuracy values above 80 % [Mitc 13, Ahma 15, Hoss 17]. The algorithms consisted of a flat classification system and were evaluated on either a pre-defined set of exercises [Ahma 15] or games [Mitc 13, Hoss 17]. The list of considered activities included dribbling, sprinting, tackling, and the two main important kick types, full-instep as well as side-foot kicks [Leva 98]. Full-instep kicks were used for the generation of fast ball speed and the ball is hit by the medial-superior portion of the instep. Side-foot kicks were used when precision is the main priority and the ball is hit by the medial aspect of the foot.

Limitations and Open Issues

Various limitations and open issues were found. Although full-instep and side-foot kicks were two of the most important soccer-specific activities, current HAR systems often achieved lower performance for kicking compared to other activities [Mitc 13, Ahma 15]. An open issue is to find solutions to increase the accuracy of kicking activities. Current approaches often do not consider a NULL class in the evaluation [Ahma 15], which would be necessary, if the HAR system is applied in a real match. Most algorithms were further evaluated using the same player in both training and test set [Mitc 13, Hoss 17], so that the application of proposed systems to unseen players is often challenging. Although current available algorithms were evaluated on game data, only pre-segmented sensor parts were mainly considered neglecting a huge amount of additional activities. Thus, the application of the proposed systems to complete game data is often unknown. Training machine learning algorithms requires the collection of a large amount of data in the target domain, e.g. in game scenarios. Nevertheless, the number of activity instances such as kicking is rather low in games [Mitc 13] compared to exercises [Ahma 15]. A high amount of games have to be acquired in order to provide sufficient training data. An open issue includes the assessment, whether exercise data alone provide enough information to train systems which can also be applied in game scenarios.

1.5.4 Ball Speed Estimation

Full-instep kicks can be classified due to the achieved ball speed, which is the main indicator of kicking success in soccer [Lees 98, Dör 02, Kell 07]. A reliable ball speed estimation system is required, which can be used to determine the intensity of the full-instep kick. Thus, a literature review about ball speed analysis was performed. The literature review is structured as follows. First, research work about dependent factors for ball speed is introduced. Second, one regression technique for IMU-based ball speed estimation is described. Third, general findings, limitations, and open issues are mentioned.

Dependent Factors for Ball Speed

Current ball speed measurements in soccer were often performed in a laboratory environment using camera-based systems [Nuno 06, Shin 09]. One research area was the investigation of dependent factors for ball speed, which was mainly performed for full-instep kicks based on stationary ball scenarios. It was concluded that the ball speed is dependent on various factors such as technique [Lees 10], approach angle [Kell 04], gender [Barf 02], support leg dynamics [Inou 14, Augu 17], and the trunk kinematics [Nait 10, Full 15]. Nevertheless, the final ball speed largely depended on the characteristics of the foot-ball impact [Lees 98, Bull 99, Kell 07, Ishi 12]. It was concluded that powerful kicks were achieved through a high foot velocity [Kell 07]. Regression analysis showed a linear relationship of ball speed and foot velocity with correlation coefficients above 0.81 [Zern 78, Nuno 06]. The results in [Bull 99] further indicated that a high ball speed can be achieved by maximizing the angular velocity of the lower leg.

In this thesis, the foot-ball impact was considered for ball speed estimation. Compared to the previously mentioned literature, an IMU sensor was used in order to determine the final ball speed.

Inertial Sensors for Ball Speed Estimation

Research work regarding IMU-based ball speed estimation is rather limited. In [Zwic 12], a biomechanical collision model was used to estimate the ball speed based on a one-axial $\pm$ 70 g accelerometer in antero-posterior direction. The accelerometer was attached to the malleolus lateralis. The velocity of the malleolus lateralis was estimated based on integrated accelerometer values followed by regression analysis. A significant correlation was found between the integrated acceleration values and the velocity of the malleolus lateralis ($r = 0.68$, $p < 0.001$). Sensor data of one player was used for evaluation. A Mean Absolute Percentage Error (MAPE) of 2.81 % was achieved between predicted and true ball speed.

In this work, a more sophisticated regression technique was applied to ball speed estimation combining accelerometer and gyroscope. The gyroscope was included in order to exploit the influence of the angular velocity on the final ball speed [Bull 99].

Findings, Limitations, and Open Issues

The literature review showed that foot-ball impact is one major dependent factor of the final ball speed [Ishi 12]. A linear relationship was found between foot velocity and ball speed [Nuno 06]. Nevertheless, research about IMU-based ball speed estimation is rather limited [Zwic 12] and various limitations were found. The data collection suffered from a minimal set of sensor information based on an uniaxi-al accelerometer attached to one player. Investigations further included only stationary ball scenarios. The regression considered the integrated acceleration values as predictor, which might not be sufficient in modeling the kicking execution. Open issues include the evaluation of more sophisticated regression techniques based on more multi-axial sensor sources and more subjects. The applicability of IMU-based systems for ball speed estimation in realistic scenarios remains further unknown.

1.6 Related Products

HAR systems based on machine learning techniques using IMU data are more and more implemented in commercial wearables in order to monitor daily and soccer activities. Six examples are introduced in the following paragraphs.

- **ActiGraph GT9X Link** (ActiGraph, Pensacola, FL): records accelerometer data to compute objective activity metrics including activity counts, energy expenditure, and body position [1]. The device can be worn on wrist, waist, ankle, or thigh.

- **fitbit alta** (fitbit, San Francisco, CA): comprises activity recognition and energy expenditure algorithms to capture activities such as walking, bicycling, as well as sleeping and to estimate consumed calories [2].

- **Misfit Shine 2** (Misfit, San Francisco, CA): uses accelerometer data to track among others number of steps, sleep duration, traveled distance, and consumed calories [3].

- **APEX** (STATSports, Newry, Northern Ireland): comprises pods worn on the upper back of athletes and multiple beacons around the pitch [4]. The pods include various sensors like accelerometer and gyroscope as well as antennas. Positional and inertial data are used to compute metrics such as distance, speed, number of sprints, and impacts. The system can be applied in both indoor and outdoor environments and is mainly used by elite sports teams.

- **Zepp Play Soccer** (Zepp, San Jose, CA): consists of an IMU inserted into a calf sleeve [5]. The system collects individual game statistics, capture real-time game reports, and create video highlight reels. The performance metrics include covered distance, maximum running speed, number of sprints, number of kicks, and kicking speed. The scenes that should be included in the highlight reel have to be selected manually.

[1] https://actigraphcorp.com/actigraph-link/
[2] https://www.fitbit.com/de/alta
[3] https://www.misfit.com
[4] https://statsports.com/apex/
[5] https://www.zepp.com/en-us/soccer/

- **adidas miCoach Smart Ball** (adidas AG, Herzogenaurach, Germany): uses motion sensors integrated in a soccer ball to estimate among others kick speed as well as spin and trajectory of the ball after leaving the foot [6].

The previously introduced devices monitor daily as well as soccer activities in order to compute various metrics. Depending on the application, the metrics include sleeping duration, activity bouts, expended energy, covered distance, number of sprints, number of kicks, and kicking speed. In this thesis, the developed soccer kick classification approach was integrated in a system, which generated video highlight reels. Compared to the Zepp Play Soccer system, the highlight reels were selected automatically based on a sensor-based HAR system. The system is described later in this thesis.

1.7 Contributions

In this thesis, nine contributions were made to the field of HAR using machine learning based on IMU data. The contributions can be divided into three groups.

The first group of five contributions deals with the development and implementation of HAR algorithms applied to **daily activities**:

1. Current state-of-the-art algorithms for HAR mostly provided flat classification systems, which relied on a single classifier decision and required a retraining of the complete system in a changing environment. Hierarchical classification systems overcome these issues, but the adjustment of existing approaches to a new environment would often be cumbersome and time-consuming. In this work, a novel, hierarchical classifier structure was developed, which relied on multiple decisions along the path through the hierarchy and provided automatically generated machine learning systems optimized for a certain activity group. The proposed hierarchical approach would enable e.g. an increased flexibility of activity trackers (see section 1.6) to react to environmental changes without exchanging the complete system. The algorithm is described in section 3.3. This contribution was published in [Schu 13b].

[6] `https://www.adidas.com/us/micoach-smart-soccer-ball/G83963.html`

2. Multiple sensor nodes are often necessary to capture the wide range of desired activities [Bao 04]. Current approaches mainly combine the information of the different sensor nodes at feature level. Feature level fusion often includes drawbacks such as the need of retraining of the whole systems, if a new sensor is added, the influence of sensor degradation during run-time, and a high computational complexity [Zapp 07, Bano 13]. To close this gap, an algorithm was developed based on fusing the information of individual sensor nodes at decision level. Compared to existing approaches, the proposed method was evaluated on a large set of daily activities including gyroscope sensor and is not restricted to certain classifier types. The system allows the absence of sensor information from different positions and might be more robust in Body Sensor Networks. The algorithm is described in section 3.4. This contribution was published in [Schu 13a].

3. Although various algorithms for HAR already exist (section 1.5.1), the identification of the best suitable technique for a specific application is still challenging. Most state-of-the-art approaches were compared based on the final achieved performance measure. Nevertheless, the explanatory power of this kind of comparison is limited, since different databases and evaluation strategies were used. To overcome this limitation, the two previously proposed algorithms were compared to five re-implemented state-of-the-art approaches [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12] using the same benchmark dataset and evaluation framework. This procedure of comparing multiple algorithms based on the same conditions can be seen as one proposed solution finding the best HAR algorithm for future activity recognition challenges. The comparison is described in section 3.5. This contribution was published in [Schu 13b]. The benchmark dataset is publicly available (https://www.activitynet.org).

4. Systems for the recognition of daily activities were mainly evaluated on isolated datasets, which reduced the number of instances for algorithm training and reduced the exploratory power of findings. In order to overcome these limitations, a novel method was developed, which increased the number of datasets without the need of a new data collection session. A workflow was introduced to merge publicly available datasets to provide a larger database. The

applicability of the proposed database fusion strategy was shown in the evaluation of one proposed HAR algorithm. A larger database for training enables the development of more robust HAR systems and increases the explanatory power of results. The database fusion and the corresponding activity recognition algorithm are described in section 3.6. This contribution was published in [Schu 14b].

5. Dynamic activities, e.g. walking can further be classified into intensity levels, such as light, moderate, and vigorous [Hend 00, Bono 09b]. The intensity level is usually defined by the expended energy. Current research work on sensor-based energy expenditure estimation mainly applied feature level fusion to accelerometer data followed by regression analysis [Bout 94]. Feature level fusion approaches often suffer from sensor degradation and low flexibility in terms of adding as well as removing sensors. To overcome the mentioned issues, an algorithm was developed and implemented, which fused the information of sensors at decision level. The proposed approach can be applied in multi-sensor systems, which should be robust against hardware failures as well as flexible in adding or removing sensors. The algorihm is described in section 3.7. The dataset used for evaluation is publicly available (https://www.activitynet.org). This contribution was published in [Schu 14a].

The second group of three contributions is split into two parts. The first part includes novel HAR algorithms applied to **soccer-specific activities**. The second part includes the integration of these HAR algorithms in a specific application:

6. Full-instep and side-foot kicks are two of the most important soccer-specific activities. Nevertheless, current algorithms often suffered from a low performance of kicks in a flat classification system. The recognition systems were further only evaluated in training sessions or manually segmented game data [Mitc 13, Ahma 15]. In this thesis, two HAR algorithms were developed and compared, which were optimized for kicks. The first method mainly performed a hierarchical classification (Figure 1.3b). The second method included additional expert knowledge about the different phases during kick execution. Both techniques were evaluated on a pre-defined set of exercises. The system, which achieved the better performance, was further evaluated on complete game data without pre-segmenting

certain parts. The proposed methods are described in section 4.3. The algorithms were published in [Schu 15] and [Schu 16].

7. The achieved ball speed during a full-instep kick is the main indicator of kicking success in soccer [Lees 98, Dör 02, Kell 07]. Nevertheless, current ball speed measurements were often performed in a laboratory environment using expensive camera-based systems and mainly considering stationary ball scenarios [Nuno 06, Shin 09]. In this thesis, a low-cost solution for ball speed estimation was developed and implemented. Compared to previous work, the approach was evaluated on dynamic exercises in unconstrained environments. The proposed ball speed estimation system is described in section 4.4 and was published in [Schu 16].

8. Video summaries are used in sports to provide among others athletes and coaches with important highlight scenes [Rehm 14]. Highlight scenes include goals, bookings, shots on goals, or penalties. Generating these summaries often required a manual, time-consuming selection and cutting of relevant sequences [DOra 10]. Automatic approaches required e.g. information of announcers' speech, availability of video shots such as close-up or replay shots, and textual overlays such as game statistics [Ekin 03, Baba 04]. Due to the previously mentioned requirements, the proposed automatic systems can mainly be applied to TV broadcast material in professional sports. Systems and methods for recreational sports are missing. To close this gap, a novel, innovative, and low-cost solution was developed for video summary generation. The previously mentioned algorithms for kick classification and ball speed estimation were partially integrated in the system. A patent application was filed encompassing parts of the algorithmic approaches and the final video summary system [Kirk 16]. An extension of the system was further published in [Schu 16]. The complete system is briefly introduced in section 4.5.

The last contribution deals with the introduction of **Big Data** to the HAR research field:

9. The IoT era provides the generation of Big Data [Swan 12]. Although Big Data offers new possibilities in HAR research, e.g. the automated detection of deviant behavior based on long-term monitoring, the current design of HAR systems did not consider Big

Data concepts and tools. In this thesis, Big Data is introduced to the HAR research field. The contribution was fourfold. First, a survey about the state-of-the-art in wearable-based HAR was provided, which was already presented in section 1.5.1. Second, a definition of Big Data in the context of HAR was given. Third, extensions of the traditional ARC were introduced, which included mandatory concepts for Big Data ready activity recognition. Fourth, open issues and future research directions were comprised. Tackling the selected items and topics enables the exploitation of the huge potential of Big Data for activity recognition challenges. This contribution was submitted [Schu 18].

1.8 Outline

The outline of the thesis is shown in Figure 1.4. Chapter 2 summarizes the fundamentals including indirect calorimetry, anatomical system, IMU, and machine learning. Chapter 3 summarizes the contributions regarding HAR with respect to daily activities. Chapter 4 summarizes the contributions regarding HAR with respect to soccer-specific activities. Chapter 5 summarizes the contributions regarding Big Data ready HAR. Chapter 6 provides the overall findings of the thesis, a general discussion, and an outlook.

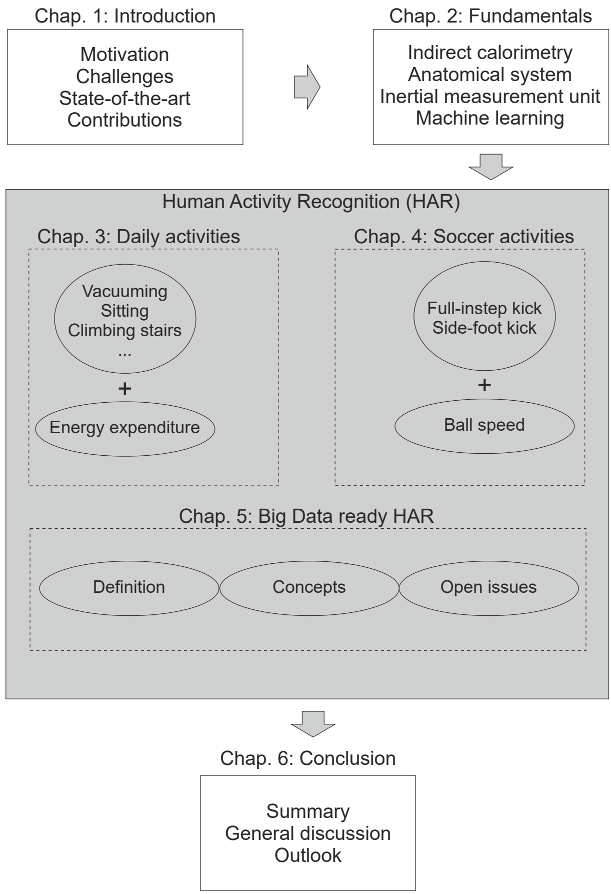


Figure 1.4.: Overview of this thesis.

Chapter 2

Fundamentals

In this chapter, the fundamentals are described which are needed to understand the concepts and components of the developed algorithms in the upcoming chapters. First, indirect calorimetry is briefly explained, since it was used as ground truth for the IMU-based energy expenditure estimation [Löl 10]. Second, the anatomical position, planes, and axes of the human body are described, since the corresponding terms were further used for the definition of the coordinate systems and the explanation of certain movements. Third, the principles of an IMU sensor are introduced, since data of this sensor type were used as input for the developed algorithms. Fourth, the applied machine learning techniques are briefly explained.

2.1 Indirect Calorimetry

Indirect calorimetry exploits the fact that energy-releasing reactions in the body depend on oxygen consumption [McAr 07]. The process of cellular respiration needs oxygen in order to produce energy and further results in the production of carbon dioxide [Kenn 11, Habe 13]:

$$Oxygen + Glucose \rightarrow Energy + Carbon\,dioxide + Water. \tag{2.1}$$

Common indirect calorimetry procedures include portable spirometry, bag technique, and computerized instrumentation [McAr 07]. Computerized instrumentation systems are often connected to a motorized treadmill and determines the actual volume per minute of oxygen consumption $\dot{V}O_2$ based on the expired air volume $\dot{V}_E$, the fraction of oxygen

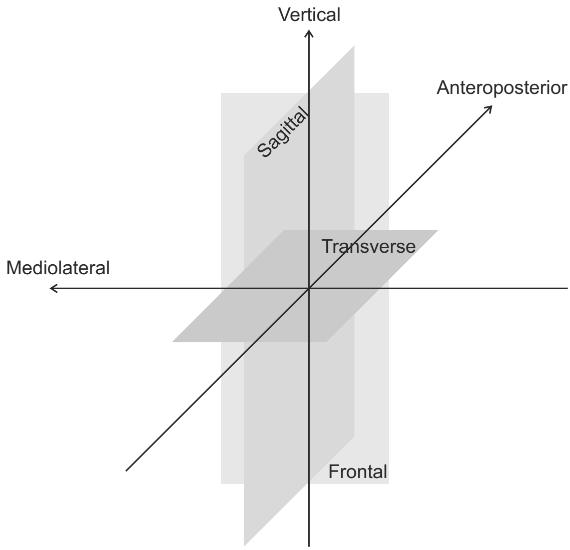


Figure 2.1.: Anatomical planes and axes (adapted from [McGi 13]).

in the expired air $F_E O_2$, and the fraction of carbon dioxide in the expired air $F_E CO_2$:

$$\dot{V}O_2 = \dot{V}_E(1 - F_E O_2 - F_E CO_2) \cdot 0.265 - (\dot{V}_E \cdot F_E O_2). \tag{2.2}$$

The corresponding expended energy is often given as MET [Schu 01] defined as

$$MET = \frac{\dot{V}O_2(\frac{mL \cdot O_2}{kg \cdot min})}{3.5 \frac{mL \cdot O_2}{kg \cdot min}}. \tag{2.3}$$

MET can be used to express the intensity of activities in the steady state as multiples of the sitting at rest state [Jett 90].

2.2 Anatomical Position, Planes, and Axes

A commonly used reference position of the human body is denoted as **anatomical position** which means standing erect, facing forward, with the feet aligned parallel to each other, toes forward, arms and hands hanging straight below the shoulders at the sides, fingers extended, and palms facing forward [McGi 13]. Movements of limbs or other anatomical structures are described in reference to the anatomical position. Specific imaginary **anatomical planes** and **axes** can be defined, e.g. in order to describe human movement (see Figure 2.1).

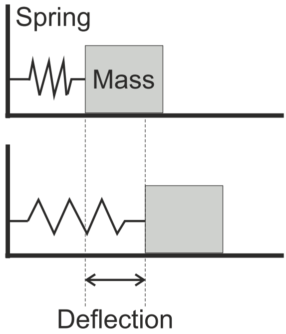
Three anatomical planes pass though the body, namely sagittal, frontal, and transverse plane [Bart 07, McGi 13, Godf 08]. The sagittal plane runs anterior (front) to posterior (back) and superior (top) to inferior (bottom). The frontal plane runs side to side and superior to inferior. The transverse plane runs from side to side and anterior to posterior.

Three anatomical axes are perpendicular to the anatomical planes, namely anteroposterior, mediolateral, and vertical axis [Bart 07, McGi 13, Godf 08]. The anteroposterior axis runs from anterior to posterior and is perpendicular to the frontal plane. The mediolateral axis runs from left to right and is perpendicular to the sagittal plane. The vertical axis runs from top to bottom and is perpendicular to the transverse plane.

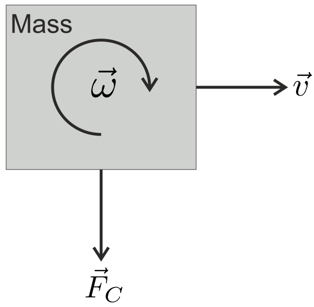
2.3 Inertial Measurement Unit

According to [Kemp 11], IMUs aim at determining the dynamic behavior of objects by exploiting inertial forces. The use of microelectromechanical systems technology results in low-cost, small-size, and lightweight IMUs [Barb 01]. The small size of IMUs further favors wearability and portability and enables a good tradeoff between comfort and cost to be applicable in various domains [Brig 11]. The two basic dynamic parameters include linear acceleration measured by an accelerometer and angular velocity measured by a gyroscope. The corresponding principles are briefly explained in the following paragraphs and are illustrated in Figure 2.2.

According to [Kemp 11], accelerometer devices implement the spring mass principle with one degree of freedom (Figure 2.2a). When acceleration occurs, a small mass system responds by applying a force to a spring [Math 04b]. This causes stretching or compressing of the spring. This deflection is transformed into an electrical signal. The output of a body-worn accelerometer is dependent on several sources [Redm 85]: body acceleration, gravitational acceleration, external vibrations (e.g. resulting from vehicles), accelerations caused by bouncing of the sensor against other objects, jolting of the sensor caused by loose attachment.



(a) Accelerometer. Spring mass principle. From top to bottom: acceleration to the left results in an expanded spring.



(b) Gyroscope. Mass is brought into vibration with momentary speed $\vec{v}$). Rotating device with angular velocity $\vec{\omega}$ results in additional displacement of mass caused by Coriolis force $\vec{F}_C$.

Figure 2.2.: Illustration of accelerometer (a) and gyroscope (b) (adapted from [Roet 06]).

The principle of the gyroscope is illustrated in Figure 2.2b. A mass is brought into vibration with momentary speed $\vec{v}$ [Roet 06]. Rotating the gyroscope with an angular velocity $\vec{\omega}$ results in an additional displacement of the mass caused by the Coriolis force $\vec{F}_C$. The direction of $\vec{F}_C$ is perpendicular to the original displacement. The Coriolis force induced motion is sensed by electrodes and converted to an electrical signal [Jone 13].

According to [Kemp 11], imperfections of the mechanical sensors influence the whole system parameters. The imperfections include manufactural imperfections, nonlinearities, and misalignment errors.

2.4 Machine Learning

The main steps of the machine learning pipeline are shown in Figure 2.3 and are described in the following section. The details were taken from [Niem 90, Duda 00, Diet 02, Bish 06, Theo 09, Witt 11, Bull 14].

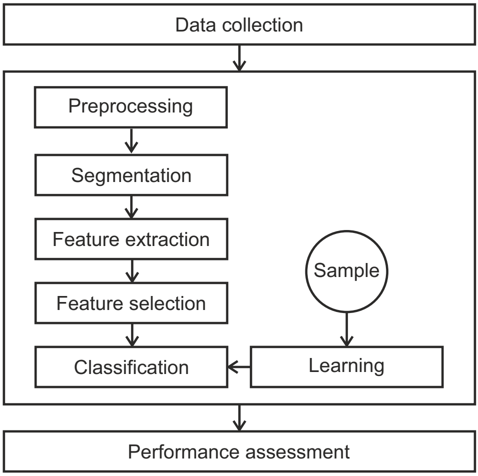


Figure 2.3.: Machine learning pipeline (adapted from [Niem 90, Duda 00, Theo 09]).

2.4.1 Data Collection

According to [Niem 90], technical sensors enable the collection of representative samples in a specific task domain. In this thesis, various inertial sensors were attached to different body parts in order to acquire raw data, while subjects performed certain daily and soccer-specific activities. The raw data is usually sampled at regular intervals and is represented as a multivariate time series [Bull 14].

2.4.2 Preprocessing

The preprocessing step performs a transformation of the raw multivariate and non-synchronous time series to a preprocessed time series [Bull 14]. Preprocessing steps include synchronization, calibration, unit conversion, normalization, re-sampling, data-level fusion, Butterworth filtering, Fourier Transform, and peak detection [Hall 97, Pall 99, Kuo 06, Bish 06, Bull 14]. A brief introduction of the preprocessing steps calibration, unit conversion, Fourier Transform, Butterworth filtering, and data level fusion is given in the following paragraphs.

Calibration and Unit Conversion

As mentioned in section 2.3, imperfections of the mechanical sensors influence the whole system parameters. In order to correct systematic errors or bias in IMU data and to map raw data to standardized units, a calibration step has to be performed [Yang 14]. The calibration step is usually performed during production time of the sensor unit. Regular re-calibration is required due to aging, decay, and damage. The single calibration procedures, which were used in this work, are described later in the corresponding chapters.

Discrete Fourier Transform

Daily and soccer-specific activities usually occur in certain frequency bands. In order to perform frequency analysis, the time-domain signal $\{x[n]\}|_{n=0,1,\ldots,N-1}$ is converted to the representation in the frequency domain $\{X[k]\}|_{k=0,1,\ldots,N-1}$ by the Discrete Fourier Transform [Kuo 06]:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j(\frac{2\pi k}{N})n}. \tag{2.4}$$

In this thesis, the Discrete Fourier Transform was applied to compute certain features in the frequency domain.

Butterworth Filtering

In order to remove noise and undesired frequency bands Butterworth filtering can be applied [Kuo 06, Bull 14]. Four different filter types include lowpass, highpass, bandpass, and bandstop filter. The Butterworth filter is an example of an Infinite Impulse Response filter [Kuo 06]. According to [Kuo 06], the design of a digital Infinite Impulse Response filter includes the definition of the corresponding filter in the analog domain and applying a bilinear transformation. The magnitude response $|H(f)|$ of a Butterworth lowpass filter is defined as

$$|H(f)| = \frac{1}{\sqrt{1 + (\frac{f}{f_c})^{2L}}}. \tag{2.5}$$

f denotes the frequency in Hz. The Butterworth filter is defined by cutoff frequency f_c and filter order L.

In this thesis, a Butterworth high pass filter was used e.g. to detect ball contacts in soccer.

Data Level Fusion: Signal Magnitude Vector Computation

The computation of the Signal Magnitude Vector (SMV) can be seen as one example for data level fusion, which neglects the direction of movement [Hall 97, Kara 06]:

$$SMV[i] = \sqrt{x[i]^2 + y[i]^2 + z[i]^2} \tag{2.6}$$

$x[i]$, $y[i]$, and $z[i]$ denote the samples at index i of the three sensor axes. The SMV was e.g. also applied to detect ball contacts.

2.4.3 Segmentation

Segmentation enables the identification of desired parts in the preprocessed data streams which are likely to include information about the target activities [Bull 14]. Three segmentation methods were used in this work depending on the application.

Sliding Window

A fixed-size window is shifted over the signal and the corresponding sensor data in the window is used for further processing [Bull 14]. Two parameters have to be defined. First, the window size, which has an influence on the delay of the activity recognition system. Second, the overlap of the windows, which has an influence on the segmentation precision and computational load. The developed algorithms for the recognition of daily activities mainly applied the sliding window technique.

Energy

The energy of the sensor signal is computed exploiting the fact that activities are executed with different intensities [Bull 14]. Signal parts are segmented which exceed a certain threshold. The developed kick classification algorithms included an energy-based segmentation routine.

Additional Sources

Additional information can be used to segment IMU data, such as GPS traces or sound records acquired by smartphone as well as diary content [Bull 14]. In this work, a smartphone app was often used to provide a coarse segmentation of the performed activities.

2.4.4 Feature Extraction

According to [Niem 90], a direct classification based on the samples is not suitable. Thus, in order to reduce the amount of processed data and to focus on important information for the classification, features are extracted. The features enable a representation of the observation which positively influence the performance of the subsequent classifier [Duda 00]. According to [Bull 14], features which correspond to the same class should be clustered in the feature space and features which correspond to different classes should be apart. Feature extraction often requires knowledge of the domain. Algorithms for HAR based on inertial sensor data often extract features in either time or frequency domain [Lara 13]. Features, which were used in this thesis, are given in Table 2.1. Time domain features include statistical measures such as mean, variance, or percentiles [Liu 12]. They are used to describe the amplitude distribution in sensor signals. One commonly used feature is further the Signal Magnitude Area (SMA), which neglects both the direction of movement and the signal distribution across sensor axes [Math 03]. Frequency domain features are often based on the previously described Discrete Fourier Transform and include e.g. spectral centroid and bandwidth [Knee 16]. In HAR, frequency domain features are often applied to detect periodic activities such as walking [Bao 04].

2.4.5 Feature Selection

According to [Guyo 03, Bull 14], advantages of feature selection includes reducing the measurement and storage requirements, reducing training and testing times, and addressing the curse of dimensionality to improve prediction performance. A manual selection of features is challenging and time-consuming. A set of methods for automatic feature ranking and selection are available, which can be grouped into wrapper, filter, and hybrid approaches [Theo 09]. In this work, a correlation-based feature selection [Hall 99a, Hall 99b, Hall 00] technique was applied.

Correlation-based feature selection belongs to filter methods, which are based on class separability criteria and are classifier-independent [Sanc 07, Theo 09].

Table 2.1.: List of extracted features in time and frequency domain.

Time-domain	**Frequency-domain**
Descriptive statistics for calculation of numerical summary statistics for observations $x[0]$,$x[1]$,...,$x[N-1]$ can be used as features [Liu 12, Devo 12]. **Mean** μ is a measure of the center of observations [Devo 12, Gupt 13]: $$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x[i] \qquad (2.7)$$ Further measures of location include **percentiles** [Devo 12]. The observations are sorted in ascending order and are ranked from 1 to N [Gupt 13]. The p-th percentile defines the observation that corresponds to the rank $p \cdot [(N+1)/100]$. In this thesis, the 10th, 25th, 50th (median), 75th, and 90th percentile were computed. **Standard deviation** δ is a measure of the variability in the observations [Devo 12]): $$\delta = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N-1} \lvert x[i]-\mu\rvert^2} \qquad (2.8)$$ **Signal Magnitude Area** (SMA) combines the three sensor axes x, y, and z [Math 03, Alle 06, Kara 06]: $$SMA = \frac{1}{N}\sum_{i=0}^{N-1} = (\lvert x[i]\rvert + \lvert y[i]\rvert + \lvert z[i]\rvert) \qquad (2.9)$$	Frequency domain features are often extracted in speech and audio processing [Theo 09, Knee 16] including **spectral centroid** SC and **bandwidth** BW. These features are defined as $$SC = \frac{\sum_{k=0}^{N-1} k \cdot X[k]}{\sum_{k=0}^{N-1} X[k]} \qquad (2.10)$$ $$BW = \frac{\sum_{k=0}^{N-1} \lvert SC-k\rvert \cdot X[k]}{\sum_{k=0}^{N-1} X[k]} \qquad (2.11)$$ SC measures the center of gravity of the magnitude spectrum. BW measures the spectral range around SC and can be seen as the variance from the mean frequency [Knee 16].

According to [Hall 99a], correlation-based feature selection ranks feature subsets according to a class separability criterion $M(S)$ defined as

$$M(S) = \frac{l\bar{c}_{yx}}{\sqrt{l + l(l-1)\bar{c}_{xx}}}. \tag{2.12}$$

S denotes the current feature subset containing l features. $\bar{c}_{yx}$ denotes the mean correlation between feature x and class y ($x \in S$). $\bar{c}_{xx}$ denotes the mean feature-feature intercorrelation. The numerator is an indicator how predictive a set of features is [Hall 00]. The denominator is an indicator how much redundancy is in the feature set. Thus, features are preferred which are highly correlated with the corresponding class and uncorrelated with each other [Sanc 07]. A best-first search is further applied, which enables heuristically searching a graph [Rich 91]. In the case of correlation-based feature selection, the heuristic evaluation function is given by equation 2.12. Each step of the best-first search consists of selecting the most promising nodes that were generated so far and expanding the chosen node by applying the rules to generate the corresponding successors. If the current explored path is less promising, backtracking is performed.

2.4.6 Classification

In this work, mainly five classifiers were applied including Naive Bayes (NB), k - Nearest - Neighbor (k-NN), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Random Forest (RF).

Naive Bayes

Classification can be formulated in probabilistic terms based on the Bayesian rule

$$p(y|\vec{x}) = \frac{p(\vec{x}|y)p(y)}{p(\vec{x})}. \tag{2.13}$$

$p(y|\vec{x})$, $p(\vec{x}|y)$, $p(y)$, and $p(\vec{x})$ denote the a posteriori probability, likelihood, a priori probability, and evidence, respectively [Duda 00]. $\vec{x} \in \mathbb{R}^l$ and $y \in \mathbb{Z}$ denote the feature vector containing l features and the corresponding class, respectively. NB is a classifier which assumes that the class-conditional densities in equation 2.13 are normally distributed and the features are conditionally independent [John 95]

$$p(\vec{x}|y) = \prod_{i=0}^{l-1} p(x_i|y) \tag{2.14}$$

$$p(x_i|y) = \mathcal{N}(x_i; \mu, \sigma). \tag{2.15}$$

$\mathcal{N}(x_i; \mu, \sigma)$ denotes a normal distribution defined by the mean μ and the standard deviation σ

$$\mathcal{N}(x_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}. \tag{2.16}$$

According to [Theo 09], the assignment of a new data point $\vec{x}_0$ to a certain class label $\hat{y}_0$ is performed by

$$\hat{y}_0 = \arg\max_{y_j} p(\vec{x}_0|y_j) \qquad j = 1, 2, ..., M. \tag{2.17}$$

M denotes the number of classes. The NB classifier, whose underlying density function is assumed to be normal, belongs to parametric classification techniques.

k-Nearest Neighbor

k-NN is a classifier which directly estimates the a posteriori probability shown in equation 2.13 [Duda 00]. For classifying a new data point $\vec{x}_0$, the k closest training points have to be determined [Cove 67], e.g. by using the Euclidean distance

$$d_i = ||\vec{x}_i - \vec{x}_0|| \qquad i = 1, ..., q. \tag{2.18}$$

q denotes the total number of stored training points. The a posteriori probability is estimated by

$$p(y_i|\vec{x}_0) = \frac{k_i}{k}. \tag{2.19}$$

k_i denotes the number of data points amongst the k nearest neighbors that belong to class y_i. k-NN classifiers require no model to be fit [Hast 09]. Majority voting among the k neighbors is applied to determine the final prediction [Duda 00].

Support Vector Machine

NB and k-NN classifier are based on Bayes decision theory. Another group of classifiers, e.g. the SVM, directly determines the discriminant functions [Cort 95]. In the case of a SVM, the discriminant function $g(\vec{x})$ is defined by

$$g(\vec{x}) = \sum_{i=1}^{N_S} \alpha_i y_i K(\vec{s}_i, \vec{x}) + b. \tag{2.20}$$

α_i, $K(\cdot)$, and b denote Lagrange multiplier, kernel function, and offset of the hyperplane, respectively. $\vec{s}_i$ and y_i denote N_S support vectors and the corresponding labels, respectively. In this thesis, a linear kernel was used defined by:

$$K(\vec{s}_i, \vec{x}) = \vec{s}_i \cdot \vec{x} \tag{2.21}$$

According to [Burg 98, Duda 00], the training of SVM aims at finding the hyperplane that separates training points of multiple classes with the largest margin. The support vectors are close to the decision boundary. In the training phase, a cost parameter C has to be optimized. A large C corresponds to assigning a higher penalty to errors.

According to [Sun 12], the SVM can be parallelized in order to reduce the training time. The SVM training is based on using hierarchical, partial SVMs. Each partial SVM is trained on a subset of the original complete training data and is used as a filter. The support vectors of two partial SVMs of the first level of the hierarchy are used as input of a partial SVM in the next level of the hierarchy. The process is repeated until one final SVM is trained.

Classification and Regression Trees

Compared to the previously mentioned techniques, non-metric methods can further be used for classification, e.g. decision trees [Duda 00]. In 2 - D, tree - based methods partition the feature space into a set of rectangles and fit a model [Hast 09]. An example of a tree-based method is CART [Brei 84]. CART constructs a tree which consists of several nodes. At each node a decision is made resulting in a binary split of the training data [Duda 00]. The decision is based on a criterion measuring the node impurity, e.g. the Gini diversity index [Hast 09]. The Gini diversity index measures the total variance across classes. A pruning of the tree is further often performed. In pruning, the whole tree is constructed until the leaf nodes have minimum impurity [Duda 00]. Subsequently, neighboring leaf nodes are investigated regarding possible elimination depending on the increase of the node impurity. The assignment of the class label is performed in the leaf nodes by majority voting [Duda 00]. Decision trees like CART often suffer from high variance [Hast 09, Jame 13].

Random Forest

The reduction of the previously mentioned variance can be performed by bootstrap aggregation (bagging) [Duda 00]. B separate randomly chosen bootstrap samples are drawn from the training data with replacement. The size of each bootstrap sample is equal to the size of the original training set. Each of the B bootstrap samples is used to train a separate classifier. Subsequently, majority voting over the B predictions is applied to determine the final prediction. Further variance reduction is achieved by the Random Forest classifier. A Random Forest classifier includes a decorrelation of the trees by randomly selecting features in the tree-growing process [Brei 01, Hast 09]. Each time a split in a certain tree is performed, a random sample of m features is chosen as split candidates from the complete set of features. Only one of these m features is selected.

2.4.7 Multiple Linear Regression

In classification problems, input data is assigned to a discrete value [Bish 06]. If the assigned value is continuous, a regression problem has to be solved. According to [Weis 05], MLR is the most commonly used type of regression and was applied in this work to estimate the expended energy and the ball speed. A MLR model $f(\vec{x})$ with n predictors $\vec{x} \in \mathbb{R}^{n+1}$ can be defined as:

$$\begin{aligned} f(\vec{x}) &= \beta_0 x_0 + \beta_1 x_1 + ... + \beta_n x_n = \begin{bmatrix} x_0 & x_1 & ... & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \\ &= \vec{x}^T \vec{\beta} \end{aligned}$$

$\vec{\beta} \in \mathbb{R}^{n+1}$ denotes the parameter vector of the model. From the pool of predictors, a set of terms can be created including the intercept/constant ($x_0 = 1$), one of the predictors (e.g. x_1), polynomials (e.g. x_1^2), or interactions (e.g. $x_1 x_2$) [Weis 05]. Ordinary least squares estimation is applied to obtain $\vec{\beta}$ based on observations $\{\vec{x}_i, y_i \in \mathbb{R}\}|_{i=0,\dots,N-1}$. N denotes the number of observations.

2.4.8 Performance Assessment

In order to evaluate the applicability of the previously mentioned classification and regression techniques, a performance assessment is required [Hast 09].

Leave-One-Subject-Out Cross-Validation

Performance assessment of classification and regression techniques are based on a splitting of the available datasets. One commonly used technique is Leave-One-Subject-Out Cross-Validation (LOSO-CV), which is a special case of the leave-one-out method [Lach 68]. Assuming a finite set of N samples are available for evaluation [Theo 09]. In the leave-one-out method, the training is performed using $N-1$ samples, and the validation or testing is performed using the excluded sample. This procedure is performed N times, each time excluding a different sample. In each trial, one performance measure is computed. The final performance of the system is determined by averaging the performance values achieved in each trial. In the case of LOSO-CV, samples of one certain subject is taken for validation or testing in each leave-one-subject-out trial, the other samples are chosen for training.

Confusion Matrix

Classification-specific performance measures are often based on the confusion matrix [Fawc 06]. The confusion matrix represents the dispositions of the set of instances, on which a classifier is tested. An example is given in Table 2.2. Each entry represents the number of classified instances. Columns indicate the true class. Rows indicate the predicted class.

Multi-Class Classification

The confusion matrix (Table 2.2) is used to compute various performance measures for multi-class classification problems [Soko 09]. Performance measures for a specific class k include sensitivity ($Sens_k$), precision ($Prec_k$), and F1-score ($F1_k$). In additional, the balanced accuracy ($BalAcc$) is often given [Brod 10]:

$$Sens_k = \frac{C_{k,k}}{\sum_{i=1}^{K} C_{i,k}} \quad (2.22)$$

$$Prec_k = \frac{C_{k,k}}{\sum_{j=1}^{K} C_{k,j}} \quad (2.23)$$

$$F1_k = 2 \cdot \frac{Prec_k \cdot Sens_k}{Prec_k + Sens_k} \quad (2.24)$$

$$BalAcc = \frac{1}{K} \sum_{k=1}^{K} Sens_k. \quad (2.25)$$

Binary Classification and ROC Curve

Performances measures of a binary classification ($K = 2$) are often applied in medial decision making [Fawc 06]. Each instance is mapped to either the positive class or the negative class. The elements in the confusion matrix (Table 2.2) are usually denoted as **true positives** ($TP = C_{1,1}$), **false positives** ($FP = C_{1,2}$), **false negatives** ($FN = C_{2,1}$), and **true negatives** ($TN = C_{2,2}$).

Receiver Operating Characteristic (ROC) graphs visualize, organize, and select classifiers based on the corresponding performance [Fawc 06].

Table 2.2.: Confusion matrix. Each entry $C_{i,j}|_{i,j \in 1,...,K}$ represents the number of classified instances regarding K classes. Columns indicate the true class. Rows indicate the predicted class.

	True class			
Predicted class	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,K}$
	$C_{2,1}$	$C_{2,2}$	$\cdots$	$C_{2,K}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$C_{K,1}$	$C_{K,2}$	$\cdots$	$C_{K,K}$

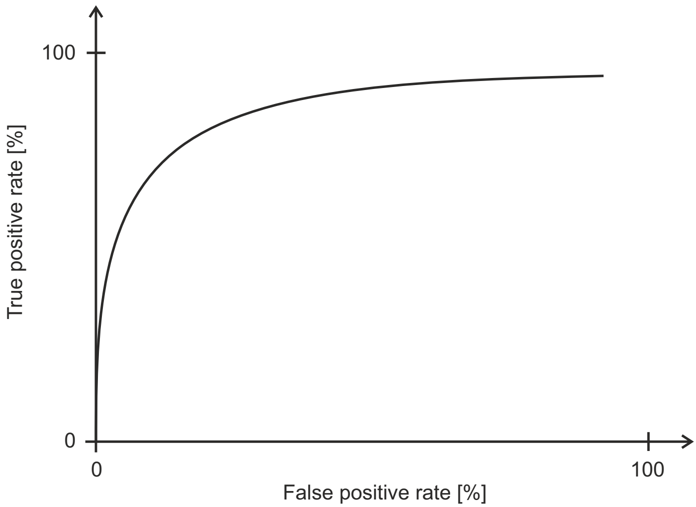


Figure 2.4.: Receiver Operating Characteristic (ROC) curve. True positive rate is plotted over false positive rate. One point on the curve depicts the result of a classifier with a specific parameter configuration.

In a ROC graph, true positive rate (TPR) is plotted over false positive rate (FPR):

$$TPR = \frac{TP}{TP + FN} \tag{2.26}$$

$$FPR = \frac{FP}{FP + TN}. \tag{2.27}$$

The ROC graph shows the relative tradeoff between benefits (TP) and costs (FP). A ROC graph can further visualize the performance of one classifier dependent on different parameter settings, which results in a ROC curve (Figure 2.4).

Regression-Specific Performance Measures

Performance assessment of regression techniques is often based on the residual e_i which is the difference between the actual value y_i and the fitted value. The corresponding performance measures include MAE, MAPE, and RMSE [Hynd 06, Mont 12] defined as

$$MAE = \frac{1}{N} \sum_{i=0}^{N-1} |e_i| \tag{2.28}$$

$$MAPE = \frac{100}{N} \sum_{i=0}^{N-1} \left| \frac{e_i}{y_i} \right| \tag{2.29}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} e_i^2}. \tag{2.30}$$

Graphical techniques such as the Bland-Altman plot are further used in the evaluation of regression algorithms [Blan 86, Blan 99]. The Bland-Altman plot illustrates the difference between the measurements of two methods against their mean. An example is given in Figure 2.5. If the differences of the measurements are normally distributed, it is expected that 95 % of the differences lie between two borders, which are dependent on the standard deviation of the difference. The Bland-Altman plot

- Allows the investigation of any possible relationship between the discrepancies and the true value. Since the true value of the quantity is not known, the mean of the measurements by the two methods is used as the best estimate.
- Shows any extreme or outlying observations.
- Can visualize e.g. an increase in variability of the differences as the magnitude of the measurements increases. A special case would be, if the mean difference is approximately proportional to the magnitude of the measurement.

Boxplots

According to [Frig 89], a boxplot enables a summary of data by displaying various main features. An example is given in Figure 2.6. The two ends of the box are determined by the lower and upper quartile $Q1$ and $Q3$, respectively. The interquartile range is computed by $Q3 - Q1$. Two fences lie at $Q1 - k(Q3 - Q1)$ and $Q3 + k(Q3 - Q1)$.

k was set to 1.5 as proposed by [Frig 89]. Points which fall outside the fences are potential outliers. A horizontal line indicating the median is

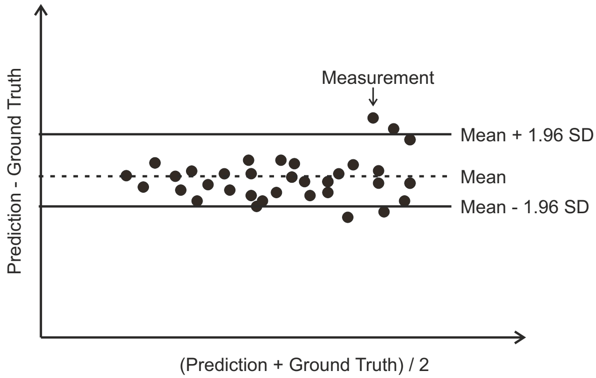


Figure 2.5.: Bland Altman plot (adapted from [Blan 99]). Difference of prediction and ground truth are plotted regarding the average of both. Mean of difference and 95 % limits of agreement are further given. SD denotes standard deviation.

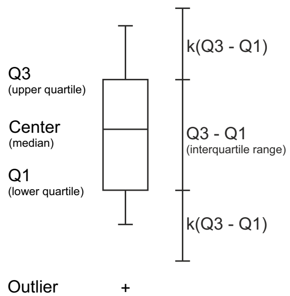


Figure 2.6.: Boxplot (adapted from [Frig 89]). $Q1$ and $Q3$ are denoted as lower and upper quartile, respectively. Horizonal line in center of box indicates median. Outlier, interquartile range $Q3 - Q1$, and two fences $Q1 - k(Q3 - Q1)$ as well as $Q3 + k(Q3 - Q1)$ are further given. k is usually set to 1.5.

further given in the center of the box. In this thesis, the boxplot was used to show the variability of classification-specific and regression-specific performance measures regarding different LOSO-CV trials.

Chapter 3

Recognition of Daily Activities

3.1 Introduction

In this chapter, five contributions regarding the recognition of daily activities are presented. The structure of the chapter is as follows. Section 3.2 describes two studies which were necessary to collect IMU data during various daily activities for training and testing machine learning algorithms. Data of the first study were used to classify daily life activities based on the activity type itself. The study was published in [Schu 13b]. Data of the second study were used to train and test an energy expenditure estimation algorithm. The study was performed during the Bachelor Thesis of Sabrina Dorn [Dorn 13] (supervised by the author of this dissertation) and was published in [Schu 14a]. Both studies are further available on https://www.activitynet.org. Section 3.3 and 3.4 introduce two algorithms for HAR based on hierarchical classification (contribution 1) and decision level fusion (contribution 2), respectively. The first algorithm was published in a journal article [Schu 13b]. The first authorship was shared with Heike Leutheuser. The method was developed by both authors and mainly implemented by the author of this dissertation. The second algorithm was published in a conference article [Schu 13a]. Section 3.5 compares the hierarchical classification and the decision level fusion approach to five state-of-the-art techniques from literature based on the same benchmark dataset (contribution 3). The idea of comparing various HAR algorithms based on the same benchmark dataset was further published in [Schu 13b]. In section 3.6, a database fusion strategy is introduced in order to increase the amount of training data for machine learning techniques (contribution 4). The method was

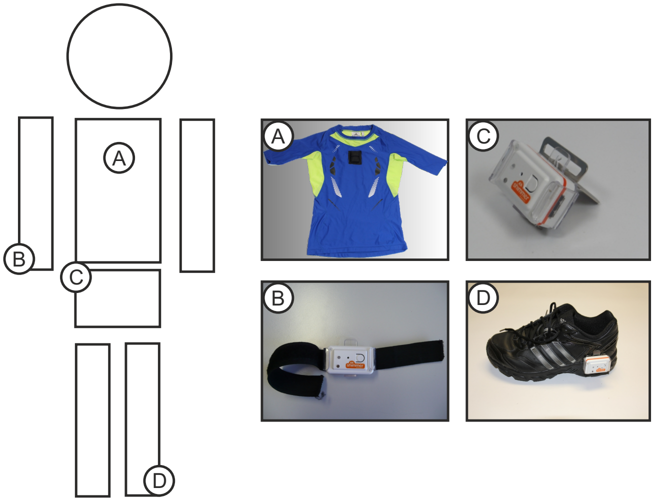


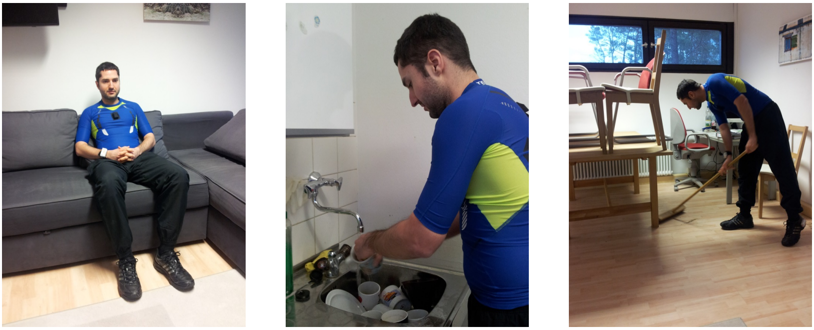
Figure 3.1.: Sensor placement used in DaLiAc study (adapted from [Schu 13b]). Left: frontal view of human and sensor positions (A: chest, B: wrist, C: hip, D: shoe); right: equipment for sensor mounting.

published in [Schu 14b]. The chapter concludes with the description of a HAR technique based on the expended energy instead of the activity type (section 3.7), contribution 5). The algorithm was published in [Schu 14a].

3.2 Data Collection

3.2.1 Study A: Daily Life Activities (DaLiAc)

The following study was called DaLiAc (**Dai**ly **Li**fe **Ac**tivities) and the corresponding data can be found on https://www.activitynet.org. In the data collection, 23 healthy subjects (10 female and 13 male) participated (age 27 ± 7 years, body mass index $24.0kg/m^2 \pm 3.5kg/m^2$, mean ± standard deviation). All subjects gave written informed consent about their participation. The Research Ethics Committee of the Friedrich-Alexander-Universität Erlangen-Nürnberg confirmed that there is no necessity to obtain the approval of the local Ethics Committee. Each subject was equipped with four SHIMMER sensor nodes (Shimmer Research, Dublin, Ireland) [Burn 10]. The sensor placement is shown in Figure 3.1.



(a) Sitting **(b)** Washing dishes **(c)** Sweeping

Figure 3.2.: Illustration of activities in DaLiac database.

One sensor was integrated in a customized shirt which ensured tight fit. A sensor on the chest was e.g. used in [Park 06]. One sensor was placed on the right wrist by a band. The reason for choosing the wrist was to cover activities, which involve movements of the upper extremities motivated by [Bao 04, Park 06, Erme 08, Liu 12]. One sensor was attached to the right hip by a clip. A sensor close to the center of mass of the body, e.g. hip, was often preferred in literature [Bout 97a, Clel 13]. One sensor was mounted on the lateral side of a regular sport shoe by a custom designed clip. The sensor position was chosen in order to cover activities, which involve movements of the lower extremities motivated by [Bao 04, Pree 09]. The used SHIMMER sensor node had a module size of 53 x 32 x 15 mm and comprised a triaxial accelerometer (MMA7361L, Freescale Semiconductors, Austin, TX, USA), a triaxial gyroscope (500 or 2000 series, InvenSense, Sunnydale, CA, USA), and a microcontroller (MSP430F1611, Texas Instruments, Dallas, TX, USA). The accelerometer range was set to $\pm$ 6 g. The gyroscope range was set to $\pm$ 500 $^\circ/s$ for the sensor nodes wrist, chest, and hip as well as $\pm$ 2000 $^\circ/s$ for the sensor node on the shoe. The reason for the higher gyroscope range for the shoe sensor was the assumption that for lower extremities higher angular velocities were expected [Bhat 80, Capp 82]. The sampling rate was set to 204.8 Hz in order to capture frequencies up to 20 Hz [Bout 97a]. The sensor data was stored on a MicroSD card.

Table 3.1.: List of 13 activities available in DaLiAc database, corresponding abbreviations, and associated MET values taken from the 'Compendium of Physical Activities' [Ains 93, Ains 00, Ains 11].

Activity	Abbreviation	Associated MET
Sitting	SI	1.3
Lying	LY	1.0
Standing	ST	1.3
Washing dishes	WD	2.5
Vacuuming	VC	3.3
Sweeping	SW	3.3
Walking	WK	3.5
Ascending stairs	AS	5.0
Descending stairs	DS	3.5
Running on treadmill	RU	9.0
Bicycling (50 watt)	BC50	3.5
Bicycling (100 watt)	BC100	6.8
Rope jumping	RJ	8.8

The subjects had to perform 13 daily life activities ranging from sedentary to vigorous-intensity activities. The selected activities were taken from the 'Compendium of Physical Activities' [Ains 93, Ains 00, Ains 11] and are listed in Table 3.1 together with, the used abbreviation, and the associated MET values. The single steps of the protocol are described below.

1. Sensor nodes were put on a plate, dropped down twice, and, in between, a sinusoidal movement was performed. This procedure was needed for offline synchronization.

2. Sensor nodes were placed on dedicated positions (see Figure 3.1).

3. Static activities were performed for one minute (**sitting**, **lying**, **standing**).

4. Household activities were performed (**washing dishes**, **vacuuming**, **sweeping**). Since the wrist sensor was placed on the right wrist, the subject was told to use the vacuum with the right hand to capture the signal corresponding to the hand movement. Washing dishes was performed for two minutes. Vacuuming was performed

for one minute. Sweeping had to be performed until a pre-defined parkour was passed.

5. **Walking** was performed outside. The subjects had to walk on the university campus from one building to another building (about 250 m).

6. **Ascending stairs** (until third floor) and **descending stairs** (back to main floor) were performed in the building.

7. **Walking** was performed outside.

8. **Running on treadmill** (h/p/cosmos quasar, h/p/cosmos & medical gmbh, Nussdorf-Traunstein, Germany) was performed indoor. The treadmill speed was set to 8.3 km/h. The activity was performed for two minutes.

9. **Bicycling** on an ergometer (sanabike 250 F, MESA Medizintechnik GmbH, Benediktbeuern, Germany) was performed indoor. Two different resistance levels (**50 W** and **100 W**) were adjusted. The subjects were told to perform the bicycling with 70 revolutions per minute. Bicycling with each resistance level was performed for two minutes.

10. **Rope jumping** was performed with five trials and at least five jumps per trial.

11. Sensor nodes were taken from the dedicated position.

Figure 3.2 illustrates the activities sitting, washing dishes, and sweeping. A study supervisor accompanied the subjects and labeled the start and end of each activity via smartphone app.

3.2.2 Study B: Energy Expenditure (EnEx)

The following study was called EnEx (**En**ergy **Ex**penditure) and the corresponding data can be found on https://www.activitynet.org. In the data collection, ten healthy subjects (seven male and three female) participated (age 49 $\pm$ 12 years, height 178 $\pm$ 10 cm, weight 80.7 $\pm$ 14.6 kg, mean $\pm$ standard deviation). All subjects gave written informed consent about their participation. Approval from the ethical committee was received (Re.-No. 181 12B, 24.07.2012, Medical Faculty, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany). The sensor setup,

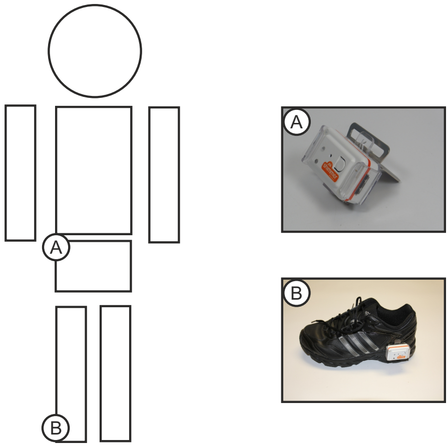


Figure 3.3.: Sensor placement used in EnEx study (adapted from [Schu 14a]). Left: frontal view of human and sensor position (A: hip mid-axilliary line, B: exterior side of shoe); right: equipment for sensor mounting.

protocol, and ground truth system were chosen according to the findings mentioned in section 1.5.2 and are described in the following. Each subject was equipped with two SHIMMER sensor nodes (Shimmer Research, Dublin, Ireland) [Burn 10]. Details of the sensor node were already given in the previous section. The sensor placement is shown in Figure 3.3. The two sensors were attached to the hip mid-axilliary line and to the exterior side of a shoe. Attaching a sensor at the hip enabled the measurement of human body movement near the center of mass, which is often done in research [Chen 97, Roth 07, Mont 16]. Since only walking activities were acquired, one sensor was further attached to the lower extremities motivated by [Godf 08]. The accelerometer range was set to $\pm$ 1.5 g for the sensor node hip and $\pm$ 6 g for the sensor node on the shoe. A higher g-range was chosen for the shoe sensor because of the assumed magnitude increase from cranial to caudal body parts [Bout 97a]. The gyroscope range was set to $\pm$ 500 $^\circ/s$ for the sensor node on the hip and $\pm$ 2000 $^\circ/s$ for the sensor node on the shoe for 70 % of the subjects and $\pm$ 500 $^\circ/s$ for 30 % of the subjects. The different gyroscope settings resulted from sensor problems. The reason for the higher gyroscope range on the shoe was the assumption that for lower extremities higher angular velocities were expected similar to the accelerometer [Bout 97a]. The sampling rate was set to 204.8 Hz in order to ensure capturing the frequencies of human movement up to 20 Hz motivated by [Bout 97a].

The subjects had to perform three speed levels ([3.2, 4.8, 6.4] km/h) on both a normal and an oscillating treadmill (h/p/cosmos model mercury med 5.0, Traunstein, Germany). The speed levels were motivated by [Leen 03, Motl 09]. The reason for using an oscillating treadmill was to impose different levels of physical activity. An inclination of 0 % was adjusted. The three walking speed levels were performed for six minutes each in an ascending order. The order of normal and oscillating treadmill was randomized by throwing a dice. Between the two trials of walking, ten minutes of quiet seated rest was performed in order to return to baseline level of energy expenditure [Leen 03].

In order to collect ground truth data, the oxygen consumption $\dot{V}O_2$ was collected by breath-by-breath analysis using the Master Screen® CPX (Viasys Healthcare GmbH, Höchberg, Germany). The expended energy, expressed in MET was determined using equation 2.3. The sampling rate was set to 0.2 Hz. The system was calibrated before the data acquisition according to the manufacturer instructions. The single steps of the protocol are described below.

1. Subjects got familiar with walking on the treadmill.
2. Subjects were equipped with gas mask over mouth and nose.
3. Sensor nodes were put on a plate, dropped down twice, and, in between, a sinusoidal movement was performed. This procedure was needed for offline synchronization.
4. Sensor nodes were placed on dedicated positions (see Figure 3.3).
5. Standing for 1 min.
6. Three speed levels ([3.2, 4.8, 6.4] km/h) were performed for six minutes each (condition normal or oscillating dependent on randomization decision).
7. Resting period of 10 minutes.
8. Three speed levels ([3.2, 4.8, 6.4] km/h) were performed for six minutes each (condition normal or oscillating dependent on randomization decision).
9. Sensor nodes were taken from the dedicated position.

3.3 Hierarchical Classification System

3.3.1 Overview

Algorithms for current HAR systems mainly provided a flat classification architecture, which applied one multi-class system in order to recognize various activities [Bao 04]. Two disadvantages could be found. First, a retraining of the complete system is often necessary, if a new activity or sensor has to be integrated. Second, the activity prediction is based on a single decision of one classifier. Hierarchical architectures are often more suitable to exploit the hierarchical structure of activities [Math 04a]. Nevertheless, current hierarchical solutions are often based on simple thresholding techniques or a manual design of components of the ARC. Thus, the adjustment of existing approaches to a new environment, e.g. new activities or sensors, would be cumbersome and time-consuming. The purpose of this section is to introduce an algorithm which provides a hierarchical, multi-sensor based approach for HAR which is flexible in adding new elements like activities and sensors. The final activity prediction of the proposed method is further based on multiple decisions along the path of the hierarchy. The algorithm was originally published in [Schu 13b] and was modified in this work. The modified parts are highlighted in the discussion.

This section is structured as follows. First, the hierarchical algorithm is introduced. Second, the experimental setup is described. Third, the results of the experiments are presented. Fourth, the results are finally discussed.

3.3.2 Methods

The proposed HAR system is depicted in Figure 3.4 and was optimized for the activities available in the DaLiAc database (section 3.2). The circles indicate the 13 activities (Table 3.1). The ellipses indicate merged activities ordered in a class hierarchy. The rectangles indicate classification subsystems. The name of the each subsystem is the name of the corresponding merged activity which is further split. The class hierarchy was motivated by the activity clusters introduced in the 'Compendium of Physical Activities' [Ains 93]. The compendium provided the following clusters for the 13 selected activities:

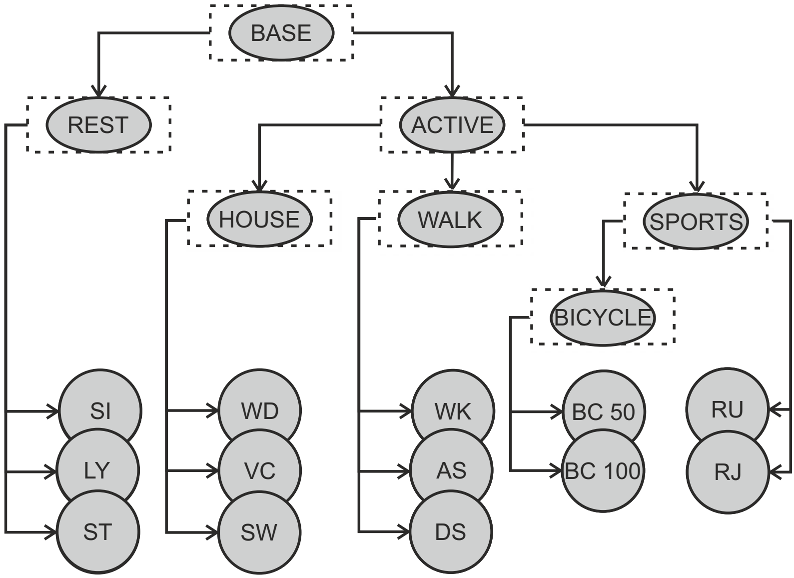


Figure 3.4.: Pipeline of hierarchical classification system (adapted from [Schu13b]). Architecture is motivated by Figure 1.3b. Circles and ellipses indicate classes of 13 final activities (see Table 3.1) and seven merged activities, respectively. Rectangles indicate classification subsystems BASE, REST, ACTIVE, HOUSE, WALK, SPORTS, and BICYCLE.

- Inactivity: sitting, lying, standing
- Home activities: washing dishes, vacuuming, sweeping
- Walking: walking (level), ascending stairs, descending stairs
- Conditioning exercises: bicycling 50 watt/100 watt, rope jumping
- Running: running 8.3 km/h

In total, seven subsystems were defined: BASE, REST, ACTIVE, HOUSE, WALK, SPORTS, and BICYCLE. In each subsystem, a different classification problem was solved. For the BASE subsystem, the rest activities (sitting, lying, standing) were combined in a first class and the other activities in the database were combined in a second class. The BASE system discriminated between these two classes. The early distinction between static and dynamic activities was motivated by [Velt 93, Fahr 97, Kara 06] The REST subsystem discriminated between the single activities sitting, lying, and standing. This subsystem offered the discrimination of different body postures motivated by [Fahr 97]. For the ACTIVE subsystem, activities were grouped according to their appearance in daily life. The household activities (washing dishes, vacuuming, sweeping) were combined in a first class, the locomotion activities (walking, ascending stairs, descending stairs) were combined in a second class, and the sports activities (running, rope jumping, bicycling with 50 watt and 100 watt) were combined in a third class. The ACTIVE subsystem discriminated between these three classes. The HOUSE subsystem discriminated between the single activities washing dishes, vacuuming, and sweeping. The WALK subsystem discriminated between the single activities walking, ascending stairs, and descending stairs. The SPORTS subsystem discriminated between the single activities running, rope jumping, and the combined bicycling activities (bicycling with 50 watt and 100 watt). The BICYCLE subsystem discriminated between biycling with 50 and 100 watt. Each of the seven classification subsystems implemented an ARC (Figure 1.1) consisting of preprocessing, segmentation, feature extraction, and classification. Details of the previously mentioned steps are given in the following sections. The single subsystems were optimized based on the corresponding considered activity set. Unseen sensor data were consecutively processed by the optimized subsystems in the hierarchy (Figure 3.4).

Preprocessing

Two preprocessing steps were applied for each of the four sensors separately, namely calibration and synchronization. The calibration of the IMUs was performed according to the manufacturer's instructions. For the accelerometer, the sensitivity and offset were computed separately for each of the three axes. The sensitivity was estimated by determining the ADC (Analog-to-Digital Converter) values representing $\pm$ 1 g and computing the mean distance of both values. The offset was estimated by dividing the sensitivity by two. The mapping from ADC values to g was performed by a linear function considering the two estimated parameters. For the gyroscope, the offset was computed separately for each of the three axes by determining the ADC values representing 0 $^\circ/s$. The mapping from ADC values to $^\circ/s$ was performed by a linear function considering the offset. In order to synchronize the four sensor nodes, the

Table 3.2.: Number of available instances per activity.

Activity	Number of instances
Sitting	431
Lying	435
Standing	430
Washing dishes	913
Vacuuming	432
Sweeping	715
Walking	2015
Ascending stairs	292
Descending stairs	249
Running on treadmill	886
Bicycling (50 watt)	897
Bicycling (100 watt)	897
Rope jumping	236

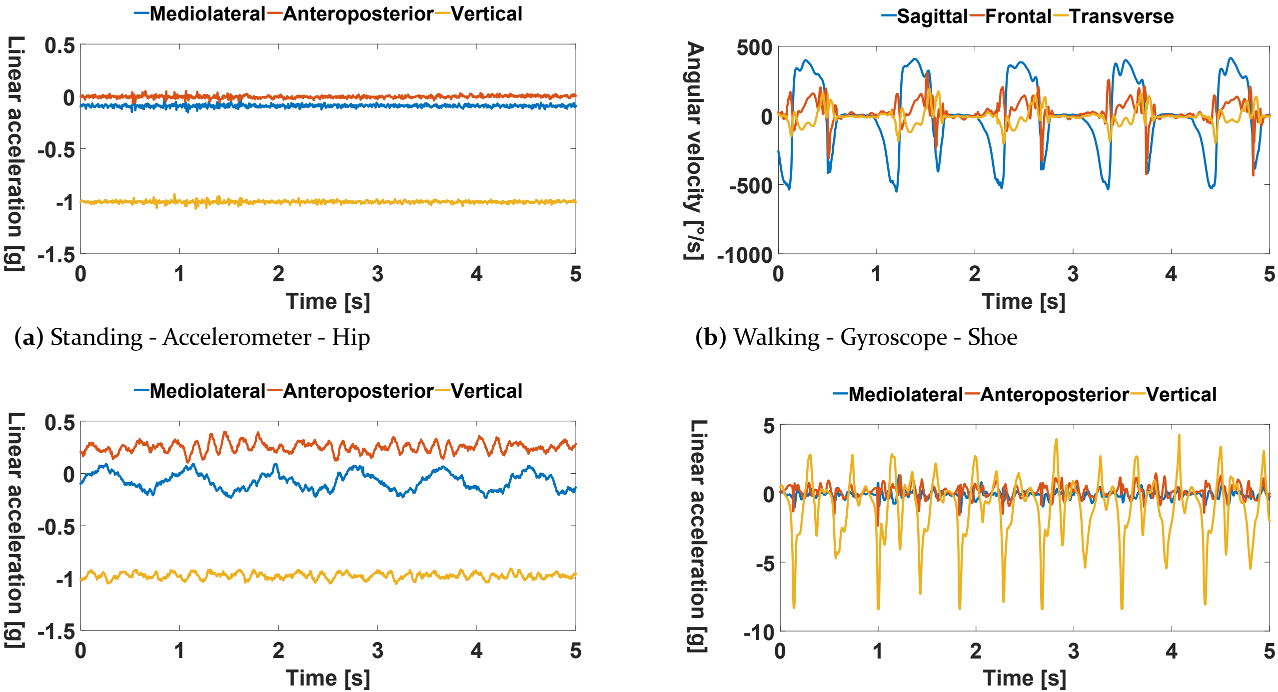
first up-down movement of the linear acceleration in vertical direction was manually selected for all four sensor nodes. The corresponding time point was defined as common start point. This procedure was motivated by [Bao 04].

Segmentation

In the segmentation step, the inertial sensor data of the 13 activities were divided in 5 s windows with 50 % overlap (sliding window method, see section 2.4). The selection of the window parameters was motivated by [Bao 04, Ravi 05, Pree 09]. Figure 3.5 shows segmentation examples for standing, walking, bicycling with 100 watt, and rope jumping.

Feature Extraction and Classification

In each sliding window, features were extracted based on the corresponding sensor data. The feature set listed in Table 2.1 was computed for each sensor type of each sensor node (accelerometer of hip, gyroscope of hip, accelerometer of chest, gyroscope of chest, ...). The features of each sensor source was fused at feature level (section 1.5.1, Figure 1.2b). In total, 224 features were extracted for each window.



(a) Standing - Accelerometer - Hip

(b) Walking - Gyroscope - Shoe

(c) Bicycling (100 watt) - Accelerometer - Chest

(d) Rope jumping - Accelerometer - Chest

Figure 3.5.: Segmented sensor data of four selected activities standing (a), walking (b), bicycling with 100 watt (c), and rope jumping (d). The considered sensor type and sensor position are further given.

In order to reduce the number of features, the correlation-based feature selection approach was applied, which is fast, simple, and was successfully used in [Hall 99b, Chow 18]. The correlation-based feature selection was described in section 2.4.5. The features were used as input for a classifier, in order to map a given window to one activity. The investigated classifiers are mentioned later in this chapter.

3.3.3 Experiments and Evaluation

The proposed hierarchical classification system for HAR was evaluated on the DaLiAc database (section 3.2.1). The datasets of four subjects had to be excluded from further processing due to hardware issues. The number of available instances for each of the 13 activities is listed in Table 3.2. Two experiments were conducted. First, the seven subsystems were compared. Second, the complete hierarchical classification system was evaluated. Details about the experiments are given below.

- **Experiment 1: Seven subsystems**
  The performance of the seven subsystems was investigated regarding classifier performance and feature selection. The five classifiers NB, SVM, k-NN, CART, and RF were applied to the selected features. Details about the classifiers can be found in section 2.4.6. The balanced accuracy was used as performance measure, which was determined by LOSO-CV (Equation 2.25). The parallelized version of the SVM was used as described in section 2.4. In the training phase of the SVM, the available datasets were randomly split in the first level of the SVM hierarchy in four disjunct subject partitions. The datasets of each partition were used as input to train one partial SVM. In each partial SVM, the linear kernel was used and the corresponding cost parameter C was optimized by grid search ($C \in \{2^N\}, N \in \{-10, ..., 10\}$). In the grid search, 70 % of the instances were randomly chosen for training and 30 % for validation. The number of neighbors k in the k-NN classifier was optimized by grid search ($k \in \{3, 5, 7\}$). The number of trees in the RF classifier was optimized by grid search ($n_{tree} \in \{5, 10, 15, 20\}$). The parameter optimization for k-NN and RF was performed in an inner LOSO-CV.

  The results of the correlation-based feature selection were further investigated. Two quantitative performance measures were selected. First, the number of selected features per subsystem was determined regarding the LOSO-CV trials (mean ±standard deviation). Second, the distribution of the selected features regarding the sensor positions wrist, chest, hip, and shoe as well as the sensor types accelerometer and gyroscope was computed for each subsystem. Therefor, the selected features of each trial of the LOSO-CV were grouped according the previously mentioned sensor positions as well as types and were summed up. In order to determine the distribution, the corresponding sums were divided by the total number of selected features.

- **Experiment 2: Complete hierarchical classification system** The performance of the complete hierarchical classification system was investigated regarding confusion matrix, activity-specific sensitivity, and balanced accuracy (section 2.4.8). Therefor, two steps were performed. First, the classifier, which achieved the highest balanced accuracy in the first experiment, was selected for each subsystem. Second, instances were processed by the hierarchical classification system and the previously mentioned performance measures were determined. The second step was performed in a LOSO-CV.

3.3.4 Results

The balanced accuracy values of the classifiers CART, SVM, NB, RF, and k-NN regarding the subsystems BASE, REST, ACTIVE, HOUSE, WALK, SPORTS, and BICYCLE are comprised in Table 3.3. RF was determined as best classifier for the BASE subsystem with a balanced accuracy of 97.3 %. k-NN was determined as best classifier for the subsystems REST, ACTIVE, HOUSE, and WALK with balanced accuracy values above 91.9 %. SVM and NB achieved the best performance for the SPORTS subsystem. NB achieved the best balanced accuracy of 58.5 % for the BICYCLE subsystem. Table 3.4 comprises the number of selected features after the correlation-based features selection. Mean and standard deviation are given for each subsystem. The feature selection routine determined in average 2.8 to 39.2 features depending on the subsystem. In Figure 3.6, the distribution of the selected features regarding sensor positions and types is depicted.

Figure 3.7 depicts mean and standard deviation of the sensitivities for all 13 activities. The confusion matrix of the hierarchical system is shown in Table 3.5. The proposed system achieved an overall balanced accuracy of 89.1 %.

Table 3.3.: Balanced accuracy [%] of CART, SVM, NB, RF, and k-NN regarding subsystems BASE, REST, ACTIVE, HOUSE, WALK, SPORTS, and BICYCLE. The best classifier per subsystem is marked in bold.

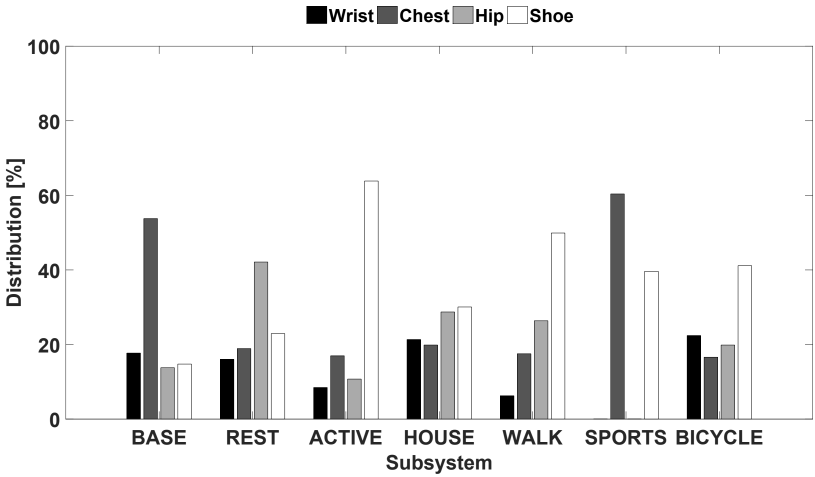
	CART	SVM	NB	RF	kNN
BASE	96.6	91.0	94.8	**97.3**	97.2
REST	93.7	95.6	96.4	96.6	**96.7**
ACTIVE	98.1	98.6	97.6	99.2	**99.7**
HOUSE	84.9	89.8	90.2	91.3	**91.9**
WALK	96.4	99.4	98.5	99.2	**99.5**
SPORTS	96.3	**98.5**	**98.5**	96.3	97.7
BICYCLE	50.6	51.3	**58.5**	55.8	55.2

Table 3.4.: Number of selected features for each subsystem after correlation-based feature selection [mean ± standard deviation].

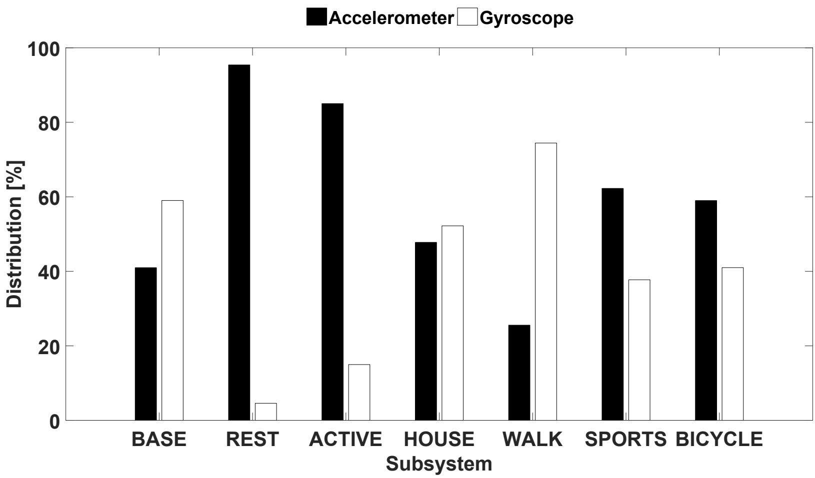
Subsystem	Number of selected features
BASE	16.1 ± 3.0
REST	18.4 ± 3.4
ACTIVE	18.6 ± 2.8
HOUSE	39.2 ± 4.5
WALK	26.2 ± 3.6
SPORTS	2.8 ± 1.0
BICYCLE	37.1 ± 3.5

Table 3.5.: Confusion matrix of hierarchical HAR system. Each entry represents the number of classified instances. Columns indicate actual class. Rows indicate predicted class. Abbreviations of the 13 activities are given in Table 3.1.

	SI	**LY**	**ST**	**WD**	**VC**	**SW**	**WK**	**AS**	**DS**	**RU**	**BC50**	**BC100**	**RJ**
SI	**396**	0	12	1	0	0	0	0	0	0	0	0	0
LY	0	**435**	0	0	0	0	0	0	0	0	0	0	0
ST	7	0	**389**	66	2	3	0	0	0	0	0	0	0
WD	23	0	25	**844**	5	5	0	0	0	0	0	0	0
VC	2	0	4	0	**378**	84	9	2	0	0	0	0	0
SW	3	0	0	2	47	**618**	1	5	0	0	0	0	0
WK	0	0	0	0	0	0	**1998**	0	2	0	0	0	0
AS	0	0	0	0	0	1	4	**285**	0	0	0	0	0
DS	0	0	0	0	0	4	3	0	**247**	0	0	0	0
RU	0	0	0	0	0	0	0	0	0	**880**	0	6	16
BC50	0	0	0	0	0	0	0	0	0	0	**574**	425	0
BC100	0	0	0	0	0	0	0	0	0	0	323	**466**	0
RJ	0	0	0	0	0	0	0	0	0	6	0	0	**220**



(a) Sensor position



(b) Sensor type

Figure 3.6.: Distribution [%] of selected features after correlation-based feature selection for the subsystems BASE, REST, ACTIVE, HOUSE, WALK, SPORTS, and BICYCLE. Distributions are grouped according to sensor positions (a) and type (b).

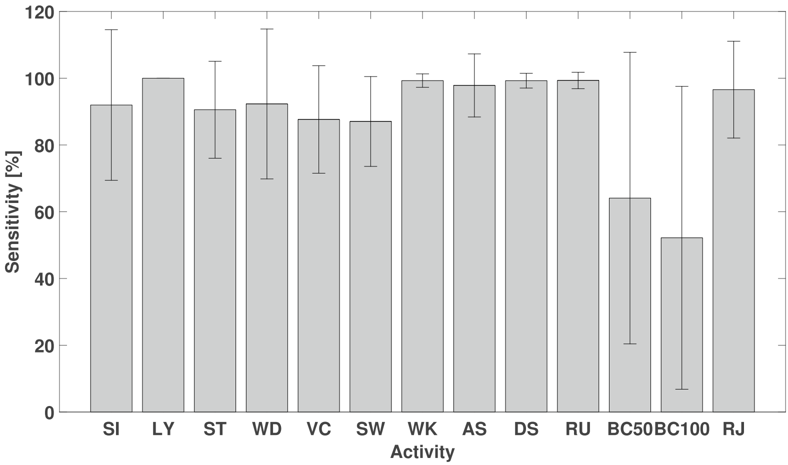


Figure 3.7.: Sensitivities of 13 activities. Mean and standard deviation are depicted. Abbreviations of 13 activities are given in Table 3.1.

3.3.5 Discussion

In this work, a hierarchical classification algorithm for HAR was developed, implemented, and evaluated. The achieved results are discussed in this section. The discussion is structured as follows. First, the modifications of the originally proposed method published in [Schu 13b] are described and motivated. Second, the performance of the seven subsystems are compared (experiment 1). Third, the performance of the whole hierarchical system is argued (experiment 2). Fourth, the proposed approach is compared to state-of-the-art approaches in literature. Fifth, examples for future work are provided.

Modifications Compared to Original Publication

Modifications of the proposed hierarchical, multi-sensor based HAR system were made compared to the original publication [Schu 13b]. All in all, the modifications slightly decreased the performance of the algorithm from 89.6 % to 89.1 %.

The depth of the hierarchy was increased from two to four which also increased the number of subsystems from five to seven. The activity groups were changed such that they better reflect the different domains in daily life, e.g. household or sports. Compared to [Schu 13b], the accuracy of the subsystems BASE, REST, and BICYCLE decreased by 0.6 %, 0.7 %, and 3.1 %, respectively. Although a slightly worse performance of the BASE and REST subsystems was observed, the proposed hierarchical architecture in this thesis allowed a better flexibility in terms of adding or removing activities to or from a certain domain. In addition, the accuracy of the HOUSE and WALK subsystems increased by 2.0 % and 1.8 %, respectively. Since identifying various household and walking activities might be more important for a human's activity profile than identifying e.g. bicycling activities, the version presented in this thesis can be seen as an improvement compared to [Schu 13b].

The configuration of the considered feature set was further modified. In [Schu 13b], the same feature set was used for all subsystems except REST. In this thesis, the number of features was increased from 152 to 224 and a correlation-based feature selection routine was applied in order to provide a subsystem-specific feature set. Instead of using expert features for the REST subsystem, as proposed in [Schu 13b], the same feature extraction and selection scheme was performed as for the other subsystems. The motivation was to reduce the burden of selecting expert features for one specific subsystem.

Performance of Subsystems

According to Table 3.3, a balanced accuracy of above 91.9 % could be achieved for all subsystems except for BICYCLE. This shows the general applicability of most subsystems to recognize a subset of activities. Since no single classifier outperformed the others, evaluating a set of classifiers and choosing the best one are mandatory. Using the same classifier for all subsystems might lower the overall performance of the system. Nevertheless, the low performance of the BICYCLE subsystem indicates that linear acceleration and angular velocity do not provide sufficient information to distinguish between different levels of bicycling. In order to further improve the system, additional sensors could be used, which measure e.g. heart rate. The architecture of the proposed hierarchical system allows to only change the BICYCLE subsystem without affecting the trained model of the remaining subsystems. It is assumed that the correlation-based feature selection routine would select features based on the integrated heart rate sensor.

Table 3.4 shows that the number of selected features differed regarding the subsystems. This might indicate the different complexities of the individual classification tasks. It seemed to be easier to distinguish between the three different merged activities in the SPORTS subsystem than to classify the three different household activities in the HOUSE subsystem. All in all, the feature selection routine could reduce the number of originally available features by at least 83 %. If the proposed activity recognition system is implemented on an embedded device, the number of features should further be reduced in order to lower the computational complexity. This could be reached by applying additional wrapper methods, which further includes the performance of the classifier in the feature selection process [Theo 09].

An advantage of applying the feature selection technique in the different subsystems is depicted in Figure 3.6. The results of the feature selection give an indication about the importance of the available sensor positions and types for the different activity sets. For the BASE subsystem, 53.8 % of the features were selected based on sensor data of the chest sensor (Figure 3.6a). In applications, which are only interested in the activity level and require a minimal number of sensors, a sensor near the chest might be sufficient. Applications, which are only interested in the posture, the REST system could be applied with a sensor on the hip, since about 42.1 % of the features were selected from sensor data acquired on the hip. Figure 3.6b indicates that for recognizing resting activities or the distinction of the three activity groups in the ACTIVE subsystem, the accelerometer sensor might be sufficient. For more complex classification problems, e.g. the distinction of the household activities, both sensor types are necessary. Information about the importance of sensor positions and types supports future HAR system engineers in the design process.

Performance of Whole Hierarchical System

The evaluation of the whole hierarchical system showed an overall acceptable performance of the system with a balanced accuracy of 89.1 %. Nevertheless, mean and standard deviation of the activity-specific sensitivities differed (Figure 3.7). Walking, stairs climbing, and running achieved standard deviations below 9 %. Thus, the proposed system can deal with the variability of different walking styles. In contrast, the higher standard deviations of washing dishes (22.5 %), vacuuming (16.1 %), and sweeping (13.5 %) across subjects showed that these activities were performed by the subjects in a different manner. Future work could integrate a subject-specific adaptation of the system as proposed by [Alle 06]. The advantage of the proposed hierarchical classification system is that such a subject-specific adaptation is only needed for certain subsystems, e.g. the HOUSE subsystem. Other subsystems like the WALK subsystem remain unchanged and does not require an adjustment.

The confusion matrix (Table 3.5) depicts the misclassification of vacuuming and sweeping, which might come from the similar execution of both activities. They include similar arm movements and a rather low intensity of the lower extremities. An additional reason of the misclassification might be the size of the window. The selected features usually describes the content of the whole window instead of certain parts inside the window. In this case, a smaller window size might be necessary, in order to only capture relevant sections of the performed activity. The hierarchical approach with individual subsystems enable the usage of different window sizes per subsystem. Future work could also investigate energy-based segmentation methods, which aim at detecting one vacuum cycle (section 2.4). The already mentioned problem of distinguishing the two resistance levels in bicycling can be also seen in Table 3.5. One disadvantage of the hierarchical structure of the algorithm can be seen in the misclassification of standing and washing dishes, which might be not intuitive at first glance. Nevertheless, the early separation of both activities in the BASE system might be one reason. Most selected features of the BASE subsystem were based on sensor data from the chest sensor (Figure 3.6a). If the upper body movement of the subjects during washing dishes was quite low, the sensor data of some segmented 5 s windows might look quite similar to standing. In order to reduce the number of misclassifications, a majority voting of the results of subsequent windows could be performed to eliminate short-term noise.

Comparison to Literature

The literature review in section 1.5.1 showed that mainly accelerometer sensors were used for HAR and the integration of the gyroscope is rather new [Amin 04, Dobk 11]. The importance of the gyroscope was shown in the proposed approach for the BASE, HOUSE, and WALK subsystem (Figure 3.6b). The advantage of the gyroscope in the WALK subsystem coincides with the findings in gait analysis research [Amin 02, McGr 12].

Most approaches in literature applied a flat classification system such as [Mant 01, Park 06, Bao 04]. Integrating a new activity or a new sensor type would require a complete retraining of the whole classification system. The proposed hierarchical classification algorithm enables a flexible integration of new activities and sensor types without affecting already trained classifiers. If 'working on a PC', 'preparing meal', or 'bicycling with 100 revolutions per minute' should be included, only the single subsystems REST, HOUSE, and BICYCLE should be re-trained, respectively. If the distinction between bicycling with 50 watt and 100 watt should be increased by using an additional heart rate sensor, only the BICYCLE subsystem has to be re-trained.

Algorithms listed in Table 1.2 mainly applied feature level fusion to combine the information of various sensor sources [Mant 01, Park 06, Liu 12]. The proposed algorithm for HAR can be seen as a combination of feature level and decision level fusion. Features of different sensor sources were merged and used as input for a certain classifier in each parent node (Figure 3.4). The subsequent classification routines along the path of the hierarchy can be seen as an additional decision level fusion.

In [Math 04a, Zhen 15], the concept of a hierarchical framework for activity recognition, which was independent regarding the activity set, was also considered. The authors in [Math 04a] applied a binary decision tree. In [Zhen 15], only one accelerometer was used and a fixed activity estimator was implemented in each level of the hierarchy. The proposed hierarchical approach, which was developed in this thesis consisted of a more generic and flexible architecture providing more than a binary split and diverse activity estimators in each level considering multiple sensors.

The achieved balanced accuracy of most subsystems was higher than the achieved performance in [Van 00, Rogg 13, Park 06, Bao 04, Ravi 05] (Table 1.2). The reason might be that grouping activities in certain levels of the hierarchy resulted in a reduction of the number of target activities. The lower number of classes might reduce the complexity of the classification problem. Nevertheless, it has to be mentioned that e.g. in the case of the merged household activities, vacuuming and sweeping might be quite similar in terms of feature distribution. Thus, the variability of features is quite low. Classifiers might benefit from this low variability. If an additional household activity with a different feature distribution is included, e.g. preparing a meal, the higher variability among the features of the household activities might negatively affect the performance of the proposed system. An additional advantage of the proposed algorithm is that a subsystem-specific window size could be chosen according to the desired activities. Currently available approaches in literature mainly used a fixed window size, sometimes optimized by grid search. This

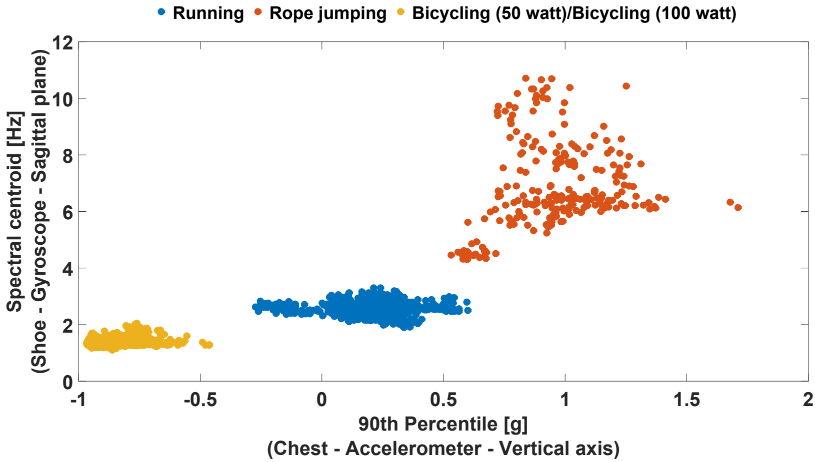


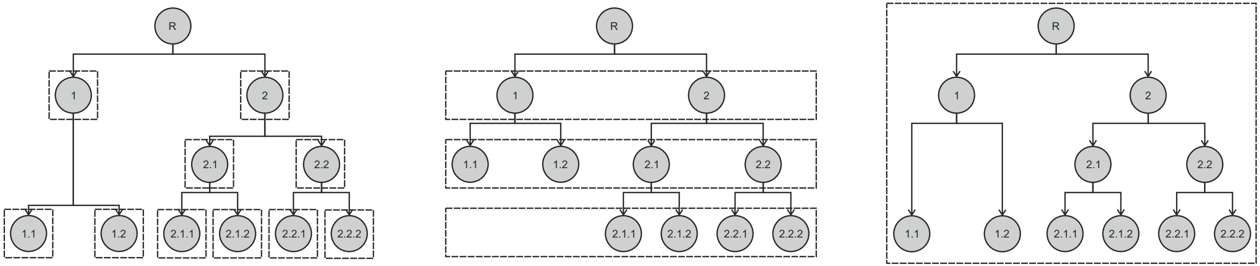
Figure 3.8.: Scatter plot of two most selected features for SPORTS subsystem. Sensor position, sensor type, and corresponding axis or plane are given for each feature.

might result in a low performance for certain systems. With the proposed architecture, the task of window size selection might be easier. The current proposed algorithm can further be combined with existing adaptive, active, self-, and co-learning methods to further increase the performance.

A detailed comparison of the hierarchical classification to state-of-the-art techniques regarding the achieved accuracy values is performed in section 3.5.

Future Work

Investigating the individual selected features per subsystem further enables suitable extensions of the proposed HAR system. An example is given for the SPORTS subsystem. Figure 3.8 shows the scatter plot of the two most selected features. Besides the visible discriminatory power, the selected features could be used to provide users with additional information about the activity execution.



(a) Local classifier per node.

(b) Local classifier per level.

(c) Global classifier model about complete class hierarchy.

Figure 3.9.: Additional hierarchical classification approaches including local classifier per node (a), local classifier per level (b), and global classifier (c) (adapted from [Sill 11]). Circles and rectangles denote classes and classification systems, respectively. R denotes root node. Leaf nodes 1.1, 1.2, 2.1.1, 2.1.2, 2.2.1, and 2.2.2 represent final desired classes.

The following example is based on the 90th percentile of the linear acceleration amplitudes in the vertical axis acquired on the chest. The lowest feature values were observed for bicycling instances ranging from about - 1 g to - 0.5 g. For bicycling, this feature gives an indication of the posture of the upper body during bicycling. An example of an instance from the left border of the mentioned range can be seen in Figure 3.5c. This subject performed the activity in an upright position of the upper body. The lower the absolute amplitude of the feature, the more the subject leans forwards or backwards. An extension of the proposed HAR system in the future could include a posture estimation of subjects during bicycling, who are provided with feedback about the actual posture and necessary posture changes. According to Figure 3.8, the values of the 90th percentile of rope jumping instances are the highest. The meaning of the feature in case of rope jumping can be shown in Figure 3.5d. Since the vertical axis points to the ground, the local minima of the vertical acceleration indicate a jump in the air. The 90th percentile of vertical acceleration might therefore be used to estimate the power of jumps.

In this thesis, one classifier was trained for each parent node (Figure 1.3b). Future researchers should further investigate other approaches for hierarchical classification, which were mentioned in [Sill 11] and shown in Figure 3.9. One classification system could be trained for each node (Figure 3.9a) or for each level (Figure 3.9b). Each single classification system should implement an ARC. The implementation of the same components, which were used in this thesis, might be a suitable starting point for further investigations. In addition, one global classifier could be investigated that learns a global classification model about the complete class hierarchy (Figure 3.9c).

3.4 Decision Level Fusion

3.4.1 Overview

Fusing multiple sensors for HAR showed an overall performance improvement [Liu 12]. Most of the algorithms presented in section 1.5.1 applied feature level fusion, which induces various problems. First, changing the number of sensors in HAR systems results in retraining of the whole system [Bano 13]. Second, the performance of activity recognition systems during runtime is affected by degradation, interconnection failures, and jitter in sensor placement and orientation of single sensors [Zapp 07].

Third, a high number of sensors often results in an increased computational complexity in certain steps of the machine learning pipeline before the actual classification [Bano 13]. Compared to feature level fusion, decision level fusion facilitates an extensible and scalable HAR system, avoids the need of retraining under different number of used sensors, is often robust against sensor problems during runtime, and reduces the computational complexity at lower levels of the machine learning pipeline [Zapp 07, Zoub 09, Bano 13]. Nevertheless, research about decision level fusion for the recognition of daily activities is rather limited [Zapp 07, Bano 13, Chow 18]. Thus, the purpose of this section is to introduce a decision level fusion approach for HAR. The algorithm was evaluated on the DaLiAc dataset including sensors on hip, chest, wrist, and shoe (section 3.2.1). The algorithm was originally published in [Schu 13a] and was modified in this work. The modified parts are highlighted in the discussion.

This section is structured as follows. First, the applied machine learning techniques are introduced. Second, the conducted experiments are explained. Third, the results of the experiments are presented. The section concludes with a discussion of the results.

3.4.2 Methods

The proposed HAR system is depicted in Figure 3.10 and consisted of five steps. Four steps were performed for each of the four sensors separately including preprocessing, segmentation, feature extraction, and classification. The fifth step was to combine the information of the individual sensors at decision level. In the following paragraphs, the five steps are explained in more detail.

The preprocessing and segmentation step were the same as described in section 3.3 : calibration, synchronization, segmenting 5 s windows with 50 % overlap. Based on the sensor data in each segmented window, 56 generic features were computed. The list of features is given in Table 2.1. In order to reduce the number of features, the correlation-based feature selection approach was applied. The features were used as input for a classifier, in order to map a given window to one of the 13 activities.

The decisions of the four sensor nodes were finally fused by majority voting [Kitt 98, Mang 10]. The class with the highest number of votes was chosen. In the case of an equal distribution of the predicted classes, the highest rank method was applied [Ho 94]. The predictions were ranked according to a certain criterion and the predicted class with highest rank

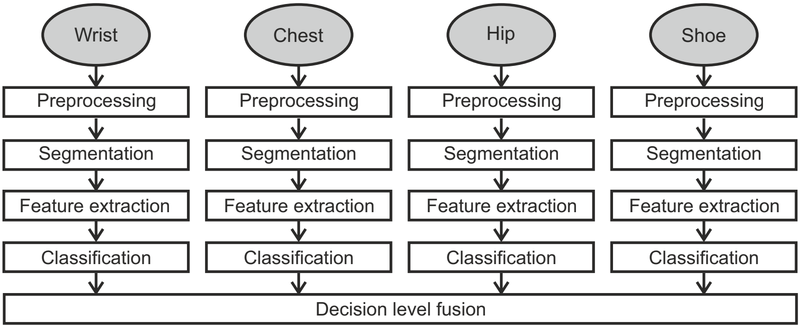


Figure 3.10.: Pipeline of decision level fusion (adapted from [Schu 13a]). Technique was applied to DaLiAc dataset (section 3.2.1).

was chosen as final prediction. In this work, the achieved sensor-specific sensitivity of the predicted activities was used as rank criterion.

3.4.3 Experiments and Evaluation

The proposed HAR system was evaluated on the DaLiAc database (section 3.2.1). The datasets of four subjects had to be excluded from further processing due to hardware issues. The number of available instances for each of the 13 activities is listed in Table 3.2. Two experiments were conducted.

- **Experiment 1: Individual sensor nodes**
  The goal was to compare the performance of the classification systems for each of the four sensors separately. The five classifiers NB, SVM, k-NN, CART, and RF were applied to the selected features. The balanced accuracy (section 2.4.8) was used as performance measure, which was determined by LOSO-CV. As shown in Table 3.2, the number of instances varied among activities. The number of walking instances was about twice as high as e.g. the number of instances of washing dishes. In order to address this class imbalance for walking, a random-undersampling of factor two was performed in each LOSO-CV trial and for each subject separately [He 09]. Further class imbalances were tackled by applying SMOTE filtering in each LOSO-CV trial [Chaw 02]. The linear kernel was used in

the SVM classifier and the corresponding cost parameter C was optimized by grid search ($C \in \{2^N\}, N \in \{-10, ..., 10\}$). In order to reduce the training time, the parallelized version of the SVM was applied (see section 2.4). The number of neighbors k in the k-NN classifier was optimized by grid search ($k \in \{3, 5, 7\}$). The number of trees in the RF classifier was optimized by grid search ($n_{tree} \in \{5, 10, 15, 20\}$). The parameter optimization of the classifiers was performed in an inner LOSO-CV. The performance of the single classification systems was further investigated regarding the 13 activities. The best classifier of each single sensor position was selected and the sensitivity was used as performance measure (section 2.4.8).

- **Experiment 2: Decision level fusion**
  The fusion of the four sensor nodes at decision level was studied. All possible combinations of two, three, and four sensors were compared regarding their balanced accuracy. For simplicity, the best classifier for each individual sensor position which was determined in the previous experiment was used in the fusion process.

3.4.4 Results

Table 3.6 depicts the balanced accuracy of the five classifiers CART, NB, k-NN, RF, and SVM regarding the individual sensor positions wrist, chest, hip, and shoe. The best balanced accuracy of 80.0 % was achieved by the chest sensor. In Table 3.7, the sensitivities of the 13 activities regarding the four sensor positions are given. Figure 3.11 shows the balanced accuracy for all possible sensor configurations. The best balanced accuracy of 85.7 % was achieved by using all four sensor positions.

Table 3.6.: Balanced accuracy [%] of CART, NB, k-NN, RF, and SVM regarding the four sensor positions wrist, chest, hip, and shoe. The best classifier per sensor position is marked in bold.

	CART	NB	kNN	RF	SVM
Wrist	72.4	68.6	74.2	**79.0**	73.5
Chest	72.0	75.4	78.8	**80.0**	77.2
Hip	71.8	73.2	75.0	**77.9**	72.1
Shoe	69.7	70.6	71.4	**75.9**	71.1

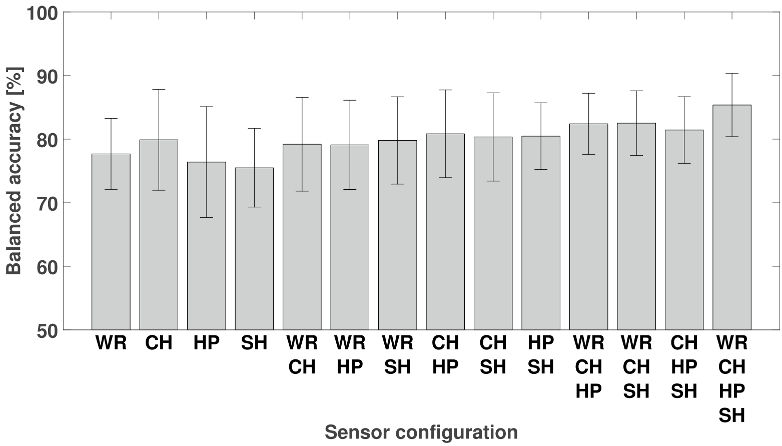


Figure 3.11.: Results of decision level fusion: balanced accuracy [%] regarding all possible sensor configurations. Mean ± standard deviation is given. WR, CH, HP, SH denote wrist, chest, hip, and shoe, respectively.

3.4.5 Discussion

A decision level fusion technique for HAR was developed, implemented, and evaluated in this work. The achieved results are discussed in this section. The discussion is structured as follows. First, the modifications of the originally proposed method published in [Schu 13a] are described and motivated. Second, the performance of the single sensors is compared (experiment 1). Third, the results of the decision level fusion are discussed (experiment 2). Fourth, the proposed approach is compared to state-of-the-art approaches in literature. Fifth, examples for future work are provided.

Modifications Compared to Original Publication

In order to evaluate the decision level fusion on a larger set of activities, the number of considered activities was increased from seven to 13. Instead of using the same set of features for all sensor-specific systems, a larger set of features was chosen in this thesis (Table 2.1) followed by a correlation-based feature selection. The aim was to provide features which are optimal for each sensor position. The proposed algorithm

Table 3.7.: Sensitivity [%] of 13 activities regarding the four sensor positions wrist, chest, hip, and shoe. The best classifier for each sensor position was chosen. The highest achieved sensitivity across the sensor positions is marked in bold.

	Wrist	Chest	Hip	Shoe
Sitting	84.5	78.0	**92.3**	69.9
Lying	90.0	89.5	89.5	**100.0**
Standing	79.5	**85.5**	75.2	44.5
Washing dishes	**90.7**	83.8	82.3	63.0
Vacuuming	**68.3**	67.2	55.1	58.7
Sweeping	72.1	**83.0**	73.8	70.8
Walking outside	85.3	92.1	94.7	**96.5**
Ascending stairs	81.4	89.0	86.9	**92.0**
Descending stairs	80.0	87.5	84.6	**87.9**
Treadmill running	**100.0**	96.9	92.5	99.6
Bicycling (50 watt)	48.2	48.5	47.9	**48.7**
Bicycling (100 watt)	47.2	38.5	41.3	**59.3**
Rope jumping	99.1	**100.0**	96.4	95.7

further included techniques for class balancing and the described method for the equality problem in the majority voting step.

All in all, the accuracy decreased by 8.2 % compared to [Schu 13a]. The performance of the single systems for wrist, chest, hip, and shoe decreased by 1.7 %, 9.6 %, 13.4 %, and 15.2 %, respectively. The reason for the worse performance might be the added activities. Vacuuming might be misclassified as washing dishes or sweeping which resulted in a low sensitivity for vacuuming (Table 3.7). The low sensitivity values for bicycling further reduced the overall accuracy. Nevertheless, in both versions of the algorithm, a performance improvement of 5.7 % was achieved by fusing all four systems compared to the best single system.

Individual Sensor Positions

The system of the chest achieved the highest balanced accuracy of 80.0 % (Table 3.6). One reason might be the proximity to the center of mass of the human body. Nevertheless, the hip position was usually preferred in literature [Bout 97a, Clel 13]. The lower balanced accuracy of the hip sensor might be explained by the sensor attachment. The sensor on the chest was integrated in a shirt which ensured tight fit (Figure 3.1). The tight fit might result in a more consistent signal characteristic over the subjects than in the case of the hip sensor. The hip sensor was attached to the trouser of the subject by a clip, which might enable the inclusion of noise vibrations caused by trouser movement. According to Figure 3.11, the mentioned noise vibrations seemed to be subject-dependent due to the high standard deviation of 10.9 % across the subjects considering the hip sensor. In order to reduce the noise vibrations, the position of the sensor could be moved e.g. to the sacrum [Pree 09] or clipped to a belt at the level of umbilicus [Liu 12].

Table 3.7 shows that the sensitivities of the 13 activities vary across the four sensor positions. Thus, each sensor position can only be used to recognize a subset of the original activity set. This also coincides with the findings in [Chow 18]. The wrist sensor of the proposed system achieved the best performance for washing dishes and vacuuming. The reason might be that the activities mainly include movements of the upper extremities. The shoe sensor system achieved the best performance for walking, ascending stairs, and descending stairs, which mainly include movements of the lower extremities.

Decision Level Fusion

Using two sensors only partially increased the balanced accuracy e.g. in the case of wrist/shoe combination compared to only using shoe sensor (80.1 % vs 75.8 %, Figure 3.11). The authors in [Bano 13] mentioned that fusion approaches drop on the performance considering a small sensor network. This behavior was also partially seen in this work. Adding the decision of the hip sensor to the decision of the chest sensor decreased the mean balanced accuracy from 79.9 % to 79.2 % and increased the standard deviation of the balanced accuracy from 6.7 % to 8.3 %. The reason might be the presence of an equal distribution of class predictions in the majority voting which results in using the proposed highest rank method. In future work, different fusion techniques should be investigated e.g. weighted majority voting, class-conscious or class-indifferent methods [Mang 10].

Nevertheless, averaging all three-sensor configurations resulted in a mean balanced accuracy of 83.8 %. The improvement of about 4 % compared to the best two-sensor configuration could be explained by the number of sensors. According to [Mang 10], majority voting is an optimal combination rule with an odd number of classifiers.

Considering the decisions of all four sensor systems in the fusion process achieved the highest mean balanced accuracy of 85.7 % with the lowest standard deviation of 4.1 % across all combinations. Thus, fusing various sensors at decision level seemed to reduce subject variability.

One major advantage of the proposed system is that no retraining is needed, if sensors are added or removed. A new sensor, e.g. on the shank, can be trained separately without affecting the already existing sensors. The decision of the shank sensor can be included in the final majority voting step.

Comparison to Literature

In [Bano 13], the best accuracy for a single sensor position was achieved by the wrist position (above 95 %). In this thesis, the system of the wrist sensor achieved the second best accuracy (78.8 %). Two reasons were found for the worse performance of the proposed method. First, the two methods applied different cross-validation settings. In [Bano 13], a 10-fold cross-validation was performed. Thus, data of one subject was available in the training and testing phase of the system, which might lead to a performance increasement. Nevertheless, LOSO-CV was applied in this thesis, which might ensure a better investigation of the generalization capabilities of the developed system. Second, both methods were applied on different activities. In [Bano 13], only nine activities were classified. Ascending as well as descending stairs were further merged to one class and no distinction between different bicycling levels was made. The combination of these classes might have a positive effect on the system performance.

The authors in [Chow 18] applied similar techniques for preprocessing, feature extraction, feature selection, and classification for the sensor-specific systems compared to the proposed algorithm in this thesis. In [Chow 18], eight activities were classified. Both methods integrated prior knowledge of the performance of the systems in the decision level fusion step. In [Chow 18], a score for each activity was computed including class-based weights, which were determined in the training phase. These weights were adapted by the posterior probabilities of the classifiers in the testing phase. The proposed approach in this thesis applied majority voting in the first step and only used prior knowledge about the training performance in case of an equal distribution of predicted classes. The algorithm described in [Chow 18] achieved a mean F1-score of 92.3 % using the PAMAP2 dataset including sensors at ankle, chest, and wrist

(Table 1.1). The proposed method in this thesis achieved a mean F1-score of 82.5 % using a comparable sensor set (wrist, chest, shoe). The better performance in [Chow 18] might show the advantage of using the posterior probability as weighting factors in the decision level fusion. Nevertheless, not all classifiers provide posterior probabilities. Thus, the method descibed in [Chow 18] is restricted in the type of classifiers. The proposed algorithm in this thesis, which is based on majority voting is not dependent on the posterior probabilities and can therefore be used for all classifiers.

Future Work

The achieved balanced accuracy of 85.7 % showed the general applicability of the proposed method for decision level fusion. Nevertheless, the approach was only tested with four sensor positions and only considering IMU sensors. In future, additional sensor positions have to be integrated in the pipeline shown in Figure 3.10. Future applications might provide a huge amount of unobtrusive sensors attached to the human's body or integrated in smart homes. Considering a large set of sensors, algorithms might be mandatory to select sensors, which negatively influence the overall performance of the system, e.g. because of sensor degradation. The proposed highest rank method could be one starting point for sensor degradation detection. Regular sensitivity analysis of the single sensors might observe lower values after a specific time frame, which indicates problems in certain sensors.

3.5 Comparison of Algorithms Using a Common Evaluation Framework

3.5.1 Overview

The identification of the best HAR algorithm for a specific application is challenging. Research papers often compared novel algorithms to existing ones based on the final achieved performance. Nevertheless, the explanatory power of this kind of comparison was limited, since different databases and evaluation strategies were applied. A fair comparison of HAR systems requires a common framework for evaluation, e.g. by applying the same set of experiments with the same performance measures to the desired algorithms. Thus, the purpose of this section is to compare various available HAR algorithms in one common evaluation framework using the same benchmark dataset. This procedure was originally published in [Schu 13b] and was modified in this work. The modified parts are highlighted in the discussion.

This section is structured as follows. First, the methods for a common evaluation framework are introduced. Second, the performed experiments are described. Third, the results of the experiments are given. The section concludes with the discussion of the results and findings.

3.5.2 Methods

The common evaluation framework for comparing various HAR algorithms is shown in Figure 3.12 and consisted of four major steps: selection of benchmark dataset, data preparation, execution of re-implemented ARC (Figure 1.1), and definition of experiments and evaluation. In the following paragraphs, the application of the proposed framework is shown by means of an example. The two previously proposed algorithms for HAR (section 3.3 and 3.4) were compared to the five state-of-the-art approaches [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12].

Selection of Benchmark Dataset

The first step was the selection of an appropriate benchmark dataset. The study design, with which the benchmark dataset was collected, should have intersections with the dataset originally used for evaluation. In this work, the DaLiAc dataset was used for benchmarking the two proposed algorithms for HAR (section 3.3 and 3.4) and the five state-of-the-art approaches [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12]. DaLiAc contained comparable sensor types, sensor placement, and performed activities (section 3.2.1).

Data Preparation

Two data preparation steps were required. First, the appropriate sensors had to be selected from the DaLiAc database. Only comparable sensor positions and types, which were found in the original papers, were used. Other sensor positions and types were disregarded. The sensor positions from the original papers [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12] and the used positions from the DaLiac database are given in Table 3.8. Since the five state-of-the-art approaches only used accelerometer sensors, the gyroscope signals were not considered as input data for these algorithms.

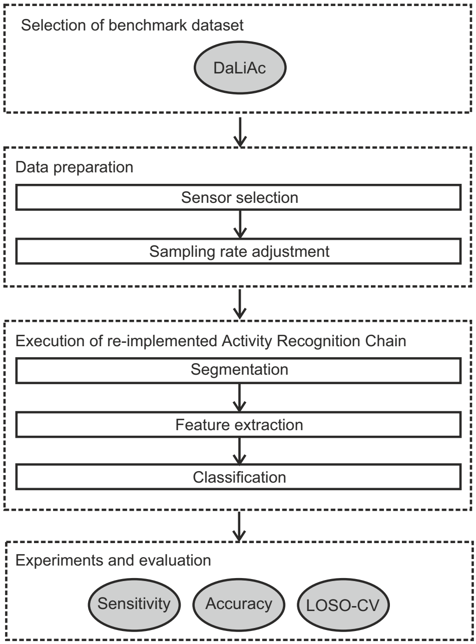


Figure 3.12.: Comparison of algorithms using a common evaluation framework. Framework consists of the selection of a benchmark dataset such as DaLiAc, data preparation, execution of re-implemented ARC, and the definition of experiments and evaluation. Performance measures in the evaluation could include e.g. sensitivity and accuracy determined by LOSO-CV.

Second, the sensor data of the DaLiAc database were re-sampled considering the sampling rates proposed in the original papers (Table3.8).

Re-Implementation of Activity Recognition Chain

The five implementations of the ARC [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12] were chosen for three reasons. First, these techniques achieved a high citation rate and thus had a sufficient research impact. Second, their study setup had intersections with the proposed benchmark dataset. Third, various applied machine learning techniques were similar compared to the two proposed algorithms (section 3.3 and 3.4), which

Table 3.8.: Re-implemented ARCs. Details regarding original sensor position, used sensor position from DaLiAc, and sampling rate are given.

Position (Original)	**Position (DaLiAc)**	**Sampling Rate**	**Reference**
Right hip Right wrist Left arm Right ankle Left thigh	Right hip Right wrist Left shoe	76.25 Hz	[Bao 04]
Pelvic	Right hip	50 Hz	[Ravi 05]
Chest Wrist (dominant)	Chest Right wrist	Chest: 200 Hz Wrist: 40 Hz	[Park 06]
Waist Thigh Ankle	Right hip Left shoe	64 Hz	[Pree 09]
Hip (dominant) Wrist (dominant)	Right hip Right wrist	30 Hz	[Liu 12]

simplified the re-implementation, since parts of algorithms could be shared. Each of the five state-of-the-art approaches was re-implemented based on the description found in the original paper. Re-implementation of the algorithms only considered the suggested final feature set and the classifier which achieved the best performance. Details about the re-implemented algorithms are described in the following paragraphs.

The authors in [Bao 04] considered two-axial accelerometer data as input. Since the orientation of the two axes were not provided, all three available axes from the DaLiAc dataset were used. An epoch size of 6.7 s with 50% overlap was used in the segmentation step. Three features were computed for each axes of the segmented sensor data: mean, energy, entropy. The mean was computed in the time domain. Energy and entropy were computed in the frequency domain. In addition, the correlation between each pair of axes across the sensor positions was computed by applying the dot product normalized by the window length. In total, 63 features were computed. The features were used as input for a C4.5 decision tree classifier.

In [Ravi 05], an epoch size of 5.12 s with 50 % overlap was used in the segmentation step. Three features were computed for each accelerometer axis: mean, standard deviation, energy. Mean and standard deviation were computed in the time domain. Energy was computed in the frequency domain. In addition, the correlation between each pair of axes was computed as ratio of covariance and product of standard deviations. In total, twelve features were extracted. The features were used as input for the following classifiers: decision tables, C4.5 decision trees, k-NN, SVM, and NB. The final prediction was determined by plurality voting.

In [Park 06], 1-s segments were used and six features were selected based on visual analysis of the feature distribution: peak frequency of vertical acceleration of chest sensor, median of vertical acceleration of chest sensor, peak power of vertical acceleration of chest sensor, variance of anteroposterior acceleration of chest sensor, sum of variances of 3-D wrist accelerations, power ratio of frequency bands 1 - 1.5 Hz and 0.2 -5 Hz measured from mediolateral axis of magnetometer. Since the DaLiAc database only contained accelerometer and gyroscope data, the magnetometer feature was not considered. In total, 5 features were computed. The features were used as input for a CART classifier. Median filtering was further applied to the classifier decisions of subsequent segments in order to remove short-duration misclassifications. A filter width of 31 s was applied.

The authors in [Pree 09] applied 2-s windows with 50 % overlap in the segmentation step. The magnitudes of the first five components of the Fast Fourier Transform were extracted as features. In total, 30 features were computed. A k-NN classifier was used to predict the activities.

In [Liu 12], the sensor data were segmented in 30-s windows. Time domain and frequency domain features were defined as follows:

- Time domain:
  Mean, standard deviation, 10th/25th/50th/75th/90th percentiles
- Frequency domain:
  Energy, entropy

The features based on the hip accelerometer included 25th percentile, standard deviation, and spectral entropy of vertical axis, spectral energy of mediolateral axis, and 90th percentile as well as standard deviation of anteroposterior axis. The features based on the wrist sensor included mean, standard deviation, 25th/50th/75th/90th percentiles and all frequency domain features of the vertical axis, all time and frequency domain features of the anteroposterior axis, and all time domain features as well as spectral energy of the mediolateral axis. In total, 31 features were computed and were used as input for a SVM.

3.5.3 Experiments and Evaluation

The two proposed algorithms described in section 3.3 and 3.4 were compared to the five, re-implemented state-of-the-art algorithms [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12]. The algorithms were applied to the DaLiAc dataset containing 13 activities (Table 3.1). Two experiments were conducted investigating the overall and activity-dependent performance.

- **Experiment 1: Overall performance**
  The overall performance was determined by the balanced accuracy and the F1-score averaged regarding the single LOSO-CV trials (section 2.4.8). The optimization of the algorithms, which were developed in this thesis, was already described in section 3.3 and 3.4. In [Ravi 05, Pree 09], the number of neighbors k in the k-NN classifier was optimized by grid search ($k \in \{3, 5, 7\}$). In [Ravi 05], the linear kernel was used in the SVM classifier and the corresponding cost parameter C was optimized by grid search ($C \in \{2^N\}, N \in \{-10, ..., 10\}$). In [Liu 12], the RBF kernel was used in the SVM. The cost parameters C and γ were optimized by grid search ($C, \gamma \in \{2^N\}, N \in \{-10, ..., 10\}$). The parameter optimization of the classifiers were performed in an inner LOSO-CV. The overall performance of the different algorithms was further investigated regarding the number of considered sensors. A boxplot of the balanced accuracy values across subjects was created.

- **Experiment 2: Activity-dependent performance**
  The activity-dependent performance was determined by computing the sensitivity per activity. The values for the sensitivity were averaged regarding the single LOSO-CV trials (section 2.4.8).

3.5.4 Results

Table 3.10a and 3.10b comprise the overall balanced accuracy as well as the F1-score and the activity-dependent sensitivities, respectively. The proposed hierarchical HAR algorithm, which was described in section 3.3, achieved the highest balanced accuracy and F1-score (89.1 % and 90.5 %). The boxplots of balanced accuracy values ordered by the number of sensors are displayed in Figure 3.13.

3.5.5 Discussion

In this work, the two proposed algorithms (section 3.3 and 3.4) were compared to five state-of-the-art algorithms [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12]. The achieved results are discussed in this section. The discussion is structured as follows. First, the modifications compared to the original publication [Schu 13b] are summarized. Second, challenges in the data preparation and re-implementation of algorithms are mentioned. Third, the overall and activity-dependent performance are discussed (experiment 1 and 2). Fourth, a comparison to literature is given. Fifth, future work is summarized.

Modifications Compared to Original Publication

In [Schu 13b], the algorithm introduced in [Kara 06] was further included in the comparison. Nevertheless, the algorithm was optimized for a specific activity set and was not applicable for all of the 13 activities provided by the DaLiAc dataset. Instead, the proposed method described in section 3.4 was included in the comparison. In the re-implementation of the algorithms, the same classifier optimization was further performed as in the two proposed approaches (section 3.3 and 3.4). In [Schu 13b], default parameters were considered.

The parameter optimization had a major impact on the accuracy values which increased by 8.1 %, 5.0 %, and 2.6 % for the techniques described in [Ravi 05], [Park 06], and [Pree 09], respectively. The performance achieved by [Bao 04] was comparable. The highest performance improvement of 27.6 % was achieved for the method described in [Liu 12]. It seemed that the SVM classifier required the parameter optimization in order to deal with the 31 features.

Table 3.9.: Comparison of two proposed algorithms (section 3.3 and 3.4) and five state-of-the-art algorithms [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12]. Overall balanced accuracy and F1-score (a) as well as activity-dependent sensitivities (b) are given in percent. Columns depict one particular algorithm.

	[Bao 04]	[Ravi 05]	[Park 06]	[Pree 09]	[Liu 12]	Sec. 3.4	Sec. 3.3
BalAcc	79.7	67.2	66.8	71.3	81.7	85.7	**89.1**
F1	81.2	72.9	68.9	72.3	89.8	85.4	**90.5**

(**a**) Balanced accuracy and F1-score

	[Bao 04]	[Ravi 05]	[Park 06]	[Pree 09]	[Liu 12]	Sec. 3.4	Sec. 3.3
SI	90.5	86.7	70.0	78.9	84.2	**92.8**	92.0
LY	**100.0**	88.8	98.9	94.7	92.1	**100.0**	**100.0**
ST	76.3	68.5	75.2	60.0	84.2	**91.2**	90.6
WD	91.1	86.5	95.4	58.8	94.7	**97.4**	92.3
VC	55.1	24.2	47.3	46.4	78.9	58.3	**87.6**
SW	84.5	69.8	87.7	66.4	**94.3**	92.8	87.0
WK	96.8	96.9	86.9	97.0	95.3	98.5	**99.3**
AS	69.3	45.1	15.1	77.5	73.7	92.5	**97.8**
DS	78.3	35.2	20.1	53.6	52.9	90.9	**99.3**
RU	99.8	98.1	93.5	98.6	**100.0**	**100.0**	99.3
BC50	52.1	51.9	54.7	57.2	60.5	49.8	**64.1**
BC100	45.0	25.1	32.8	50.3	**55.3**	51.1	52.2
RJ	**100.0**	99.5	90.8	90.1	**100.0**	**100.0**	96.6

(**b**) Sensitivity. Abbreviations for activities were defined in Table 3.1.

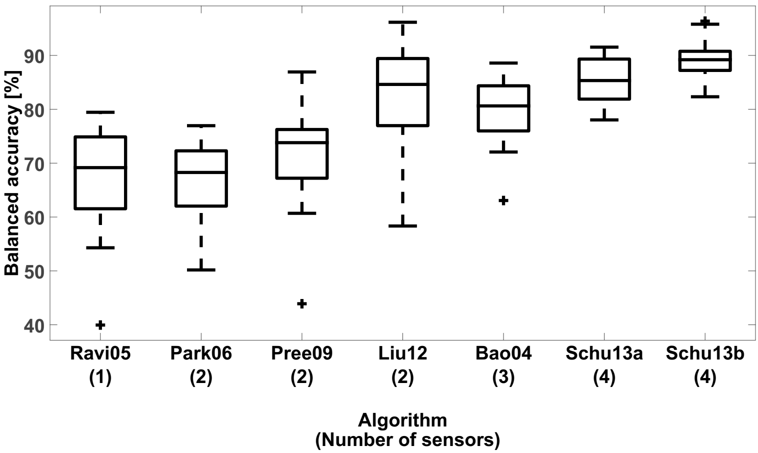


Figure 3.13.: Comparison of algorithms. Boxplot of balanced accuracy values of LOSO-CV trials is displayed over seven algorithms (number of sensors is written in brackets). For each algorithm median, 25th quartile, 75th quartile, 1.5 of interquartile range, and outliers are given.

Data Preparation and Re-Implementation

Various challenges were identified in the data preparation and re-implementation step of the proposed evaluation framework (Figure 3.12). In [Bao 04], a sensor was attached to the right ankle acquiring data of the lower extremities. Since the DaLiAc dataset did not include an ankle sensor, the sensor on the shoe was chosen instead (Table 3.8). Nevertheless, the different sensor position could have an effect on the extracted features since the amplitudes often increase from cranial to caudal body parts [Bhat 80, Capp 82]. The authors in [Bao 04] used a 10 g accelerometer which might not reach saturation. In the DaLiAc datasets, sensor data were acquired with a lower g-range and might reach the saturation state in some activities. Another challenge is the detailed description of the sensor position. Descriptions in literature for a sensor close to center of mass included waist (at sacrum) [Pree 09], hip [Schu 13a], and pelvic region [Ravi 05]. Signal characteristics and computed features might be slightly different depending on the exact position. The effect might be minimal, if the same features are computed for each accelerometer axis like in [Bao 04, Ravi 05, Pree 09]. Nevertheless, extracting axis-specific features like in [Park 06, Liu 12] could have a negative influence on the performance of algorithms. If the sensor axis points in a slightly different direction in the new dataset compared to the originally used dataset, the computed feature might not be suitable anymore. A standardized description of the sensor position might further be beneficial to avoid errors in the axis-specific feature selection. An example could be the description of the hip sensor in [Liu 12], which was attached to the 'abdomen at level of umbilicus in line with the anterior auxiliary line'.

Overall Performance

Both methods proposed in this thesis achieved higher balanced accuracy than the five re-implemented algorithms (Table 3.10a). The balanced accuracy values of above 85 % showed the general applicability of the systems for HAR. The proposed hierarchical classification system further achieved the lowest interquartile range (Figure 3.13). This indicates that the algorithm can deal with subject variability. If the algorithm is part of a deployed system, further optimization might not be necessary after the release of the product. Other algorithms with a higher interquartile range such as [Ravi 05, Liu 12] might have to apply user-specific training as proposed by [Bao 04], in order to reduce the subject variability. Nevertheless, four IMUs sensors comprising accelerometer and gyroscope were needed in the proposed hierarchical classification system, which was the highest number of used sensors across all algorithms (Figure 3.13). Thus, in real applications, a trade-off between number of sensors and system accuracy has to be made.

Although the proposed decision level fusion approach achieved the second best balanced accuracy, the algorithm introduced in [Liu 12] achieved the second best F1-score. This shows one of the major advantages of the proposed common evaluation framework (Figure 3.12). The identification of the best algorithm for a specific application can be based on a quantitative assessment using different metrics. Whereas the balanced accuracy is only based on the sensitivity measure, the F1-score also includes the precision (section 2.4.8). Precision gives an indication about the confidence of the prediction output of the system. Thus, an algorithm can be chosen according to certain requirements of a specific application.

Figure 3.13 shows that in [Ravi 05, Pree 09, Bao 04] outliers occurred below the lower whisker. In [Ravi 05, Pree 09], the outlier subject was the oldest and heaviest person included in the pool of subjects. The selected features might not describe the movement pattern of this person. Thus, both algorithms might have problems, if applied to elderly and overweight persons. In order to further improve these systems, anthropometric data could be used as additional features. If algorithms are tested on different datasets including e.g. different age groups, the above described behavior could not be detected.

The best mean balanced accuracy among the five re-implemented algorithms was achieved by [Liu 12]. The extracted features were similar to the feature set defined in this work (Table 2.1). Compared to other algorithms which used statistical features, the 10th/25th/50th/75th/90th percentiles were applied both in [Liu 12] and the two algorithms proposed in this work. Thus, describing the amplitude distribution in more detail might be beneficial for this set of activities. The advantages of the method provided by [Liu 12] compared to the two proposed algorithms in this work include the lower sampling rate, the lower amount of sensors, the lower amount of sensor types, and the larger window size. These factors reduce the computational complexity of the HAR system.

Although the balanced accuracy achieved by [Bao 04] was 2.0 % lower than the balanced accuracy achieved by [Liu 12], the interquartile range was also lower (Figure 3.13). This means that the method provided by [Bao 04] can better deal with the subject variability. Another reason might be that the arm and thigh sensor, which were originally part of the sensor setup, had to be removed (Table3.8). Since the feature set was optimized considering these two sensor positions, the accuracy might increase, if these sensor positions are included.

Although the re-implemented algorithms based on [Park 06] and [Pree 09] also used two sensor positions, both methods achieved a lower balanced accuracy than the method provided by [Liu 12]. The reason might be the extracted features which were optimized for a different set of activities. In [Park 06] a visual and statistical analysis was performed for feature selection based on an initial feature set which was similar to the feature set proposed in this thesis. Applying an automatic approach for feature selection, e.g. based on correlation, might improve the performance.

The method proposed by [Ravi 05] achieved a rather low balanced accuracy. Nevertheless, only one sensor was used e.g. compared to [Liu 12]. In [Ravi 05], data of only two subjects were used and the classifier was optimized by 10-fold cross-validation, which is prone to over-fitting. In this thesis, the approach was tested on 19 subjects considering a LOSO-CV.

Thus, investigating the generalization capabilities of the system might be better suitable with the proposed evaluation framework.

Activity-Dependent Performance

The lowest sensitivity regarding sitting was achieved by the method introduced in [Park 06] (70.0 %, Table 3.10b). One reason might be that sitting and standing were merged in the original paper due to known misclassification of both activities. The higher sensitivities of the other approaches could be further explained by the additional sensor attached to the hip, which might provide more information about the static activities.

The lowest sensitivity regarding standing was achieved by the re-implemented algorithm proposed in [Pree 09]. One reason might be the applied feature set. In [Pree 09], only components of the Fast Fourier Transform were chosen as features, since activities with mostly periodic movements were included in the original dataset. This feature set might not be suitable for various kinds of static activities.

According to Table 3.10b, similar sensitivity values were achieved by the two proposed algorithms across the activities except for vacuuming. The difference of 29.3 % could be explained by the number of available classes for the classifiers. In the decision level fusion algorithm, the classifier had to decide among 13 activities. The hierarchical classification algorithm merged certain activities and only three activities remained in the HOUSE system. This shows the advantage of merging activities and dividing the multi-class classification to smaller sub-problems. Nevertheless, the hierarchical system based on feature level fusion is less flexible than the decision level fusion approach, if the information of one or more sensors is not available. It is proposed to combine the advantages of both systems in the future. Different subsystems which merge certain target activities should be trained for each sensor and a decision level fusion should be applied in the final step.

The algorithm described in [Liu 12] was originally evaluated on datasets which did not contain various activities from the DaLiAc database, e.g. sweeping and washing dishes. Due to the high achieved sensitivities of these activities, the implemented features seemed also be suitable to recognize activities which were not considered in original feature engineering step.

Regarding the locomotion activities walking, ascending stairs, descending stairs, and running, a comparable performance was achieved for walking and running by all algorithms. Thus, a good detection of both activities could be performed by multiple approaches varying in sampling rate, segmentation windows, extracted features, and chosen classifiers. Nevertheless, sensitivity values below 78.3 % were achieved for the stair climbing activities by all re-implemented algorithms. Only the two proposed methods in this thesis could distinguish between the upward and downward direction. Nevertheless, four sensors were needed. If the distinction between ascending stairs and descending stairs is not important, the method introduced in [Bao 04] should be used due to the highest averaged sensitivity across both activities.

All algorithms had problems to distinguish between the two bicycling activities with sensitivities below 64.1 %. Nevertheless, the original method proposed by [Liu 12] also included features based on an ventilation sensor. Adding this sensor type would further improve the recognition rate, especially for the classification of the two bicycling levels. The method provided by [Liu 12] outperformed the hierarchical classification algorithm developed in this work considering the performance regarding individual activities. Table 3.10b shows that the sensitivity regarding sweeping was higher using the method proposed in [Liu 12]. The reason might be that the 30-s windows used in [Liu 12] were more appropriate for detecting single sweeping actions than the 5 s windows used in the proposed approaches in this thesis.

The above mentioned examples showed the advantage of the proposed evaluation framework. Depending on the target activities and application requirements, the best suitable implementation of the ARC can be chosen. HAR system engineers should first define the requirements of an application, select available implementations of the ARC, and apply the proposed evaluation framework, to identify the best suitable method.

Comparison to Literature

Currently, novel algorithms for HAR were mainly compared to state-of-the-art approaches based on the performance measures provided in the corresponding research papers. Although this type of comparison enables a first hint of advantages and disadvantages of new HAR algorithms, several important aspects are not considered in the evaluation. The compared methods were evaluated on different datasets containing a different set of activities. The reason of a better performance of one specific algorithm can be the design of the method itself, but also the lower complexity of the activities in the datasets. A fair comparison should include the application of the algorithms to and the evaluation on a common benchmark dataset. The proposed evaluation framework could be one example for such a fair comparison. This proposed procedure would strengthen the meaningfulness of the achieved findings.

Future Work

The proposed evaluation framework could further be extended in the future. The re-implemented algorithms should be evaluated on more datasets, which were e.g. listed in Table 1.1. Thus, the behavior of algorithms under more conditions could be investigated. The set of algorithms should be extended by additional methods listed in Table 1.2. In the future, a common platform could be set up to provide open access to implemented algorithms. The platform could integrate the proposed evaluation framework to compare algorithms at a larger scale. In the future, the evaluation of the original proposed HAR algorithms could be extended by varying certain components or parameters of the ARC such as sampling rate, segmentation window width, and feature set. Besides testing algorithms, the evaluation framework could further be used to test new proposed performance measures in machine learning, e.g. the technique introduced in [Kaut 17]. The new technique can be tested on a large set of algorithms using the same benchmark dataset.

3.6 Database Fusion Strategy

3.6.1 Overview

Current algorithms for HAR were mainly evaluated on isolated databases (Table 1.1). These databases are often limited in at least one of the following aspects: number of subjects, number of activities, or number of sensors. The mentioned limitations of isolated databases reduce the number of instances which can be used for the training of HAR systems. The explanatory power of the achieved findings are further limited, since the algorithms were tested on a small population. The aim of this section is twofold. First, a database fusion method is developed and implemented which overcomes the limited amount of training instances by fusing various publicly available databases. Second, the applicability of the proposed database fusion technique is shown for one example algorithm for HAR.

The proposed database fusion and activity recognition algorithm were originally published in [Schu 14b] and were modified in this thesis. The modifications are described later in this section. This section is structured as follows. First, the database fusion technique is presented. Second, the proposed algorithm for HAR is introduced. Third, the conducted experiments are introduced. Fourth, the results of the experiments are presented. The section concludes with a discussion of the results.

3.6.2 Database Fusion Algorithm

The pipeline of the database fusion algorithm is shown in Figure 3.14 and consisted of two major tasks. The first task was to select the data, which was needed for a certain application. As an example, the following application should be considered. A flexible integration of new sensor positions and sensor types should be possible without retraining the complete system. The system should provide an activity profile including periods of human's inactivity as well as higher level locomotion, sports, and household activities. A detailed knowledge e.g. about the human's posture and the direction of movement are not required. The second task of the database fusion algorithm was to prepare the data for further

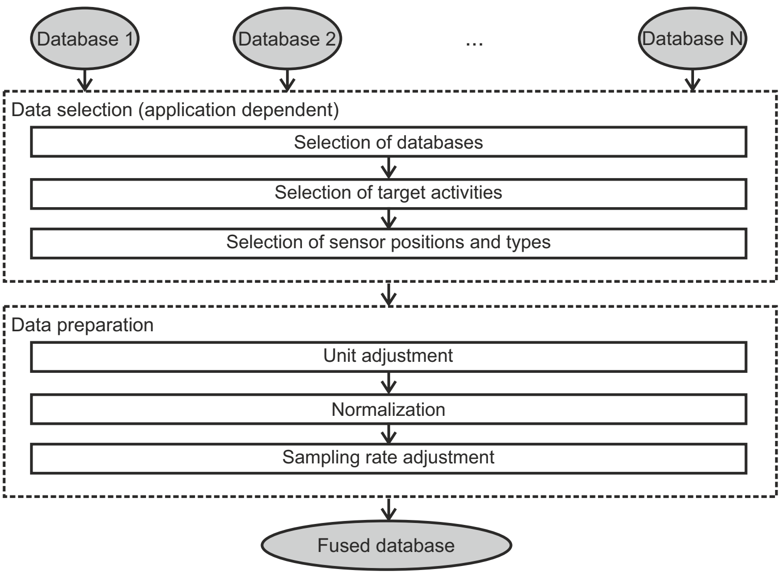


Figure 3.14.: Pipeline for database fusion.

processing. In the following paragraphs, the two tasks are described in more detail and applied to three example databases.

Data Selection

The data selection task consisted of three steps. The first step was to select appropriate isolated databases, which fulfilled the needs of the considered application and can be used for evaluating the corresponding HAR algorithm. The selected databases in this thesis included DaLiAc, PAMAP2, and USC-HAD (Table 1.1). All three databases provided sensor positions, sensor types, and a set of activities, which fulfilled the requirements of the specific application. The DaLiAc database was already described in section 3.2.1. PAMAP2 and USC-HAD are introduced in the next paragraphs.

In PAMAP2, the subjects were equipped with three Inertial-Magnetic Measurement Units attached to chest, wrist, and ankle as well as one heart rate monitor [Reis 12]. Each Inertial-Magnetic Measurement Unit consisted of two triaxial accelerometers, a triaxial gyroscope, and a triaxial magneto-resistive magnetic sensor. The range of the accelerometer, gyroscope, and magnetometer was $\pm$ 16 g / $\pm$ 6 g, $\pm$ 1500 $^{\circ}/s$, and $\pm$ 400 μT, respectively. The accelerometer and gyroscope data were given in m/s^2 and rad/s, respectively. The inertial data was sampled with 100 Hz. The dataset included nine subjects (1 female and 8 male, age 27.2 $\pm$ 3.3 years, Body-Mass-Index (BMI) 25.1 $\pm$ 2.6 kgm^{-2}). One subject was left-handed, the others were right-handed. Each subject had to perform 18 activities (lying, sitting, standing, walking, running, cycling, Nordic walking, watching TV, computer work, car driving, ascending stairs, descending stairs, vacuuming, ironing, folding laundry, house cleaning, playing soccer, rope jumping). Compared to DaLiAc, cycling and running were performed outside. The dataset is publicly available [1].

In USC-HAD, one Inertial-Magnetic Measurement Unit was attached to the subjects' hip [Zhan 12]. The inertial data was sampled with 100 Hz. The range of the accelerometer was set to $\pm$ 6g. The gyroscope range was set to $\pm$ 500 $^{\circ}/s$. The dataset included 14 subjects (7 female and 7 male, age 30.1 $\pm$ 7.2 years, height 170 $\pm$ 6.8 cm, weight 64.6 $\pm$ 12.1 kg). Each subject had to perform 12 activities (walking forward, walking left, walking right, walking upstairs, walking downstairs, running forward, jumping, sitting, standing, sleeping, elevator up, elevator down). Day-to-

[1] http://www.pamap.org/demo.html

Table 3.10.: List of selected activities (adapted from [Schu14b]). The availability of the activities regarding the three databases USC-HAD, DaLiAc, and PAMAP2 is indicated by '✓'.

Activity	USC-HAD	DaLiAc	PAMAP2
Static	✓	✓	✓
Walking	✓	✓	✓
Climbing stairs	✓	✓	✓
Running	✓		✓
Jumping	✓		
Vacuuming		✓	✓
Bicycling		✓	✓
Rope jumping		✓	✓

day activity variations were considered by performing five trials for each activity on different days at various indoor and outdoor locations. The dataset is publicly available [2].

The second step of the data selection task was to select the target activities which should be recognized by the system. Eight activities were selected, which typically appear in daily life and can be integrated in the previously mentioned activity profile. The eight selected activities and the availability in the three different databases are comprised in Table 3.10. The first activity was defined as **static** and comprised sitting, lying, sleeping, as well as standing. Detecting instances of the static activity gives an indication about human's inactivity during the day. Locomotion activities were represented by **walking** and **climbing stairs**. The later one consisted of ascending as well as descending stairs and were combined to one class, since knowledge of the direction of movement was not required. **Vacuuming** was chosen as an example of a typical household activity. Sports activities included **running**, **jumping**, **bicycling**, and **rope jumping**. In DaLiAc, the subjects performed bicycling on on a stationary bike with two resistance levels (50 and 100 watt). In PAMAP2, the cycling activity took place outside. In this work, instances of all bicycling types were merged.

The third step of the data selection task was to select the sensor positions and the sensor types. In this work, the triaxial accelerometer and gyroscope data of all four sensor positions from the DaLiAc database, all

[2] http://sipi.usc.edu/had/

three sensor positions from the PAMAP2 database, and the hip sensor from the USC-HAD database were used in the fusion process. In the PAMAP2 database, data of two triaxial accelerometers were available. Due to the author's recommendation, the high-range accelerometer was chosen for further processing.

Data Preparation

The second task was to prepare the data for further processing and consisted of three steps. The first step was to define common sensor units across the three different isolated databases. PAMAP2 used m/s^2 and rad/s as units for the accelerometer and gyroscope, respectively. In contrast, USC-HAD and DaLiAc used g and $°/s$ as units for the accelerometer and gyroscope, respectively. Thus, the units of accelerometer and gyroscope were converted to g and $°/s$, respectively. The second step was to define common sensor ranges across the three different isolated databases. The minimal available range per sensor type was chosen as common range. The accelerometer range was set to $\pm$ 6 g for all selected sensor positions. For the sensor positions wrist, chest, and hip, the gyroscope range was set to $\pm$ 500 $°/s$. For the sensor position ankle, the gyroscope range was set to $\pm$ 1500 $°/s$. All amplitudes outside the adjusted range were cut to the maximal border amplitudes. The third step was to define a common sampling rate across the databases. The minimal available sampling rate of 100 Hz was chosen for all sensor data. The adjustment of the sampling rate was based on linear interpolation.

3.6.3 Human Activity Recognition System

Figure 3.15 shows the pipeline of the proposed HAR system. The algorithm was based on the decision level fusion technique described in section 3.4.

Decision level fusion was proven to enable a flexible integration of additional sensor positions. In order to further add or remove various sensor types, the data acquired on the single sensor positions were further split according to the sensor types, i.e. accelerometer and gyroscope. Preprocessing, segmentation, feature extraction, and classification were performed for each sensor type of each individual sensor position, separately. The preprocessing step included the SMV computation for each sensor type (Equation 2.6). Examples of a SMV are shown in Figure 3.16 and 3.17 for climbing stairs and bicycling, respectively. The motivation for using

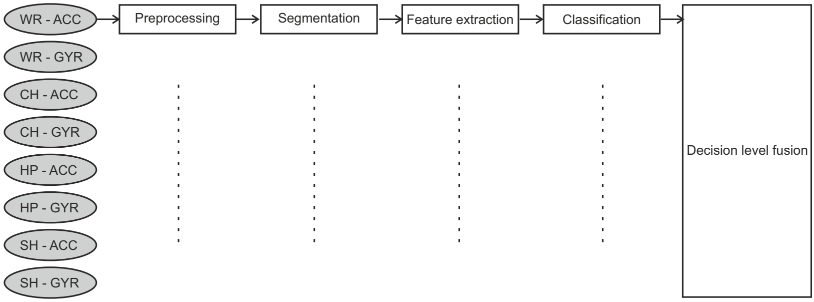


Figure 3.15.: Pipeline for decision level fusion (adapted from [Schu 14b]). WR, CH, HP, and SH denote wrist, chest, hip, and shoe, respectively. ACC and GYR denote accelerometer and gyroscope, respectively.

the signal magnitude vector instead of each axis of the sensor types was twofold. First, further processing was based on one axis, which reduced the computational complexity. Second, the knowledge of the movement direction was removed, since it was not required in the previously described application. A sliding window segmentation was performed with a window length of 5 s motivated by [Bao 04, Ravi 05, Pree 09]. In each window, the features listed in Table 2.1 were computed except the SMA, since this feature required three axes as input. In total, nine features were extracted for each sensor type of each sensor position. For the classification of the desired activities, the SVM classifier was applied. In the final step, the classifier decisions of each available sensor type were fused. The decision level fusion was based on majority voting. In case of equal votes, the decision of the sensor source was chosen, which achieved the highest sensitivity regarding the predicted activity.

3.6.4 Experiments and Evaluation

Table 3.11 comprises the number of instances for each activity regarding the three considered databases and the total amount of instances. For the subsystems of wrist, chest, and shoe 28 subjects were available. For the subsystems of hip, 33 subjects were available. Two experiments were conducted.

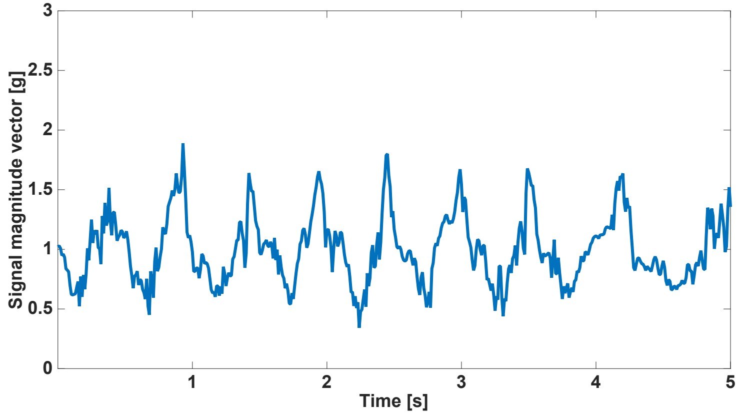
- **Experiment 1: Individual databases**
  The activity recognition algorithm was applied to the individual

Table 3.11.: Number of available instances for database fusion regarding the three databases PAMAP2, DaLiAc, and USC-HAD as well as total number of instances. Table contains activity-specific information.

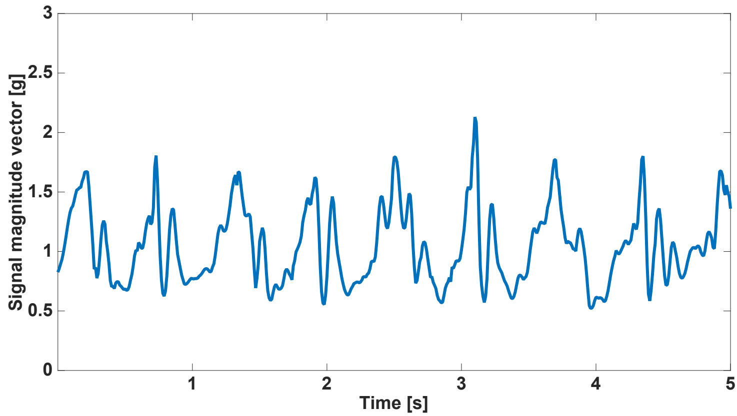
Activity	PAMAP2	DaLiAc	USC-HAD	Total
Static Walking	1131	678	1745	3554
Climbing stairs	475	1006	758	2239
Running	441	288	812	1541
Jumping	194	0	348	542
Vacuuming	0	0	208	208
Bicycling Rope	347	222	0	569
jumping	326	916	0	1242
	95	123	0	218

databases. The performance of the algorithm was determined by the sensitivities regarding the eight selected activities and the balanced accuracy (section 2.4.8). The performance measures were computed by LOSO-CV. The linear kernel was used in the SVM classifier and the corresponding cost parameter C was optimized by grid search ($C \in \{2^N\}, N \in \{-10, ..., 10\}$). The parameter optimization was performed in an inner LOSO-CV.

- **Experiment 2: Fused database**
  The three databases were combined using the proposed database fusion algorithm. The developed HAR system was applied to the fused database. The same performance measures were computed as in experiment 1.

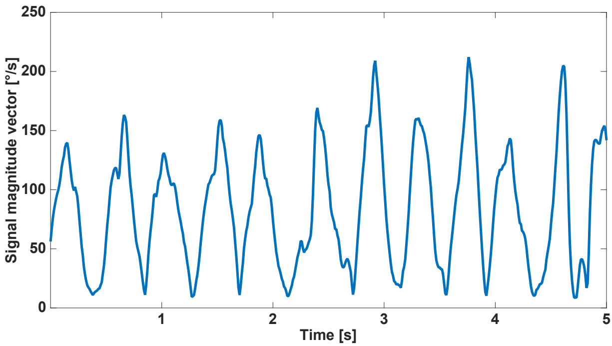


(a) DaLiAc.

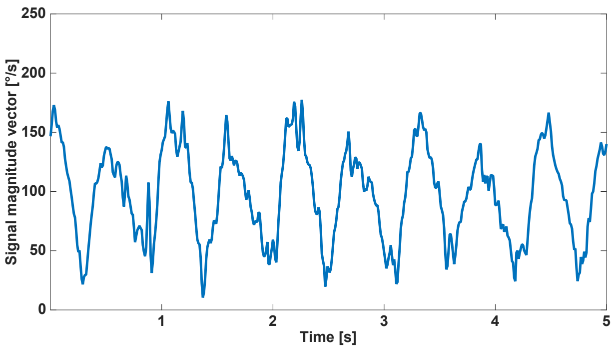


(b) USC-HAD.

Figure 3.16.: SMV of climbing stairs based on accelerometer data acquired on hip. Activity was taken from DaLiAc (a) and USC-HAD (b).



(a) DaLiAc. Sensor data acquired on shoe.



(b) PAMAP2. Sensor data acquired at ankle.

Figure 3.17.: SMV of bicycling based on gyroscope data. Activity was taken from DaLiAc (a) and PAMAP2 (b).

Table 3.12.: Sensitivity [%] of eight activities regarding the four database configurations. Balanced accuracy [%] is further given as mean of sensitivities.

Activity	Fusion	DaLiAc	PAMAP2	USC-HAD
Static	99.6 ± 0.9	99.5 ± 1.1	98.9 ± 0.7	**99.9 ± 0.5**
Walking	97.9 ± 3.8	**99.1 ± 2.3**	92.5 ± 7.9	95.6 ± 4.4
Climbing stairs	49.2 ± 24.5	64.7 ± 20.9	**78.8 ± 8.2**	36.6 ± 25.8
Running	86.8 ± 25.0	-	87.7 ± 8.3	**91.6 ± 22.3**
Jumping	85.2 ± 31.5	-	-	**87.8 ± 28.9**
Vacuuming	86.4 ± 15.7	86.2 ± 15.4	**88.3 ± 5.1**	-
Bicycling	95.2 ± 9.6	**100.0 ± 0.0**	89.5 ± 7.1	-
Rope jumping	97.2 ± 7.1	**100.0 ± 0.0**	86.1 ± 10.5	-
Mean	87.2	**91.6**	88.8	82.3

3.6.5 Results

Table 3.12 contains the sensitivities of the eight activities regarding the four database configurations. The best balanced accuracy regarding the isolated databases was achieved using DaLiAc. Considering the fused database, a balanced accuracy of 87.2 % was reached.

3.6.6 Discussion

In this section, three publicly available datasets were fused in order to increase the number of instances, which can be used to train machine learning algorithms. A proposed HAR algorithm was evaluated on the individual, isolated databases as well as the fused database. The concept of database fusion and the achieved results of the HAR approach are discussed in the following paragraphs.

The discussion is structured as follows. First, the modifications compared to the original publication [Schu 14b] are summarized. Second, the performance of the proposed HAR system is discussed. Third, the database fusion strategy is assessed. Fourth, future work is summarized.

Modifications Compared to Original Publication

Compared to the original publication [Schu 14b], one modification was made. Since the application of the proposed feature set in Table 2.1 to activity recognition challenges was proven in section 3.3 and 3.4, these features were also used in the proposed HAR system. The updated feature set increased the accuracy values by 1.2 %, 4.0 %, and 0.1 % considering PAMAP2, DaLiAc, and USC-HAD, respectively.

Human Activity Recognition System

The general practicability of the proposed HAR system to the mentioned application was proven with balanced accuracy values above 82.3 % (Table 3.12). The proposed architecture further allowed a flexible integration of new sensor types, which were provided by different studies. The balanced accuracy increased from 82.3 % to 88.8 % to 91.6 % using one, three, and four sensors of the individual databases. A higher number of sensors seemed to increase the performance of the HAR, which coincides with the findings in literature [Liu 12].

The six activities available in DaLiAc were classified with a sensitivity of above 86.2 % except for climbing stairs. Climbing stairs might be similar to walking, if the direction is removed by computing the signal magnitude vector (Equation 2.6). Future research should include features based on the three accelerometer axes in order to include the direction of movement. The second best balanced accuracy across the three isolated databases was achieved using PAMAP2. Although the balanced accuracy was 2.8 % lower than using DaLiAc, applying the method to PAMAP2 data resulted in lower standard deviations in the sensitivities of climbing stairs and vacuuming. This might indicate a more similar execution of the activities among the subjects. The worst balanced accuracy across the three isolated databases was achieved using USC-HAD. The main reason is the low sensitivity of climbing stairs (36.6 %) meaning that the sensor on the hip was not able to detect instances of ascending and descending stairs. The common set of activities among all isolated databases included static, walking, and climbing stairs (Table 3.10). The achieved mean sensitivities regarding climbing stairs differed from 36.6 % (USC-HAD) to 78.8 % (PAMAP2). Compared to climbing stairs, the mean sensitivities regarding static and walking differed less than 7 % across the isolated databases. Thus, the proposed HAR system was able to detect static and walking instances, even from different studies. Nevertheless, the applicability of the example algorithm to detect instances of climbing stairs remains unknown due to large discrepancy of the individual results.

Table 3.12 indicates that the decision level fusion approach could deal with the inter-study subject variability regarding static, walking, bicycling, and rope jumping activities due to low standard deviations. It can be assumed that the proposed algorithm can deal with different walking styles of subjects, even if the subjects were acquired in different studies. In DaLiAc, bicycling was performed indoor on an ergometer. In PAMAP2, cycling was performed on a real bike outside. The high mean sensitivity of bicycling using the fused database further indicates, that the proposed HAR can also deal with activities in different locations, indoor and outdoor. A further improvement might be achieved by considering e.g. location data. Nevertheless, higher standard deviation values for the sensitivities were achieved for climbing stairs, running, jumping, and vacuuming using the fused database. Due to a low mean sensitivity and a high standard deviation regarding climbing stairs, it can be concluded that the proposed HAR is not applicable to ascending and descending stairs. Since the performance was only adequate considering the PAMAP2 database, it can be assumed that the sensor data acquired in DaLiAc and USC-HAD negatively influenced the trained system. One reason might be that the activity was performed in different buildings, which vary in terms of number of steps or number and length of resting places. If the resting place is long, subjects are forced to walk in a straight line and not upstairs or downstairs. In the future, a person-specific training might reduce the high standard deviation as proposed in [Bao 04].

Database Fusion Strategy

The general applicability of the proposed database fusion strategy was shown for one example algorithm. The amount of training data for machine learning techniques could be increased without the acquisition of new data. The number of subjects raised from minimum nine (USC-HAD) to maximum 33 subjects. Although the number of subjects could be increased, the mean age was comparable across the individual databases. In the future, more databases should be considered which include more age categories. The number of activities could be increased from minimum five (USC-HAD) to eight. Activities which were not available in the original database can now be investigated. The number of sensors could be increased from minimum one (USC-HAD) to four. Nevertheless, fusing multiple databases might not guarantee an automatic improvement of the system performance. The HAR system was applied to the fused database and a balanced accuracy of 87.2 % was achieved, which was 4.4 % lower than the best performance considering individual databases. The reason for the lower balanced accuracy might be the higher subject variability by merging individual databases. Thus, fusing isolated databases could be used to identify certain problems of current HAR systems. In the case of high subject variability person-specific training methods should be developed as proposed by [Bao 04].

Future Work

In the future, the proposed database fusion system should further be extended. In the mentioned application, information about the direction of movement was not required and thus the SMV computation was suitable. If all axes of a sensor are considered, a further step in the data preparation might include an automatic coordinate system transformation across various databases. An extension of the pipeline could further include an automatic classification of the sensor type as proposed in [Scho 17] instead of a manual selection.

3.7 Energy Expenditure Estimation

3.7.1 Overview

Activities like walking can further be classified regarding the intensity, with which the activity was performed [Ains 93, Ains 00, Ains 11]. Activity intensity is usually defined by energy expenditure. Current approaches of energy expenditure estimation in literature mainly applied feature level fusion to accelerometer data followed by regression analysis [Bout 94]. The application of the methods might be cumbersome in case of a degradation of a set of sensors and hardware failures. In addition, systems have to be completely re-trained, if sensors are added or removed. Thus, the purpose of this section is to develop and implement an IMU-based system for energy expenditure estimation based on decision level fusion using multiple sensor types. The proposed algorithm was evaluated against indirect calorimetry, whose fundamentals were explained in section 2.1. The conducted study was already introduced in section 3.2.2. The proposed algorithm was originally published in [Schu 14a] and was modified in this work. The modified parts are highlighted in the discussion.

This section is structured as follows. First, the proposed algorithm for energy expenditure estimation is introduced. Second, the conducted experiments are described. Third, the results are given. The section concludes with a discussion of the results.

3.7.2 Methods

Figure 3.18 shows the pipeline of the proposed energy expenditure system. Four sensor sources provided accelerometer and gyroscope data acquired on hip and shoe. The data of the four sensor sources were processed independently. The processing steps included preprocessing, segmentation, feature extraction, and regression. The predicted energy expenditure values of each single sensor source were fused in the final step (decision level fusion).

The preprocessing step included calibration and synchronization, which were already described in section 3.3. As pointed out in section 1.5.2, usually IMU data were evaluated at the end of the activity stages because the oxygen consumption reached a steady state. Thus, the last three minutes of the six-minute period of each speed level was considered for further processing motivated by [Motl 09]. Figure 3.19a shows an example of the steady state segmentation. Non-overlapping sliding windows were further applied to the IMU. The window width was set to 30 s motivated by [Bout 94, Liu 12]. As previously mentioned, the sampling rate of the ground truth system was 0.2 Hz. Thus, one energy expenditure value was provided each 5th second. In order to get the ground truth energy expenditure of the considered 30 s interval, the mean of the corresponding six samples in the ground truth data were computed. In each previously determined windows, the features comprised in Table 2.1 were computed. In total, 28 features were extracted for each of the four sensor systems. Figure 3.19b shows an example of the 10th percentile of the gyroscope data (sagittal plane) on the shoe during running on a normal treadmill. MLR was applied in order to predict the expended energy based on IMU data (section 2.4.7). The predicted MET values of the four systems were fused in the final step. Therefor, the mean of all predictions was computed.

3.7.3 Experiments and Evaluation

The proposed energy expenditure estimation system was evaluated on the previously described EnEx database (section 3.2.2). Mean, standard deviation, and range of the ground truth expended energy are comprised

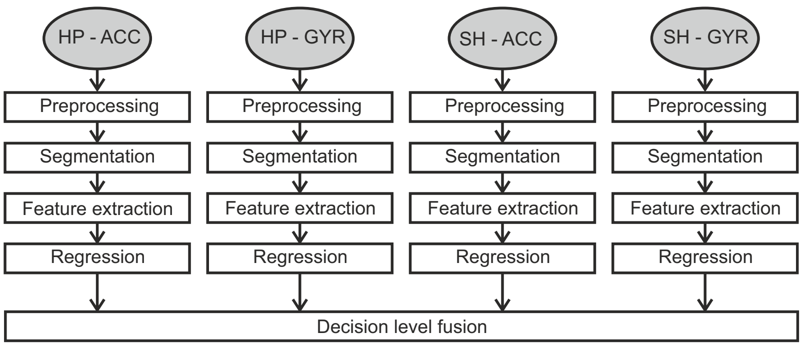
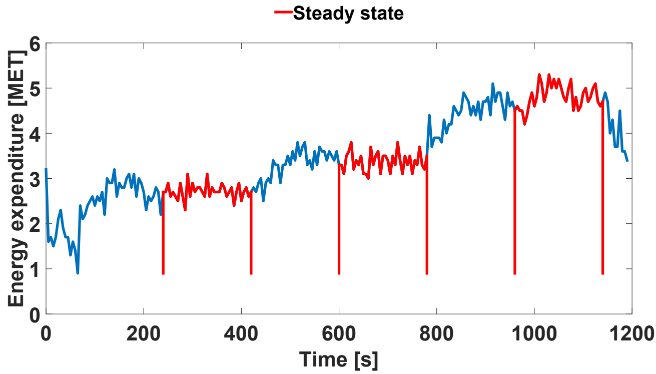
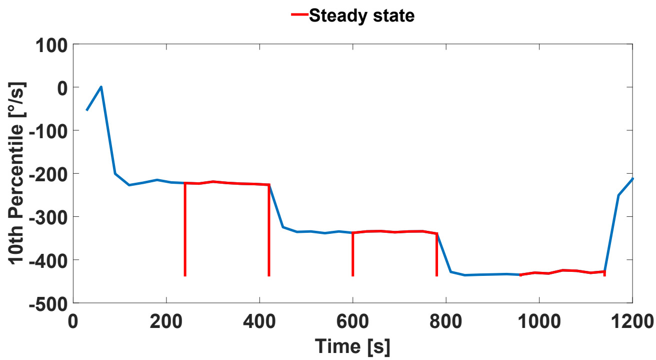


Figure 3.18.: Pipeline of energy expenditure estimation system (adapted from [Schu 14a]). Preprocessing, segmentation, feature extraction, and regression were performed for the four sensor sources HP-ACC, HP-GYR, SH-ACC, and SH-GYR. HP, SH, ACC, and GYR denote hip, shoe, accelerometer, and gyroscope, respectively. The predicted energy expenditures of each sensor source were fused in the final step (decision level fusion).



(a) Ground truth energy expenditure determined by Master Screen® CPX.



(b) 10th percentile signal based on gyroscope data of shoe (sagittal plane).

Figure 3.19.: Illustration of steady state segmentation in ground truth system (a) and 10th percentile signal (b). The steady states corresponding to the three speed levels [3.2, 4.8, 6.4] km/h are between subsequent vertical, red lines.

in Table 3.13. For each speed level, 120 segmented windows were available for evaluation. Two experiments were conducted.

- **Experiment 1: Overall performance**
  The overall performance of all single sensor systems as well as all two, three, and four sensor combinations were compared. In total, 15 sensor configurations were investigated, which are given

Table 3.13.: Mean, standard deviation (SD), and range of expended energy mea-sured by indirect calorimetry with respect to speed levels.

Speed level [km/h]	Mean [MET]	SD	Range [MET]
3.2	3.23	0.47	2.37 - 4.72
4.8	3.93	0.55	2.90 - 5.52
6.4	5.58	0.92	4.45 - 8.23

in Table 3.14. LOSO-CV was applied to determine the RMSE. Boxplots of the RMSE were further provided for each system in order to investigate the distribution of the RMSE across the subjects. The fitted model included constant, linear, and quadratic terms (section 2.4.7). In order to further compare the proposed decision level fusion approach to algorithms from literature, the overall mean correlation and RMSE were further computed for specific sensor configurations.

- **Experiment 2: Speed-dependent performance**
  The Bland-Altman plot was used in order to investigate the speed-dependent performance (section 2.4.8). The sensor combination with the lowest mean RMSE was considered.

3.7.4 Results

Figure 3.20 shows the boxplots of the RMSE regarding the 15 possible sensor configurations, which are comprised in Table 3.14. The lowest median RMSE of 0.40 MET was achieved by fusing all four sensor systems (configuration 15). Fusing both accelerometer types of hip and shoe achieved a mean RMSE of 0.42 MET. This combination further achieved the lowest interquartile range of 0.12 MET.

Table 3.14.: Possible sensor configurations considering four sensor sources HP-ACC, HP-GYR, SH-ACC, and SH-GYR. HP, SH, ACC, and GYR denote hip, shoe, accelerometer, and gyroscope, respectively. Each column comprises selection of sensor systems per configuration. Selected systems are indicated by '✓'.

Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
HP-ACC	✓				✓	✓	✓				✓	✓	✓		✓
HP-GYR		✓			✓			✓	✓		✓	✓		✓	✓
SH-ACC			✓			✓		✓		✓	✓		✓	✓	✓
SH-GYR				✓			✓		✓	✓		✓	✓	✓	✓

The lowest mean RMSE of 0.54 MET was achieved by combining both sensor types of the shoe position and the accelerometer of the hip position (configuration 13). A mean RMSE of 0.70 MET was achieved fusing all four sensor systems. A mean correlation of 0.76 and 0.94 were achieved by HP-ACC (configuration 1) and SH-ACC/SH-GYR (configuration 10), respectively.

Figure 3.21 depicts the Bland Altman plot of sensor configuration 6 (HP-ACC and SH-ACC). A mean difference of -0.06 MET as well as 95 % limits of agreement of -1.5 MET and 1.4 MET were achieved.

3.7.5 Discussion

In this work, an energy expenditure estimation algorithm was developed, implemented, and evaluated, which was based on decision level fusion. The achieved results are discussed in the following paragraphs.

The discussion is structured as follows. First, the modifications compared to the original publication [Schu 14a] are summarized. Second, the performance of individual sensor sources is compared. Third, the

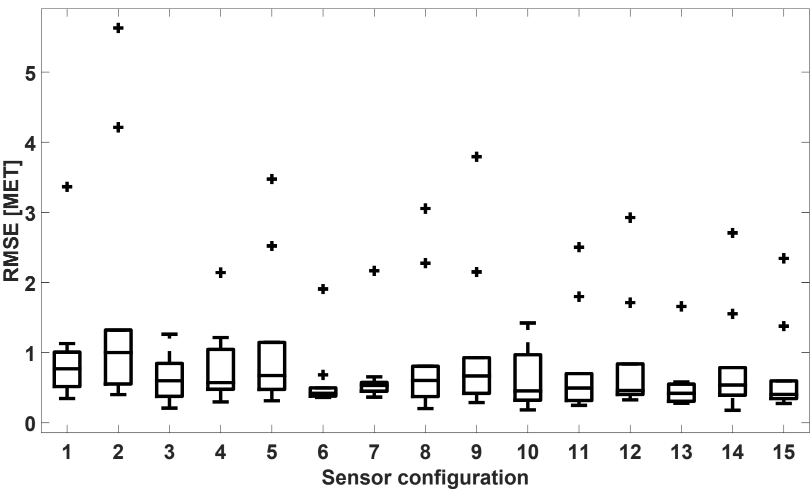


Figure 3.20.: Boxplots of RMSE values. Each boxplot considered the single achieved RMSE values of the LOSO-CV trials. RMSE is refered to the 15 sensor configurations comprised in Table 3.14.

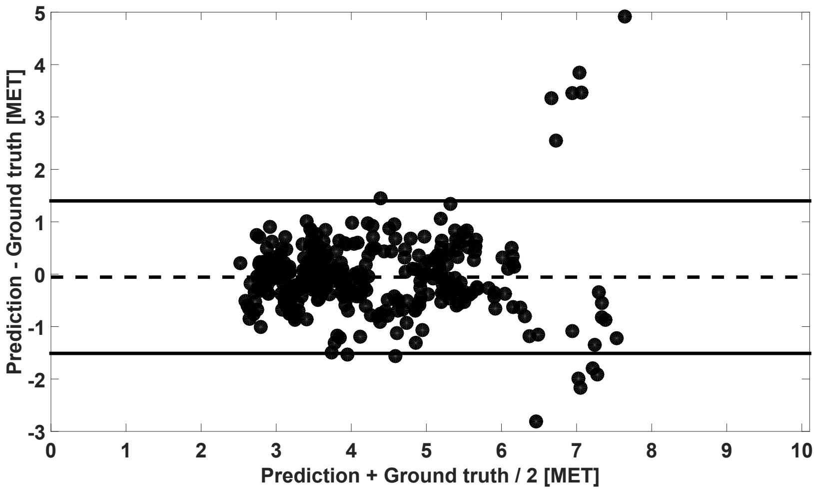


Figure 3.21.: Bland Altman plot for proposed approach (decision level fusion of HP-ACC and SH-ACC).

performance of decision level fusion is discussed. Fourth, the findings are compared to literature. Fifth, examples for future work are provided.

Modifications Compared to Original Publication

Compared to the original publication [Schu14a], several modifications were made. The original feature set was exchanged by the feature set listed in Table 2.1. Instead of applying and comparing three different regression algorithms, MLR was investigated in more detail. All in all, the MAE slightly decreased by 0.1 MET. The reason for the worse performance might be that only MLR was performed in this thesis compared to a combination of Support Vector Regression, CART, and MLR in [Schu14a].

Individual Sensor Sources

Regarding the individual sensor sources, the lowest median RMSE was achieved by SH-GYR (Figure 3.20, configuration 4). The reason might be that only walking was acquired in the study. Walking mainly includes rotational movements performed by the lower extremities. Nevertheless, SH-ACC achieved the second lowest median RMSE (configuration 3). In comparison to SH-GYR, the interquartile range was lower and no outlier was present. Thus, the SH-ACC system might perform better for an unseen subject since the subject variability is reduced compared to SH-GYR. Using accelerometer or gyroscope of the hip sensor alone achieved higher median RMSE values compared to the corresponding sensor sources from the shoe. Thus, only using the hip sensor might not be sufficient to estimate the expended energy for all subjects and a subject-dependent system should be preferred.

Decision Level Fusion

Although the two sensor types of the shoe achieved the best single performance, combining the decisions of the two accelerometer sensor types achieved the lowest median RMSE regarding two sensor sources. In addition to a lower median RMSE, the interquartile range was further reduced from 0.57 (SH-GYR) to 0.12 (HP-ACC and SH-ACC). Thus, fusing acceleration information from different sensor positions increased the performance of energy expenditure estimation systems. Nevertheless, the median of RMSE increased for the sensor configurations 5, 8, and 9 compared to SH-GYR. Thus, a general improvement from one to two sensors was not achieved. Using three sensors did not reduce the median RMSE. A small reduction of 0.016 MET was achieved by using all four sensors.

All in all the median RMSE varied from 0.4 MET to 1.0 MET across all sensor configurations. In relation to the range of the ground truth values [2.37 MET - 8.23 MET], the median RMSE was below 17.1%. This shows the general applicability of the proposed pipeline independent on the chosen sensor configuration. Nevertheless, due to the good trade-off between low median RMSE and low interquartile range, the combination of both accelerometer types of hip and shoe is proposed (Figure 3.20, configuration 6).

According to the Bland Altman plot in Figure 3.21, no systematic error was found using the previously proposed sensor configuration (HP-ACC, SH-ACC). The performance of the system decreased at the higher border of the ground truth speed values. Nevertheless, the instances outside the 95 % limits of agreement for higher speed levels mainly came from the same subject. The subject had the highest weight across all subjects. The high weight might result in accelerometer signals reaching the saturation state, which might affect the regression performance. Thus, the proposed system was not able to deal with subjects of higher weights. In order to further improve the system, anthropometric information should be included in the regression model. A further improvement could also be the application of a high range accelerometer in order to avoid saturation state of sensors. The new sensor could be included in the decision level fusion approach without retraining the already existing system.

Comparison to Literature

Since a stationary ground truth system was used, the set of activities was restricted to walking on a treadmill. Research work which also considered different walking bouts on a treadmill included [Bout 94, Motl 09, Leen 03]. The authors in [Bout 94] placed one triaxial accelerometer in an elastic waist belt. A correlation of 0.96 was achieved only considering walking activities. The proposed method in this work achieved a lower mean correlation of 0.76 using only an accelerometer on the hip. Three reasons were identified for the worse performance of the proposed approach. First, the distribution regarding the anthropometric data of the subjects varied across the studies. In [Bout 94], the subjects were in average 23.5 ± 1.8 years old and their mean weight was 68.6 ± 9.9 kg. The subjects, who were evaluated in this thesis, were older (49 ± 12 years) and heavier (80.7 ± 14.6 kg). Second, the aim in [Bout 94] was a statistical analysis of accelerometer features as predictors for regression. Thus, the data of all subjects were used in order to estimate the regression coefficients. In this thesis, the aim was to evaluate the performance of the proposed system on unseen subject data. Therefore, the correlation was given as an average value across all LOSO-CV trials. For one subject a high error was achieved especially for higher speed levels (Figure 3.21), which might have an influence on the correlation. Third, accelerometer data was acquired during five different speed levels. The higher number of speed levels might result in a better accuracy of the regression equation. In order to further improve the proposed method, anthropometric data such as age should be included in the feature set.

The authors in [Motl 09] used the same three walking bouts and achieved a correlation of 0.89. The age and weight distribution was closer to the distribution in this work (40.9 ± 11.4 years and 72.5 ± 15.0 kg), which reinforce the argument that the correlation drops with an increased age and weight. In [Motl 09], cut-points based on the acceleration counts were further proposed. The cut-point procedure enables the classification of activities regarding intensity levels. The same cut-off points could be used in the proposed approach to classify activities regarding intensity levels.

In [Leen 03], the Tritrac and CSA systems were attached to the hip and achieved a RMSE of 0.46 and 0.53 MET, respectively. In this thesis, the HP-ACC system achieved a mean RMSE of 0.98 MET. The reason for the worse performance might also be the different age and weight distribution (23.7 ± 3.9 years and 67.2 ± 13.5 kg).

All in all, the investigations so far were based on comparing the performance of HP-ACC to the results from literature. Fusing different sensor sources increased the performance. The 2-sensor system SH-ACC/SH-GYR achieved the highest correlation of 0.94 which is comparable to the achievements found in [Bout 94] and even higher compared to [Motl 09]. Adding HP-ACC to the two shoe sensor sources resulted in the highest mean RMSE of 0.54 MET which is comparable to the achievements found in [Leen 03]. The main advantage of decision level fusion might be that the influence of anthropometric data to the performance of energy expenditure systems is reduced. Future systems, which do not want to consider anthropometric data, might have to include more sensor sources. In addition to IMU sensors, other sensor types could be integrated such as a ventilation sensor [Liu 12]. The proposed decision level fusion enables an easy integration of such a new sensor type without the need of retraining the current system.

In the following paragraph, the proposed system is compared to research work using more advanced machine learning algorithms such as Artificial Neural Networks and Random Forest regression. For the comparison, the fusion of all four sensor sources was considered which achieved a mean RMSE of 0.70 MET. Artificial Neural Networks were applied in [Stau 09] as well as in [Mont 16]. RMSEs of 1.22 MET and 2.16 MET were achieved, respectively. The authors in [Elli 14], applied Random Forest for first classifying an activity followed by predicting the expended energy. An accelerometer was placed around the wrist. The system achieved a RMSE of 1 MET. Although the performance in this thesis was better than in [Stau 09, Elli 14, Mont 16], a lower amount of activities were acquired in this work.

Future Work

In future work, the described decision level fusion approach should be applied to further activities in more realistic settings outside the laboratory. In order to acquire ground truth data, a portable indirect calorimeter has to be used like in [Elli 14]. Two evaluation phases are proposed.

First, the algorithm should be evaluated on data acquired in a controlled environment with activities similar to study A (section 3.2.1). It should be investigated, if the two sensor-setup is sufficient to perform a reliable estimation of the expended energy or if more sensors on different sensor locations should be integrated. In addition, activity-dependent regression models might be necessary.

Second, the approach should be evaluated during activities in a non-controlled environment without instructions of a supervisor. In order to deal with the large variety of captured activities, which are often unknown during regression model training, shallow architectures as the proposed decision level fusion approach might be combined with deep architectures. Deep learning methods might be applied to data of single sensor types and combined by decision level fusion. The proposed decision level fusion approach would further allow a hybrid system. Deep and shallow architectures could be compared for each sensor type separately and the best architecture could be used for further processing.

Chapter 4

Recognition of Soccer-Specific Activities

4.1 Introduction

In this chapter, three contributions regarding the recognition of soccer-specific activities and corresponding applications are presented. Soccer-specific activities included full-instep and side-foot kicks [Leva 98]. The goal was to recognize the two kick types by attaching IMU sensors to soccer players and applying machine learning techniques. Full-instep kicks can further be classified based on the achieved ball speed. Thus, an IMU-based ball speed estimation is further presented. The HAR algorithms were integrated in an automatic, sensor-driven video summary application.

This chapter is structured as follows. First, three studies are introduced, in which IMU data were collected for training and testing the proposed machine learning techniques. The first two studies were used to acquire data for the proposed kick classification approaches. The data in the two studies were acquired during training exercises (study A) as well as during an 11-a-side game (study B), respectively. Both studies were published in [Schu 15]. The third study was used to acquire data for the proposed ball speed estimation approach during training exercises (study C). The data collection was performed during the Master Thesis of Carolin Jakob [Jako 16] (supervised by the author of this dissertation). Second, the two proposed kick classification routines were described (contribution 6). The algorithms were published in [Schu 15] and [Schu 16].

Third, the proposed ball speed estimation algorithm is presented (contribution 7), which was published in [Schu 16]. Fourth, the mentioned video summary application is briefly introduced, which integrated parts of the previously mentioned algorithms (contribution 8). A patent application was filed encompassing parts of the algorithmic approaches and the final video summary system [Kirk 16]. An extension of the system was further published in [Schu 16].

4.2 Data Collection

4.2.1 Study A: Training Exercises

Subjects

In the conducted study, eleven male amateur players participated (age 29.6 ± 9.2 years, height 182.3 ± 6.3 cm, weight 77.7 ± 9.68 kg, mean $\pm$ standard deviation). All subjects gave written informed consent about their participation.

Hardware Setup

Each player was equipped with one customized system per leg [1]. Each system comprised a sensor unit and a storage unit (Figure 4.1). The sensor unit consisted of a MPU-6050 motion processing unit (InvenSense Inc., Sunnyvale, CA) [MPU 11] including a triaxial accelerometer and a triaxial gyroscope. The accelerometer range was set to $\pm$ 16 g. The gyroscope range was set to $\pm$ 2000 $°/s$. The customized system of the manufacturer provided already calibrated values in g and $°/s$. The sampling rate was adjusted to 1000 Hz. The sensor unit was placed in a cavity of an adidas F50 adizero soccer shoe. The cavity ensured a standardized sensor position and was robust against additional vibrations. The storage unit consisted of a MicroSD card integrated in a shin guard. Sensor and storage unit were connected by cable.

Description of Exercises

Each player had to perform 14 exercises which are comprised in Table 4.1. Exercise ID, exercise name, number of trials, and the corresponding event leg are given. Event leg was defined as the leg, with which the

[1] Developed by the University of Applied Sciences, Zweibrücken.

kick was performed. The other leg was defined as supporting leg. The exercises contained basic soccer-specific activities with and without the ball. Exercises 1 - 13 were performed outdoors on a grass pitch. Figure 4.2 shows the pitch setup including field dimensions and persons who are involved in the data collection session. Exercise 14 was performed indoors. All 14 exercises are described in Figures A.1 - A.11 in the appendix.

Protocol

The single steps of the protocol are described below.

1. Sensor nodes were powered on.
2. Player clapped heels together. This procedure was needed for offline synchronization. The synchronization method is described later in this chapter.
3. Exercises listed in Table 4.1 were performed. Exercise 1 - 7 were performed in the given order and can be seen as a warm up. The execution sequence of the exercises 8 - 13 were randomized in order to reduce the influence of the ordering on the algorithm performance. Exercise 14 was last executed.
4. Player clapped heels together.
5. Sensor nodes were powered off.
6. Sensor data were stored on a PC for offline processing.

A study supervisor accompanied the players and labeled the start and end of each exercise via smartphone app. During the data acquisition, the players were further filmed by a video camera.

4.2.2 Study B: 11-A-Side Game

In the conducted 11-a-side game, 17 male amateur players participated (age 29.6 ± 6.6 years, height 182.3 ± 6.4 cm, weight 80.6 ± 9.2 kg, mean $\pm$ standard deviation). All subjects gave written informed consent about their participation. The game took place in a stadium on the adidas site in Herzogenaurach, Germany. Only one player, who participated in the previous described study, was involved in the 11-a-side game.

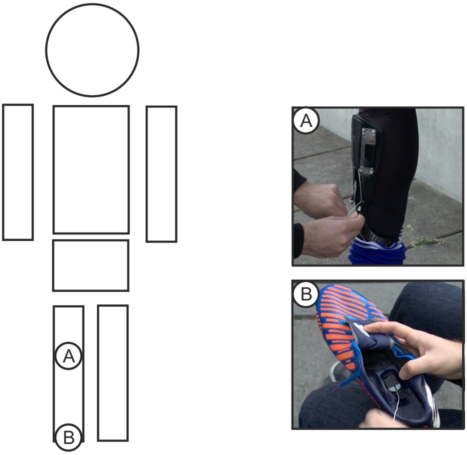


Figure 4.1.: Hardware setup used in study A (adapted from [Schu 15]). Storage unit integrated in shin guard (A). Sensor unit placed in cavity of an adidas F50 adizero soccer shoe (B).

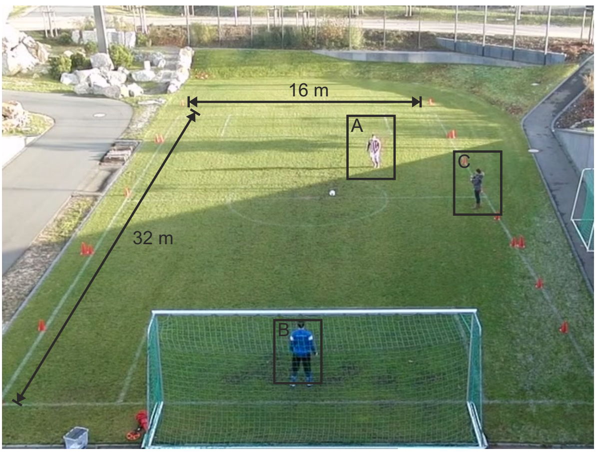


Figure 4.2.: Pitch setup with field dimensions. Rectangles A, B, and C show the player equipped with sensors, player without sensors acting as teammate/opponent/goal keeper, and study supervisor, respectively.

Table 4.1.: List of exercises. ID, name of exercise, number of trials, and considered event leg are given. Event leg is defined as the leg, with which the kick was performed.

ID	Exercise	# Trials	Event Leg
1	Walking	1	-
2	Running	3	-
3	Side-stepping	2	-
4	Eight subsequent side-foot kicks	3	Left and right
5	Dribbling	3	Preferred
6	Slalom	3	Preferred
7	Full-instep kicks on stationary ball	3	Left and right
8	Slalom - side-foot kick - jogging	3	Preferred
9	Control with thigh - side-foot kick	3	Preferred
10	Control with chest - side-foot kick	3	Preferred
11	Dribbling - side-foot kick - dribbling	3	Preferred
12	Dribbling - side-foot kick - full-instep kick	3	Preferred
13	Dribbling - tackling - full-instep kick	3	Preferred
14	Interactive player test	1	Preferred

One player got injured and was not considered for further processing. Inertial sensor data of both shoes were not available for four players, e.g. due to sensor problems during the game. Since the proposed algorithms required information of both shoes, these players were not considered for further processing. Gamedata of 45 minutes were captured. No instructions were given to the players, how to behave on the pitch.

4.2.3 Study C: Ball Speed Measurements using Hawk Eye

Subjects

In the conducted study, ten male amateur players participated (age 26 $\pm$ 2 years, height 180 $\pm$ 4 cm, weight 80 $\pm$ 5 kg, mean $\pm$ standard deviation). The subjects played in the current or past season in a team attending the 'Landesliga' (Bavarian sixth league) or a higher league. All subjects gave written informed consent about their participation.

Hardware Setup

Each player was equipped with nine sensor units (miPod, Portabiles, GmbH, Erlangen, Germany) [Blan 14]. Four on the left and right lower extremity (thigh, shank, ankle, heel) and one on the pelvis. In this thesis, only sensor data of the right ankle sensor was considered (Figure 4.3) motivated by [Zwic 12]. The sensor unit consisted of a MPU-9150 motion processing unit including a triaxial accelerometer and a triaxial gyroscope (InvenSense Inc., Sunnyvale, CA) [MPU 13]. The accelerometer range was set to $\pm$ 16 g. The gyroscope range was set to $\pm$ 2000 $°/s$. The sampling rate was adjusted to 1000 Hz. Accelerometer and gyroscope data were stored on a NAND flash memory inside the miPod. The IMU was placed in a strap two fingers above the lateral malleolus in the sagittal plane.
The Hawk-Eye system was used as reference system (Hawk-Eye Innovations, Basingstoke, UK) [Sher 01]. Sixteen cameras were placed around the pitch capturing the scene. Each camera acquired 100 frames per second. In each frame of each camera, the ball was segmented using among others the information of size and shape of the ball. The corresponding 3-D position of the ball was estimated from at least two different cameras. A ball-flight-path was predicted from the 3-D ball position enabling the computation of the ball speed. A calibration had

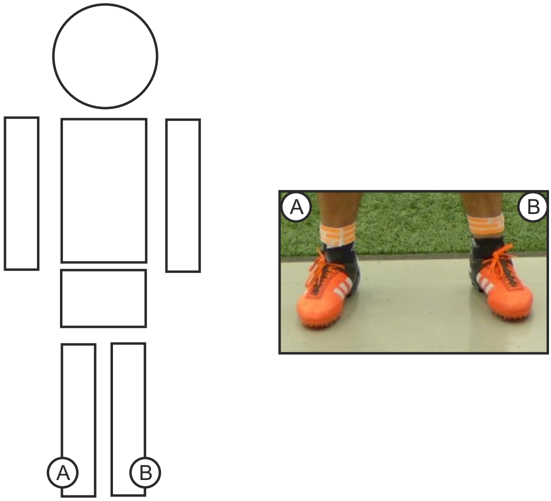


Figure 4.3.: Hardware setup used in study C (adapted from [Jako16]). IMU sensor was placed in strap two fingers above the lateral malleolus in the sagittal plane.

to be performed before the actual acquisition defining among others the position and field of view for each camera and anchor points.

Description of Exercises

Each player had to perform four exercises including a full-instep kick. The kicks were executed towards an empty goal from a distance of minimal 11 meters. In the first exercise, the player was asked to perform a full-instep kick on a **stationary ball**. The approach angle and amount of steps were chosen freely by the player. In the second exercise, the player executed a full-instep kick after a **dribbling sequence** (Figure 4.4). The approach angle was approximately $45°$. At least three ball contacts were required before the kick. The speed of dribbling was chosen individually by the player. In the third exercise, the player **passed** the ball to a teammate (Figure 4.5). The teammate passed the ball back to the player who finished the exercise with a full-instep kick. In the fourth exercise, the player had to perform the **interactive player test** and finish the test with a full-instep kick (Figure 4.6).

After executing a pre-defined and varying number of passes, an acoustic signal indicated that the player had to perform a full-instep kick. Each exercise was repeated nine times. To obtain a wide range of captured ball speeds, the players were told to perform the kicks three times softly, three times with a medium intensity, and three times as hard as possible.

(a) **(b)** **(c)** **(d)**

Figure 4.4.: Full-instep kick after dribbling sequence (adapted from [Jako 16]).

(a) **(b)** **(c)** **(d)**

Figure 4.5.: Full-instep kick after passing to teammate (adapted from [Jako 16]).

(a) **(b)** **(c)** **(d)** **(e)** **(f)**

Figure 4.6.: Full-instep kick after interactive player test (adapted from [Jako 16]).

Protocol

The single steps of the protocol are described below:

1. Sensor nodes were mounted with double adhesive tape on a firm metallic ruler.
2. Sensor nodes were powered on and the ruler was hit on a table. This procedure was needed for offline synchronization.
3. Four exercises were performed. The order of the second and third exercise was changed for every second subject.
4. Sensor nodes were mounted on the ruler and the ruler was hit on a table.
5. Sensor nodes were powered off.
6. Sensor data were stored on a PC for offline processing.

The data collection took approximately 70 minutes. A study supervisor accompanied the players and labeled start and end of each exercise via smartphone app. During the data acquisition, the players were further filmed by a video camera.

4.3 Kick Classification

4.3.1 Overview

Current IMU-based algorithms in soccer often applied a flat classification architecture (Figure 1.3a) to classify various soccer-specific activities such as dribbling, running, and kicking [Mitc13, Ahma15, Hoss17]. These systems often did not consider a NULL class, achieved a low performance for kicking, and were mainly evaluated on exercises or pre-segmented game data. In this section, two kick classification systems are introduced, which are able to distinguish between full-instep and side-foot kicks. Both approaches implemented an ARC, but differed in various components. The first approach explicitly modeled a NULL class containing various activities such as dribbling and running. The NULL class was considered in the preprocessing and classification step. The classification step was based on a hierarchical architecture. For the rest of this thesis, this algorithm is denoted as 'hierarchical approach'. The second approach relied on an expert-driven kick phase segmentation. For the rest of this

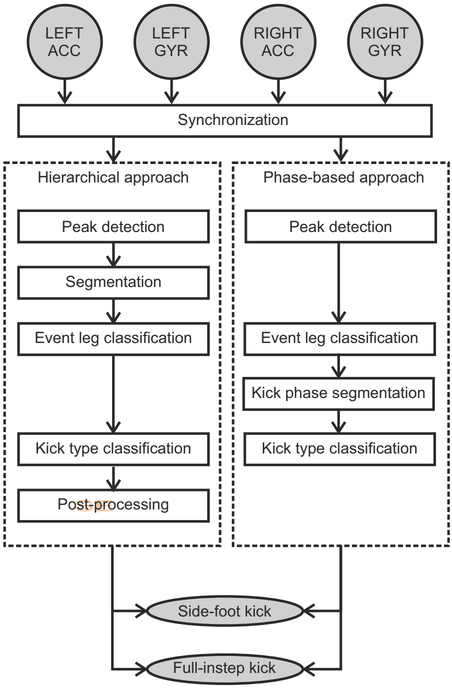


Figure 4.7.: Two proposed pipelines for kick classification: hierarchical approach (left), phase-based approach (right). ACC and GYR denote accelerometer and gyroscope, respectively.

thesis, this algorithm is denoted as 'phase-based approach'. Parts of the two algorithms were originally filed in a patent application [Kirk 16]. The two approaches were further published in [Schu 15] as well as [Schu 16] and are compared in this thesis. Both techniques were modified in this work. The modified parts are highlighted in the discussion.

This section is structured as follows. First, the two proposed kick classification systems are explained. Second, the conducted experiments and evaluation are described. Third, the results are presented. Fourth, the results are finally discussed.

4.3.2 Hierarchical Kick Classification System

Figure 4.7 (left) shows the pipeline of the proposed hierarchical kick classification system. The pipeline consisted of five major steps, which are explained in the following sections.

Preprocessing and Segmentation

The preprocessing step included synchronization and peak detection. The synchronization of the two IMUs was performed manually. As previously mentioned, each player was asked to clap both shoes together. The clapping of the shoes resulted in high peaks in the accelerometer signal of the sensors in both shoes. The peaks were used as common synchronization point.

The peak detection exploited the fact that biomechanical data of kicks are characterized by a sudden transition from low-frequency (before and after impact) to high-frequency (during impact) [Shin 09]. It was assumed that the kick candidates were represented as peaks in the high-frequency components of the accelerometer data. In order to detect the peaks, five steps were performed. First, a Butterworth high-pass filter was applied to the accelerometer data acquired for the left and right leg (section 2.4). High-pass filtering enabled the removal of low frequency content of NULL class activities such as running. The Butterworth high-pass filter was applied with an order L and a cutoff frequency f_c. Second, the SMV of the high-pass filtered signal was computed for both legs (Equation 2.6). Computing the SMV enabled the determination of movement intensity and removal of movement direction [Kara 06]. Third, the SMV signals were subtracted from each other and the absolute values of the resulting signal amplitudes were computed. It was assumed that during the execution of a kick, the signal intensity of the event and supporting leg differed. In the fourth step, thresholding was applied. Peaks below a certain threshold θ were removed from further processing. In the fifth step, peaks above the threshold were sorted in descending order regarding the corresponding amplitude. Peaks, which occurred 0.5 s before and after a detected, higher-ranked peak, were removed. It was assumed that no additional kick can occur during this period. The optimization of the parameters of the peak detection (L, f_c, θ) is described later in this section.

In the segmentation step, a window centered around each previously detected peak was defined. The authors in [Brop 07] measured a total duration of 790 ms for a kick. Thus, a window size of 1 s was chosen. Linear acceleration and angular velocity data of both left and right legs were considered for further processing.

Event Leg Classification

A side-foot-kick and a full-instep kick can be performed with the left or the right leg. It was assumed that the kick execution was bilaterally symmetrical to the sagittal plane. In order to avoid developing and training one system for each leg configuration, an early leg classification (event vs. supporting leg) was performed.

The leg classification procedure consisted of three steps. First, a symmetry transformation was performed with reference to the sagittal plane. In the symmetry transformation, the sign of the linear acceleration amplitude in mediolateral direction and the sign of the angular velocity amplitude in the frontal and transverse plane were changed (Figure 2.1). Second, the features listed in Table 2.1 were extracted. In total, 112 features were computed. A correlation-based feature selection routine was applied to reduce the number of features (section 2.4). In the third step, the selected features were used as input for a classifier. The classifier mapped left and right leg to event and supporting leg. The set of tested classifiers is described later.

Kick Type Classification

In the kick type classification step, a hierarchical recognition approach (Figure 1.3b) was applied considering two subsystems denoted as BASE and KICK. First, the BASE subsystem aimed at distinguishing between a merged side-foot/full-instep kick class and a NULL class. The NULL class included activities such as tackling, running, and side-stepping. Predicted NULL class instances were removed from further processing. Second, the KICK subsystem aimed at distinguishing between side-foot and full-instep kick. Both subsystems consisted of the same feature extraction and classification procedure as in the event leg classification.

Post-Processing

Four post-processing steps were applied in order to refine the decisions of the kick type classification routine. The steps were defined after the evaluation of the algorithm on the 11-a-side game and are motivated in the following paragraphs.

1. The peak detection routine determined more than one peak as kick candidate around the actual kick in the sensor data. Therefore, peaks in a 1-s window which shared the same classifier decision were considered as a peak chunk. In order to reduce the final number of peaks, the mean of the peaks in each chunk was considered for further processing.

2. It was assumed that a side-foot kick cannot follow a full-instep kick in a time interval of 2 s. Thus, side-foot kicks, which did not fulfill the requirement, were removed.

3. One major characteristic of a full-instep kick is that the vertical acceleration saturates before the actual ball contact (Figure 4.8a). Predicted full-instep kicks, for which a saturation phase of below 10 ms were detected, were reclassified as side-foot kicks. Predicted NULL instances, for which a saturation phase of above 10 ms were detected, were reclassified as full-instep kicks.

4. One major characteristic of a full-instep kick is that the angular velocity in the sagittal plane reaches a maximum before the ball contact [Reil 03, Bull 99]. Predicted full-instep kicks, for which a maximum angular velocity below 1000 $^\circ/s$ was reached, were classified as side-foot kicks.

4.3.3 Phase-Based Kick Classification System

Figure 4.7 (right) shows the pipeline of the proposed phase-based kick classification system. The pipeline consisted of four major steps, which are explained in the following sections.

Preprocessing

The preprocessing step included synchronization and peak detection. The synchronization of the two IMUs was performed manually and was already described in section 4.3.2.

The main difference of the peak detection to the method described in section 4.3.2 was that kick candidates were determined for both legs separately and fused in the final step. The proposed technique consisted of four main steps including sensor type selection, Butterworth high-pass filtering, SMV computation, and thresholding. These four steps were performed for the data of each leg, separately. Figure 4.8 shows an example of the single peak detection steps based on the linear acceleration of a full-instep kick. The sensor type selection offered the possibility to select the most appropriate sensor type ST for peak detection (accelerometer or gyroscope). High-pass filtering was performed in order to extract only high frequency parts in the sensor data, which were assumed to be candidates of kicks (Figure 4.8b). The Butterworth high-pass filter was applied with an order L and a cutoff frequency f_c. The SMV was computed (Equation 2.6) in order to remove the direction information (Figure 4.8c). Thresholding was performed in order to remove regions with a corresponding peak below a threshold θ. The positions of detected peaks from left and right leg were fused. The optimization of the parameters of the peak detection (ST, L, f_c, θ) is described later in this section. It was assumed that only side-foot kicks and full-instep kicks instances passed the peak detection step.

Event Leg Classification

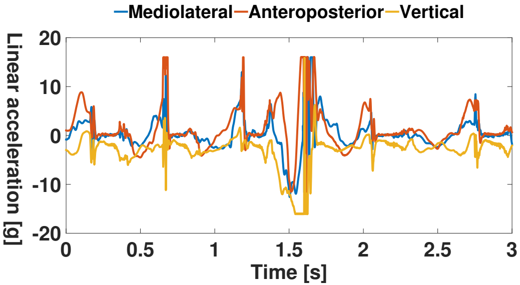
The main difference of the event leg classification to the method described in section 4.3.2 was the reduction of the computational complexity. The leg classification procedure consisted of three steps. First, a window was defined around the previously determined peaks. The start of the window was defined 0.1 seconds before and 0.4 seconds after the peak motivated by [Brop 07].

Second, a symmetry transformation was performed with reference to the sagittal plane. In the symmetry transformation, the sign of the linear acceleration amplitude in mediolateral direction and the sign of the angular velocity amplitude in the frontal and transverse plane were changed.

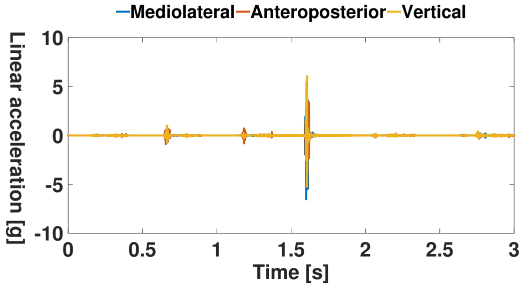
Third, the legs were classified into event and supporting leg. The discrimination was based on the SMA (Equation 2.9), which was computed for the accelerometer signal of both legs. The leg, which showed the highest SMA value, was assumed to be the event leg. Further processing was performed with sensor data of the event leg.

Kick Phase Segmentation

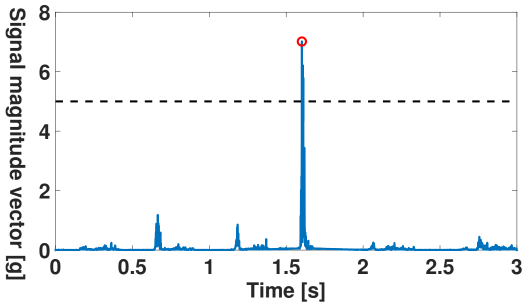
The kick phase segmentation aimed at dividing the signals of the event leg into phases driven by biomechanical expert knowledge. Instants and phases were extracted during the approach and kicking motion. The phases were defined for the full-instep kick. Instances, which did not pass the kick phase segmentation, were assumed to be side-foot kicks. Two instants (heel strike, mid swing) were computed based on the last step of the approach before the actual kick [Ounp 94, Duga 05]. Two instants (start of leg acceleration, ball contact) were computed based on the kicking motion [Nuno 02, Brop 07]. All instances were determined by signals in the sagittal plane motivated by [Dör 99, Nuno 02, Kell 07]. The definitions of the instants are given below. The processing was performed according to the given order. The kick phase segmentation is further illustrated in Figure 4.9.



(a) Linear acceleration.

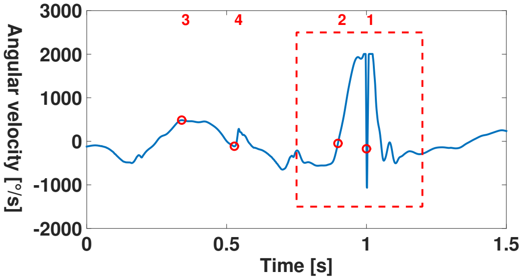


(b) Butterworth high-pass filtering ($L = 2$, $f_c =$ 250 Hz).

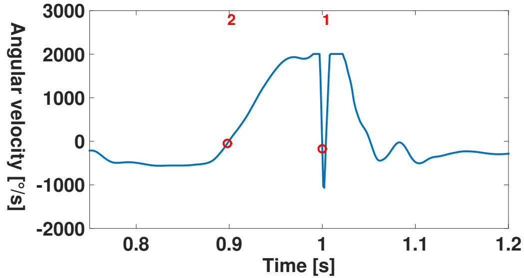


(c) SMV and thresholding ($\theta = 5$ g).

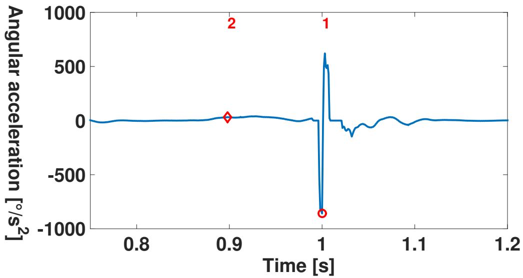
Figure 4.8.: Illustration of peak detection. Linear acceleration during full-instep kick (a), high-pass filtered signal (b), and SMV as well as chosen threshold (c) are given.



(a) Angular velocity. Red dashed rectangle indicates main part of kicking motion.



(b) Angular velocity of main part of kicking motion.



(c) Angular acceleration of main part of kicking motion.

Figure 4.9.: Illustration of kick phase segmentation. Angular velocity of approach and kicking phase in sagittal plane (a), angular velocity of kicking phase (b), and angular acceleration of kicking phase (c) are shown. The numbers indicate the instances ball contact (1), start of leg acceleration phase (2), mid swing (3), and heel strike (4).

1. **Ball contact**: first extremum of the angular acceleration motivated by [Nuno 06].
2. **Start of leg-acceleration phase**: last zero-crossing of angular velocity before ball contact motivated by [Nuno 02].
3. **Mid swing**: local maximum angular velocity before leg-acceleration motivated by [Amin 02].
4. **Heel strike**: next local minimum angular velocity after mid swing motivated by [Amin 02].

The segmented part between start of leg-acceleration phase and ball contact was the considered region of interest for feature extraction. The location of the heel strike was needed for the mentioned video summary application, which is described in section 4.5.3.

Kick Type Classification

The main difference of the kick type classification compared to the method described in section 4.3.2 was the reduction of the computational complexity. The kick type classification consisted of two steps. First, feature extraction was performed based on sensor data of the event leg. The leg-acceleration phase determined in the previous step was considered as region of interest for feature extraction. The absolute sum of the signal was computed for each accelerometer and gyroscope axis. The features reflected the different movement directions of side-foot and full-instep kicks [Leva 98]. In total, six features were extracted. The features were used as input for a classifier. The set of investigated classifiers is described later in this chapter.

4.3.4 Experiments and Evaluation

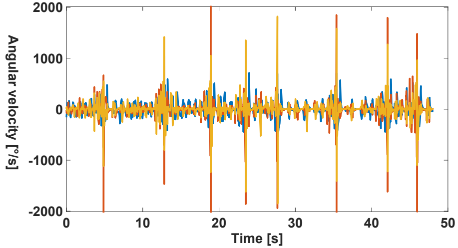
The two previously described kick classification systems were tested on data of a pre-defined protocol containing specific exercises (section 4.2, study A) and data of an 11-a-side game (section 4.2, study B). Data of study A was used to compare peak detection, event leg classification, and kick type classification routines of both algorithms separately. Pre-segmented sensor data was used as input. Data of study B was used to test a complete pre-trained kick classification system on non-segmented game data.

This section is structured as follows. First, the preparation of data from study A (exercises) and data from study B (11-a-side game) are summarized. Second, the optimization procedure of the peak detection routines are comprised. Third, the optimization of the classifier parameters is introduced. Fourth, the conducted experiments are described.

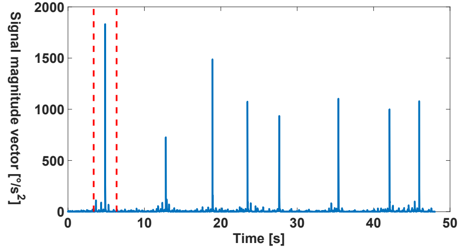
Data Preparation: Study A

The following paragraphs explain, how the sensor data were pre-segmented. The captured video data of the exercises were used to label 'full-instep kick', and 'side-foot kick' in exercises 8 - 14 (Table 4.1). The time point of 'full-instep kick' and 'side-foot kick' executions were selected and the corresponding sensor data around the time points were directly used for further processing.

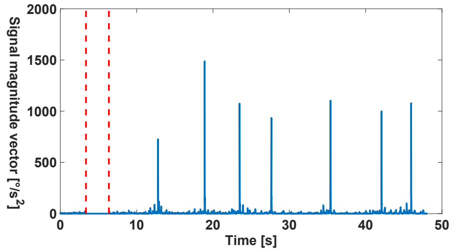
Sensor data of 'full-instep kick' and 'side-foot kick' based on exercises 4 and 7 (Table 4.1) as well as NULL class activities, which appeared in all exercises, were automatically segmented. NULL class activities contained dribbling, walking, jogging, moving back, side-stepping, running, and slalom (Table 4.1). The motivation of an automatic data segmentation procedure was twofold. First, an automatic technique reduced the time to label kick instances. Second, a manual labeling of NULL class instances was challenging, since it was not clear which part of the sensor data should be taken for segmentation. The automatic segmentation algorithm consisted of five steps based on the angular velocity of the event leg and is illustrated in (Figure 4.10). First, the angular acceleration was computed for each gyroscope axis. Second, the SMV of the angular acceleration was computed (Figure 4.10b). Third, the maximum amplitude of the SMV was determined. Fourth, segmentation borders were defined (Figure 4.10b). The start and end point was set 1.5 s before and 1.5 s after the time of the maximum amplitude, respectively. The corresponding linear acceleration and angular velocity signals between the segmentation borders were extracted and considered for further processing. Fifth, the SMV amplitudes between the segmentation borders were set to zero (Figure 4.10c). Step three to five were iteratively performed until a stopping criterion was reached. The stopping criterion was a pre-defined number of extracted instances dependent on the exercise, e.g. eight for exercise 'eight subsequent side-foot kicks'.



(a) Angular velocity.



(b) SMV of angular acceleration with segmented event (red dashed lines). First iteration of automatic procedure is shown.



(c) SMV of angular acceleration with segmented event (red dashed lines). Result after first iteration is shown.

Figure 4.10.: Illustration of automatic segmentation of instances. Angular velocity during exercise 'eight subsequent side-foot kicks' (a) and SMV of angular acceleration before (b) as well as after first iteration of automatic procedure (c) are given. Segmented instances were used to train and test proposed kick classification systems.

Data Preparation: Study B

The captured video data of the 11-a-side game were used to select the time points of 'full-instep kick' and 'side-foot kick' executions. No further automatic segmentation steps were applied compared to the data preparation of study A.

Peak Detection Parameter Optimization

The peak detection techniques of both the hierarchical and the phase-based kick classification aimed at determining time points which belong to candidates of full-instep and side-foot kicks. In the training phase of both peak detection algorithms, the pre-segmented sensor data described in the previous section was considered. The segmented data was grouped into two classes. Instances, which contained either full-instep or side-foot kicks, belonged to the true class. Instances from NULL class activities belonged to the false class.

In the hierarchical kick classification L, f_c, and θ had to be optimized. In order to reduce the number of parameters in the optimization procedure, L was set to 2. The optimization procedure of f_c and θ was performed in a LOSO-CV and is described below:

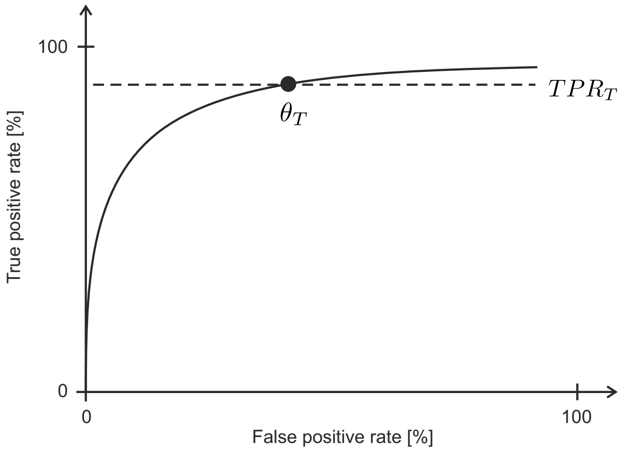
- **Cut-off frequency** f_c: it was assumed that the peaks occurred in higher-frequency bands. The magnitude of the Discrete Fourier Transform coefficients (Equation 2.4) was computed for each of the three accelerometer axes. Instances of full-instep and side-foot kicks were considered. The energy was computed by summing the squared magnitudes of coefficients. The determined cut-off frequency of each instance was set to the frequency up to which 90 % of the energy was reached. The final chosen cut-off frequency was the median of the single cut-off frequencies of the considered kick instances.

- **Threshold** θ: in order to optimize θ, the first three steps of the peak detection routine were applied to all instances (Butterworth filtering using the previously determined cut-off frequency, SMV computation, computation of absolute difference). The maximum amplitude of each instance was considered for further processing. A grid search was applied regarding minimum and maximum of the previously determined amplitudes. In every grid search iteration, True Positive Rate (TPR) and the False Positive Rate (FPR) were computed and added to a ROC curve as illustrated in Figure 4.11a. One point on the ROC curve represented one selected threshold θ. In order to determine the final parameter setting, a target true positive rate TPR_T = 95 % was selected. All parameter combinations, which resulted in a TPR higher than TPR_T were considered for selection. The threshold, for which the lowest FPR was achieved, was set as the final threshold.

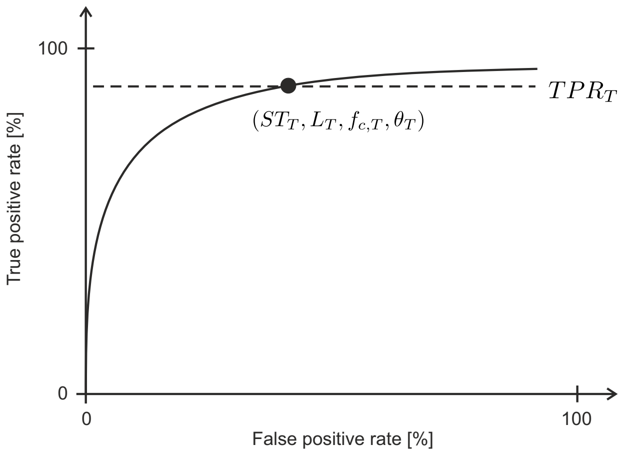
In the phase-based kick classification, ST, L, f_c, and θ had to be optimized. Compared to the previously described procedure, all four parameters were optimized by ROC analysis, which is explained in the following paragraphs. The ROC curve is illustrated in Figure 4.11b. A nested grid search was applied. The first three loops considered ST, L, and f_c. The values of the parameters of ST, L, and f_c were fixed and were determined before grid search execution:

$$\begin{aligned} ST &\in \{accelerometer, gyroscope\} \\ L &\in \{2, 3, 4, 5\} \\ f_c &\in \{50 \cdot k\}|_{k \in \{5,\ldots,10\}} \end{aligned}$$

In each grid search iteration, sensor type selection, high-pass filtering, and SMV computation were applied to event leg instances of true and false class for the current parameter setting of the first three loops. The maximum amplitudes of the SMV of all instances were determined. The range of the fourth loop over θ was set from the minimum amplitude to the maximum amplitude of the previously computed maximum signal magnitude values. In every grid search iteration, TPR and the FPR were computed and added to a ROC curve as illustrated in Figure 4.11. One point on the ROC curve represents one selected parameter configuration and was determined by LOSO-CV. In order to determine the final parameter setting, a target TPR $TPR_T = 95$ % was selected. All parameter combinations, which resulted in a TPR higher than TPR_T, were considered for selection. The parameter configuration with the lowest FPR was defined as best parameter setting.



(a) Hierarchical system



(b) Phase-based system

Figure 4.11.: ROC analysis of hierarchical (a) and phase-based (b) system. True positive rate is plotted over false positive rate. TPR_T denotes the target true positive rate. The parameters sensor type selection, order as well as cut-off frequency of Butterworth filter, and threshold are denoted as ST_T, L_T, $f_{c,T}$, and θ_T, respectively.

Classifier Parameter Optimization

The classifier parameter optimization in the event leg and kick type classification routines of the hierarchical system as well as in the kick type classification routine of the phase-based approach was performed in a similar manner. The five classifiers NB, SVM, k-NN, CART, and RF were applied to the corresponding algorithm-specific features. The class imbalance was tackled by SMOTE filtering [Chaw 02]. The linear kernel was used in the SVM classifier. The cost parameter C in the SVM, the number of neighbors k in k-NN, and the number of trees n_{tree} in the RF classifier were optimized by grid search ($C \in \{2^N\}, N \in \{-10, ..., 10\}$, $k \in \{3, 5, 7\}$, $n_{tree} \in \{5, 10, 15, 20\}$). The parameter optimization for all classifiers was performed in a LOSO-CV.

Experiments

Four experiments were conducted. The first three experiments were based on pre-segmented exercise instances (section 4.2, study A). The fourth experiment was based on data from the 11-a-side game (section 4.2, study B). The experiments are described below.

- **Experiment 1: Comparison of peak detection**
  In order to evaluate the two peak detection routines, the pre-segmented instances were divided into instances of a true class and instances of a false class. The true class included instances of full-instep and side-foot kicks. The false class included instances of additional activities such as dribbling or moving forward without the ball. In total, 986 instances of the true class and 989 instances of the false class were available. The evaluation was based on LOSO-CV. In each LOSO-CV loop, the corresponding pre-described optimization schemes were applied. The corresponding best parameter setting was used for left-out subjects. The three performance measures TPR, FPR, and precision were determined for each LOSO-CV trial (section 2.4.8). The mean and standard deviation of the three performance measures were computed.

- **Experiment 2: Comparison of event leg classification**
  In order to evaluate the two event leg classification routines, pre-segmented sensor data were used based on the two exercises 'eight subsequent side-foot kicks' and 'full-instep kicks' (Table 4.1). In

total, 593 instances were available. The mean and standard deviation of the sensitivities for left and right leg as well as the balanced accuracy were defined as performance measures determined by LOSO-CV (section 2.4.8). For the hierarchical system, the pre-described classifier parameter optimization scheme was applied in each LOSO-CV trial. The number of selected features across the LOSO-CV trials was further determined for the hierarchical system (mean $\pm$ standard deviation).

- **Experiment 3: Comparison of kick type classification**
  In order to evaluate the two kick type classification routines (side-foot vs. full-instep kick), all pre-segmented instances of side-foot and full-instep kicks were considered. In total, 870 instances of side-foot kicks and 116 instances of full-instep kicks were available. The mean and standard deviation of the sensitivities of side-foot and full-instep kicks as well as the balanced accuracy were computed in a LOSO-CV. In each LOSO-CV trial, the pre-described classifier parameter optimization scheme was applied. For the hierarchical approach, the BASE subsystem was evaluated separately. In total, 986 and 989 instances of kick and NULL were available, respectively. The number of selected features across the LOSO-CV trials was further determined for all classification systems of the hierarchical approach (mean $\pm$ standard deviation).

- **Experiment 4: Evaluation on game data**
  In order to evaluate a complete kick classification system on game data, the following procedure was applied to the system, which achieved the best balanced accuracy in experiment 3. The single systems for peak detection, event leg classification, and kick type classification were trained based on all the pre-segmented instances acquired in study A (training exercises). The trained systems were tested on the non-segmented sensor data of the 11-a-side game. In total, the test set included 279 instances of side-foot kicks and 18 instances of full-instep kicks from 12 players. The evaluation of the complete kick classification system included the performance assessment of peak detection and kick type classification routine. The computed performance measures were based on the confusion matrix, shown in Table 4.2. The confusion matrix considered the three classes 'side-foot kick', 'full-instep kick', and 'NULL'. Each of the nine entries is described below. Since the hierarchical approach

achieved the highest balanced accuracy in experiment 3, components of the proposed algorithm are mentioned.

- $C_{1,1}$: time between predicted side-foot kick and ground truth side-foot kick is less than δ s.
- $C_{1,2}$: time between predicted side-foot kick and ground truth full-instep kick is less than δ s.
- $C_{1,3}$: time between predicted side-foot kick and any ground truth kick type (side-foot kick or full-instep kick) is larger than δ s.
- $C_{2,1}$: time between predicted full-instep kick and ground truth side-foot kick is less than δ s.
- $C_{2,2}$: time between predicted full-instep kick and ground truth full-instep kick is less than δ s.
- $C_{2,3}$: time between predicted full-instep kick and any ground truth kick type (side-foot kick or full-instep kick) is larger than δ s.
- $C_{3,1,1}$: time between one detected peak position, which was classified as NULL by BASE subsystem, and one ground truth side-foot kick was less than δ s.
- $C_{3,1,2}$: ground truth side-foot kick did not pass peak detection routine.
- $C_{3,2,1}$: time between one peak position, which was classified as NULL by BASE subsystem, and one ground truth full-instep kick was less than δ s.
- $C_{3,2,2}$: ground truth full-instep kick did not pass peak detection routine.

Table 4.2.: Confusion matrix for three-class problem (side-foot kick vs. full-instep kick vs. NULL class). Each entry represents the number of classified instances. Columns indicate actual class. Rows indicate predicted class.

	Side-foot kick	**Full-instep kick**	**NULL**
Side-foot kick	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$
Full-instep kick	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$
NULL	$C_{3,1,1} + C_{3,1,2}$	$C_{3,2,1} + C_{3,2,2}$	$C_{3,3}$

- $C_{3,3}$: instances which passed peak detection routine and were correctly classified as NULL by BASE subsystem.

In this experiment, δ was set to 0.5 s. For simplicity, the following sums of matrix entries were defined:

$$\begin{aligned}
C_{1,*} &= C_{1,1} + C_{1,2} + C_{1,3} \\
C_{2,*} &= C_{2,1} + C_{2,2} + C_{2,3} \\
C_{3,*} &= C_{3,1,1} + C_{3,1,2} + C_{3,2,1} + C_{3,2,2} + C_{3,3} \\
C_{*,1} &= C_{1,1} + C_{2,1} + C_{3,1,1} + C_{3,1,2} \\
C_{*,2} &= C_{1,2} + C_{2,2} + C_{3,2,1} + C_{3,2,2} \\
C_{*,3} &= C_{1,3} + C_{2,3} + C_{3,3} \\
C_k &= C_{1,1} + C_{1,2} + C_{2,1} + C_{2,2}
\end{aligned}$$

In order to evaluate the peak detection routine, the total number of peaks is given, which coincides with the number of 1 s windows considered for further processing. Based on the confusion matrix shown in Table 4.2, the TPR and FPR were computed for the peak detection routine as defined below:

$$\begin{aligned}
TPR &= \frac{C_k + C_{3,1,1} + C_{3,2,1}}{C_{*,1} + C_{*,2}} \\
FPR &= \frac{C_{1,3} + C_{2,3}}{C_{*,3}}
\end{aligned}$$

The evaluation of the kick type classification routine was twofold. First, sensitivity and precision of the BASE subsystem were computed:

$$\begin{aligned}
Sensitivity &= \frac{C_k}{C_k + C_{3,1,1} + C_{3,2,1}} \\
Precision &= \frac{C_k}{C_{1,*} + C_{2,*}}
\end{aligned}$$

Second, sensitivity regarding side-foot and full-instep kick as well as balanced accuracy of the KICK subsystem were computed:

$$Sensitivity(side-foot) = \frac{C_{1,1}}{C_{1,1} + C_{2,1}}$$
$$Sensitivity(full-instep) = \frac{C_{2,2}}{C_{1,2} + C_{2,2}}$$

Balanced accuracy was defined as the mean of the sensitivities of both kick types.

4.3.5 Results

Experiments 1 - 3

The achieved performance measures for peak detection, event leg classification, and kick type classification (side-foot vs. full-instep kick) are comprised in Table 4.4a, 4.4b, and 4.4c, respectively. Table 4.5a summarizes the performance measures for the kick type classification in the hierarchical system (BASE subsystem). The number of selected features in the single steps event leg classification and kick type classification (hierarchical system) are given in Table 4.5b.

Experiment 4

The confusion matrix after applying the proposed hierarchical kick classification system on the 11-a-side game data is depicted in Table 4.5. In total, 1355 peaks were detected by the proposed hierarchical kick classification system across all players. The peak detection routine achieved a TPR of 90.2 % and a FPR of 24.3 %. The BASE subsystem of the kick type classification routine achieved a sensitivity of 88.4 % and a precision of 47.3 %. The KICK subsystem of the kick type classification routine achieved a sensitivity of 87.7 % and 82.4 % for side-foot and full-instep kick, respectively. A balanced accuracy of 85.0 % was reached.

4.3.6 Discussion

In this section, two kick classification systems were developed, implemented, and evaluated (Figure 4.7). The achieved results are discussed in the following paragraphs. The discussion is structured as follows.

Table 4.3.: Comparison of two kick classification methods (hierarchical vs. phase-based approach) regarding peak detection (a), event leg classification (b), and kick type classification (c).

Performance measure	Hierarchical	Phase-based
True positive rate	95.0 ± 7.2	94.8 ± 2.6
False positive rate	19.2 ± 9.3	11.8 ± 7.2
Precision	83.9 ± 6.2	89.1 ± 6.1

(a) Peak detection. True positive rate, false positive rate, and precision are given in percent.

Class	Hierarchical	Phase-based
Left leg	99.7 ± 1.1	99.0 ± 1.7
Right leg	99.7 ± 1.1	99.7 ± 1.1
Mean	99.7 ± 0.0	99.3 ± 0.5

(b) Event leg classification. Sensitivities and balanced accuracies are given in percent. The highest balanced accuracy was achieved by k-NN classifier for the hierarchical algorithm.

Class	Hierarchical	Phase-based
Side-foot kick	98.4 ± 2.0	96.1 ± 4.3
Full-instep kick	94.7 ± 8.2	92.1 ± 11.6
Mean	96.5 ± 2.6	94.1 ± 2.8

(c) Kick type classification. Sensitivities and balanced accuracies are given in percent. The highest balanced accuracy was achieved by NB and SVM classifier for hierarchical and phase-based algorithm, respectively.

First, the modifications compared to the original publications [Schu 15, Schu 16] are comprised. Second, the peak detection, event leg classification, and kick type classification routines are discussed (experiment 1 - 3). Third, evaluation of the game data is investigated (experiment 4). Fourth, the proposed approaches are compared to currently available literature. The section concludes with future work.

Modifications Compared to Original Publications

Compared to the original publications [Schu 15, Schu 16], modifications were made. In [Schu 15], 48 statistical features were extracted in the

Table 4.4.: Evaluation of hierarchical system. Performance of BASE classification system (a) and correlation-based feature selection of BASE and KICK subsystem (b) are given.

Class	Sensitivity
Kick	97.0 ± 1.9
NULL	98.0 ± 2.0
Mean	97.5 ± 0.7

(a) BASE subsystem. Sensitivities of Kick and NULL class as well as balanced accuracy are given in percent. k-NN achieved highest balanced accuracy.

Classification algorithm	Number of selected features
Event leg	12.5 ± 2.7
Kick type (BASE subsystem)	16.6 ± 1.9
Kick type (KICK subsystem)	7.6 ± 1.0

(b) Correlation-based feature selection results for event leg and kick type classification (BASE and KICK subsystems). Number of selected features (mean ± standard deviation).

Table 4.5.: Confusion matrix for three-class problem (side-foot kick vs. full-instep kick vs. NULL class). Each entry represents the number of classified instances. Columns indicate actual class. Rows indicate predicted class. Definition of matrix entries is given in Table 4.2.

	Side-foot kick	Full-instep kick	NULL
Side-foot kick	193	3	211
Full-instep kick	27	14	53
NULL	31 + 28	0 + 1	823

event leg and kick type classification steps. The feature set in this thesis consisted of 112 feature followed by a feature selection routine, in order to provide an optimized feature set for each classification problem. The set of tested classifiers were further extended in this thesis compared to [Schu 15, Schu 16]. Compared to [Schu 16], the set of instances used in the evaluation was extended by including more exercises. The conducted experiments described in [Schu 15, Schu 16] were also extended.

The accuracy of BASE and KICK subsystem (experiment 4) were comparable to the performance achieved in [Schu 15]. The performance of the BASE and KICK subsystem slightly decreased by 1.1 % and increased by 0.8 %, respectively.

The TPR of the peak detection and the sensitivity for side-foot kicks of the kick type classification (Table 4.3) were further comparable to the performance achieved in [Schu 16]. The performance of the kick type classification routine for full-instep kicks decreased by 3.5 %. The reason might be that more instances were used in this thesis. Some instances might be misclassified as side-foot kicks.

Comparison of Kick Classification Components

Both **peak detection routines** included similar signal processing components such as high-pass filtering, SMV computation, and thresholding. TPRs of above 94 %, FPRs below 20 %, and precision values above 83 % showed the general applicability of both methods to detect kick candidates (Table 4.4a). One advantage of both methods is the generic architecture due to the applied optimization scheme. The algorithms can e.g. also be applied to sensor data acquired on a different sensor position such as the ankle position shown in Figure 4.3. Both methods exploited the fact that biomechanical data of kicks are characterized by a sudden transition from low-frequency (before and after impact) to high-frequency (during impact). Nevertheless, the performance of the peak detection routine in the phase-based approach seemed to be slightly better indicated by a lower FPR and a higher precision . The reason might be that the ROC analysis of the peak detection in the phase-based approach included more parameters in the optimization, e.g. the signal type and the order of the Butterworth filter (Figure 4.11b). In order to further reduce the number of false positives, the corresponding soccer-specific activities which fall in the false positive group should be investigated. An automatic cluster analysis might identify a group of activities, e.g. ball contacts during dribbling, which might be eliminated by an additional processing step.

Both **event leg classification routines** were able to recognize the correct leg, which was used to perform either a side-foot or a full-instep kick (Table 4.4b). One main difference is the number of extracted features. The correlation-based feature selection in the hierarchical approach determined 12.5 features on average (Table 4.5b) compared to only one computed features in the phase-based approach. Thus, computing the

SMA seemed to be sufficient for event leg classification. An additional advantage is that the classifier is only based on one rule without any parameters which have to be optimized in a training phase. Nevertheless, both approaches were only evaluated on data of kicks. In future, the event leg classification could be extended to more complex activities such as dribbling. Statistics about the leg usage in dribbling scenes might give coaches further insights in the performance of individual players. In the more complex scenarios, the event leg classification routine in the hierarchical approach might outperform the corresponding expert-driven one due to the more generic architecture.

Both **kick type classification routines** achieved balanced accuracy values of above 94 % (Table 4.4c). Thus, side-foot and full-instep kicks seemed to be distinguishable from each other, at least in a training session with pre-defined exercises. The sensitivity of side-foot kicks were higher in both routines. The reason might be that more instances were available and the SMOTE filtering was not able to generate adequate synthetic instances of full-instep kicks. The correlation-based feature selection in the hierarchical approach determined almost the same number of features than in the phase-based approach, which used a fixed feature set (Table 4.5b). Both feature sets might be interchangeable. The single mean sensitivities were further slightly higher in the hierarchical approach and the corresponding standard deviations were lower (Table 4.4c). This indicates that the hierarchical approach achieved a better performance regarding unseen players and seemed to be more robust regarding player variability. The BASE subsystem in the hierarchical approach further showed an overall good performance indicated by a balanced accuracy of 97.5 % (Table 4.5a). The BASE system can be used to remove non-kick activities which passed the peak detection routine.

Both systems showed advantages in different routines. The precision of the peak detection routine was e.g. better in the phase-based approach, whereas the kick type classification in the hierarchical approach achieved a better performance. In the future, a fusion of both algorithms should be investigated. The module-like architecture of both methods might benefit the fusion process, since the single routines are interchangeable (Figure 4.7).

Evaluation on Game Data

The hierarchical approach was further applied on game data. The performance of the peak detection dropped by about 5 % regarding both TPR and FPR. The lower TPR might be explained by the different executions of side-foot kicks and full-instep kicks in the game. Since the hierarchical system was trained on exercise data, the mentioned variability was not modeled. The higher FPR indicates that additional activities occur in a game scenario which were also not considered in the training phase. Future systems should include game data in the training process of HAR systems in order to improve the performance. The proposed algorithm might be used as first version of a kick detection system, which further implements online learning techniques to gather game scenarios. Game data can be used to re-train the system and improve the accuracy. Nevertheless, detecting 1355 peaks which included 90 % of the desired events showed the general applicability of the peak detection routine to find kick candidates. One major advantage of using peak detection is the reduction of windows that are considered for further processing. Assuming a game duration of 45 minutes and a sliding window approach (1 s width) for sequence data processing, 2700 windows have to be processed. With the proposed peak detection, the number of processed windows could be reduced by about 50 %. The lower number of processed windows might enable an implementation on embedded devices with a reduced computational complexity. Future research should investigate the performance of the peak detection routine with a lower sampling rate. The sampling rate could be one additional component in the ROC analysis (Figure 4.11a).

The BASE subsystem removed 31 instances of side-foot kicks for further processing (Table 4.5). The reason might be that the side-foot kick execution differed from the execution during exercises. Side-foot kicks occurring in a game might be part of a high dynamic sequence of movements, which could not be captured in a training session. Since the instances passed the peak detection, the selected parameters for the Butterworth filter and the threshold seemed to be appropriate. Nevertheless, the selected features in the BASE subsystem did not capture the variability of side-foot kick executions. Future systems should consider more side-foot kicks, which are performed e.g. with the lateral side or other areas of the foot. Nevertheless, the BASE subsystem correctly removed 823 instances for further processing (Table 4.5). The BASE subsystem can be seen as a second outlier detection apart from the peak detection routine. Although a high number of outlier instances could be removed, 264 instances were wrongly predicted as kicks resulting in a low precision of 47.3 %. The false positives were further investigated. Out of the

264 false positives, 90 and 52 peak-aligned segmentation windows were close to a control and a long pass, respectively. These activities were not considered in the training phase of the system. Thus, the BASE subsystem further recognized additional low-intensity and high-intensity ball contacts. In the future, the hierarchy of the proposed algorithm could be extended by additional ball contact types. The first level might distinguish between instances with and without the involvement of the ball. The second level might further divide the ball contact instances into low-intensity and high-intensity ball contacts. Examples of low-intensity ball contacts might be controls and side-foot kicks. Examples of high-intensity ball contacts might be long passes and full-instep kicks. The third level might further distinguish the instances among the different intensity ball contacts. Neglecting the controls and long passes results in 122 remaining false positives based on 12 players. This means that in 45 minutes about 10 instances were wrongly predicted as kicks by the proposed BASE subsystem. The instances might include peak-rich activities such a dribbling and running. The low number of player-specific false positives in reference to the low achieved precision indicates one major challenge of IMU-based HAR in soccer. The target activities, in this case kicks, rarely appear in game scenarios and only a few misclassifications result in a low performance of the complete system. Thus, applications in professional team sports, which require a high accuracy of the detected events, might not use IMU-based systems as alternative to expensive camera-based systems to compute game statistics. Nevertheless, amateur teams, which would like to assess the performance of players in training sessions might benefit from such a low-cost IMU-based HAR system. Although it is challenging to apply the proposed approach for player performance assessment, various other application domains might be possible. One example is shown later in this chapter.

The high balanced accuracy of 85.0 % shows the general applicability of the KICK subsystem and the subsequent post-processing steps to distinguish between side-foot kick and full-instep kick. The combination of the two step classification scheme seems to be beneficial. First, the rather generic KICK subsystem provides a first distinction between the two kick types based on an automatic, data-driven approach. Second, certain rules and expert knowledge were included in the system for a decision refinement. The thresholds, which were currently used in the post-processing step, were manually set. Future research should determine the thresholds by an automatic optimization scheme based on a large amount of game data.

Comparison to Literature

The best single kick classification technique, the hierarchical approach, is further compared to state-of-the-art algorithms. The comparison to literature is divided into two parts. The first part deals with the different sensor configurations. The second part considers the applied algorithmic concepts and the corresponding evaluation.

Compared to the current literature, a higher sampling rate was used in this thesis which might be beneficial to detect high frequency signal parts. It was assumed that high frequency signal parts might indicate ball contacts. Using a much lower sampling rate, e.g. 16 Hz as in [Mitc 13] might remove important signal characteristics. The number of attached sensors and considered sensor types were also higher in this thesis compared to commonly available state-of-the-art approaches. In this work, two sensors were used placed inside a soccer shoe cavity (Figure 4.1). Each sensor comprised both an accelerometer and a gyroscope. In [Mitc 13, Ahma 15, Hoss 17], only one accelerometer was used attached to the upper back, to the shank, and to the wrist, respectively. Advantages of the proposed sensor setup are that (a) information of both event and supporting leg can be considered in the algorithms and (b) the players are often not aware of wearing a sensor unit, if it is inside the sports equipment. Attaching a sensor on the wrist or at the upper back might influence the player during the execution of soccer-specific activities.

The authors in [Mitc 13] implemented an ARC for a multi-class classification problem (stationary, walking, jogging, sprinting, hitting the ball, attempting a standing tackle, and dribbling the ball). The windows were not centered around peaks. The algorithm was evaluated on five five-a-side soccer games each lasting one hour. In total, data from 15 players were recorded and a F1-score of about 74 % was achieved for hitting the ball. Compared to the 7-class problem in [Mitc 13], the BASE subsystem solved a 2-class problem and achieved a F1-score of 61.6 % considering game data. Although the proposed approach achieved a lower performance, the defined experiment might be more realistic. In [Mitc 13], data were manually segmented not considering the huge amount of additional soccer-specific activities. Furthermore, instances of the same player were used in the training and test set. Compared to [Mitc 13], the complete game data was considered for evaluation and the players in training and set mainly differed. Only one player took part in both data collection sessions. Future research should compare both methods using the same dataset.

In [Ahma 15], an ARC was implemented solving a 6-class problem. The set of target activities included football free kicks. The algorithm was evaluated on pre-defined exercises. The proposed system achieved a sensitivity of 83.9 % regarding kicking. In this thesis, the KICK subsystem achieved a sensitivity of 94.7 % and 82.4 % considering exercise and game data, respectively. Thus, the proposed approach outperformed the system introduced in [Ahma 15] using a comparable study setup and achieved a similar performance using a more complex game scenario. This means that the proposed system might be more applicable in realistic settings than the method described in [Ahma 15]. Nevertheless, two sensors with two sensor types were needed compared to one sensor and one sensor type. In future, a trade-off between number of sensors and accuracy has to be made.

In [Hoss 17], a deep learning approach was applied compared to the shallow architectures proposed in [Mitc 13, Ahma 15] and in this thesis. The developed method was evaluated on data from 5-a-side games and sensitivity values of about 75 % and 82 % were achieved for passing and kicking, respectively. In this thesis, the KICK subsystem achieved sensitivities of 87.7 % and 82.4 % for side-foot and full-instep kick, respectively. The reason for the lower performance of the method introduced in [Hoss 17] might be the low amount of training data. Deep learning architecture usually requires a huge amount of data. Nevertheless, data of only six players were available.

Future Work: From Shallow to Deep Architectures

The proposed kick classification methods in this thesis as well as the algorithms described in [Mitc 13, Ahma 15] provided a shallow architecture [Beng 09]. Hand-crafted feature engineering was performed using features which were typically applied in the field of activity recognition [Bull 14]. They achieved good results under rather controlled conditions. Nevertheless, the performance of the shallow machine learning techniques dropped under more realistic conditions, e.g. during a game (experiment 4). The high number of False Negatives (FN) and False Positives (FP) resulted in lower sensitivity and precision values. The problem of FN might come from a different execution of activities, e.g. a side-foot kick, which is also subject-dependent. The problem of FP might result from a lot of 'unseen' activities which were implicitly labeled as NULL class. To better model different executions of the same activity and NULL class instances, deep architectures should be used in the future. The applicability of deep architectures to detect soccer-specific activities was proven by [Hoss 17]. Nevertheless, the algorithm was only tested with pre-segmented game data based on a low number of players. Future research work should include a comparison of different deep learning approaches evaluated on complete game data. Techniques which use Restricted Boltzman Machines should be compared e.g. to Convolutional Neural Networks. A further performance improvement might be achieved by using Recurrent Neural Networks, which are very powerful for modeling sequence data [Deng 14], e.g. text or speech. Recurrent Neural Networks should include the Long-Short-Term Memory architecture in order to deal with the problem of gradient vanishing and gradient explosion. The comparison might be amended by Generative Adversarial Nets which were introduced by [Good 14]. The proposed framework simultaneously trained both a generative and a discriminative model. The generative model captured the distribution of the sensor data. The discriminative model estimated the probability that an instance came from the training data rather rather than from the generative model.

One reason, why deep learning techniques were not investigated in this thesis, was the low amount of available training data. One major challenge in soccer is that the target activities such as kicks appear rather rarely in relation to the game duration (experiment 4). Thus, a high number of games have to be acquired, in order to provide a suitable amount of labeled instances. Nevertheless, labeling of such games is usually very time-consuming. In order to speed up the labeling of games, parts of the proposed approach, e.g. the peak detection could be used to provide the labeling person with a set of candidates for the target activities. The labeling person could then set the class labels to the candidates. Shallow architectures like the proposed algorithm developed in this thesis should not be replaced by deep architectures but maybe used as a supporting tool in order provide e.g. training data in a shorter amount of time.

Future Work: From Offline to Online Analysis

The proposed algorithms in this thesis and the state-of-the-art approaches [Mitc 13, Ahma 15, Hoss 17] applied offline analysis. In the future, algorithms might be implemented directly on the sensor node for online analysis. One possible application of such an online analysis was introduced in a patent application [Zwic 16]. The invention included an instrumented shoe for ball sports like soccer, which can change its surface properties dependent on the match situation. IMU-based machine learning methods were required in order to detect e.g. soccer-specific activities as early as possible. One example algorithm was developed and implemented in the Master Thesis of Eva Dorschky [Dors 14] (supervised by the author of this dissertation). The algorithm was designed to detect kicks as early as possible using phase segmentation, feature extraction, and a hybrid HMM detector. The approach achieved a sensitivity and precision of 76 % and 96 %, respectively. The algorithm was further published at a conference [Dors 15]. The reason for the low sensitivity might be the shallow architecture, which was e.g. based on hand-crafted feature engineering. Additional algorithmic approaches have to be investigated in order to improve the performance of the system. An application of deep architectures as described in the previous section should be investigated. Deep architectures could e.g. be used to better model phases before the kick.

4.4 Ball Speed Estimation

4.4.1 Overview

The estimation of the ball speed after a full-instep kick enables the categorization of the kick according to the intensity. The ball speed in soccer is usually investigated by vision-based systems. Research work regarding low-cost, IMU-based solutions for ball speed estimation was limited and focused on one-axial accelerometers evaluated on stationary full-instep kicks [Zwic 12]. In this section, a ball speed estimation approach is introduced, which used accelerometer and gyroscope data to determine the ball speed. The technique was evaluated on ten subjects who performed dynamic exercises. Parts of the algorithm were originally filed in a patent application [Kirk 16]. The proposed approach was further published in [Schu 16] and was modified in this thesis. The modified parts are highlighted in the discussion.
This section is structured as follows. First, the proposed ball speed estimation algorithm is introduced. Second, the conducted experiments and evaluation are explained. Third, the results are comprised. Fourth, the results are finally discussed.

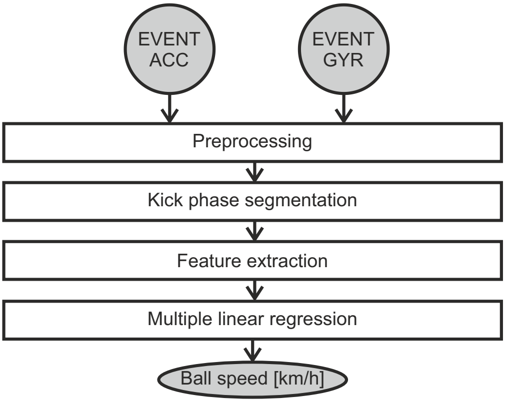


Figure 4.12.: Pipeline of ball speed estimation. ACC and GYR denote accelerometer and gyroscope, respectively. EVENT denotes event leg.

4.4.2 Methods

The proposed ball speed estimation pipeline is shown in Figure 4.12. The approach used sensor data of the event leg as input (see section 4.3). The system consisted of four steps:

1. **Preprocessing:** calibration of the IMU sensor based on [Ferr 95].
2. **Kick phase segmentation:** already introduced in section 4.3.3.
3. **Feature extraction**: computation of three features during leg acceleration phase exploiting the fact that the final ball speed is dependent on the foot velocity before impact [Nuno 06]:
   - SMV signal was computed (equation 2.6) based on accelerometer data followed by integration motivated by [Zwic 12].
   - Maximum angular velocity in sagittal plane motivated by [Reil 03, Bull 99].
   - Aggregated number of samples in the saturation phase of all three accelerometer and gyroscope axes. It was assumed that the length of the saturation phase correlated with the foot velocity.
4. **Multiple linear regression** based on extracted features (section 2.4.7).

4.4.3 Experiments and Evaluation

The proposed ball speed estimation algorithm was evaluated on the datasets acquired in study C (section 4.2.3). The kick instances were manually segmented based on the camera data. In total 360 instances were available. The range of the ground truth ball speed was [55.3 km/h, 119.7 km/h]. Three experiments were conducted.

1. **Experiment 1: Comparison of term configurations**
   The performance of the MLR was compared regarding three different term configurations (section 2.4.7):

   (a) Constant and linear.

   (b) Constant, linear, and interaction.

   (c) Constant, linear, interaction, and quadratic.

   The three configurations were compared by boxplots of the MAPE. The MAPE was determined in a LOSO-CV. For the configuration with the lowest MAPE, the MAE was further provided.

2. **Experiment 2: Ball speed-dependent performance**
   The Bland-Altman plot was determined for the configuration with the lowest MAPE (section 2.4.8). The aim of the second experiment was to investigate the systematic error of the ball speed estimation algorithm and the ball speed-dependent performance.

3. **Experiment 3: Subject-dependent performance**
   Mean and standard deviation of the subject-dependent MAPE were computed in order to investigate the performance of the ball speed estimation algorithm regarding individual subject characteristics, e.g. the BMI.

4.4.4 Results

Figure 4.13 shows the MAPE regarding the three considered model term configurations. The lowest mean MAPE of 6.7 ± 1.4 % was achieved by the model using constant, linear, and interaction terms (configuration b). The MAE was 5.5 ± 0.9 km/h. The corresponding Bland-Altman plot is further depicted in Figure 4.14. Table 4.6 comprises mean and standard deviation of the subject-dependent MAPE. The BMI of each subject is further provided.

4.4.5 Discussion

In this section, an IMU-based ball speed estimation algorithm was introduced and compared to a Hawk Eye ground truth system. The results are discussed in the following paragraphs. The discussion is structured as follows. First, the modifications compared to the original publication are described. Second, the applicability of the Hawk Eye system is discussed. Third, the findings of the three experiments are comprised. Fourth, the proposed approach is compared to the literature. The section concludes with future work.

Modifications Compared to Original Publication

Compared to the original publication [Schu 16], modifications were made. The feature set was extended. In [Schu 16], only the number of samples in saturation phase was computed for the accelerometer data. In this thesis, the gyroscope signal was integrated in the sum. Two additional features were extracted in this work: maximum angular velocity in sagittal plane and sum of accelerometer-based SMV. A different dataset for evaluation was further chosen, since only 45 instances were available

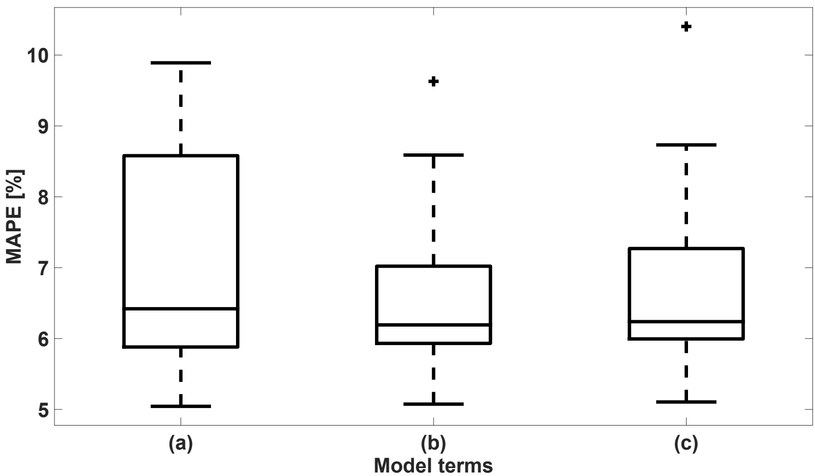


Figure 4.13.: Boxplot of MAPE regarding three considered model term configurations. (a) constant and linear, (b) constant, linear, and interaction, (c) constant, linear, interaction, and quadratic.

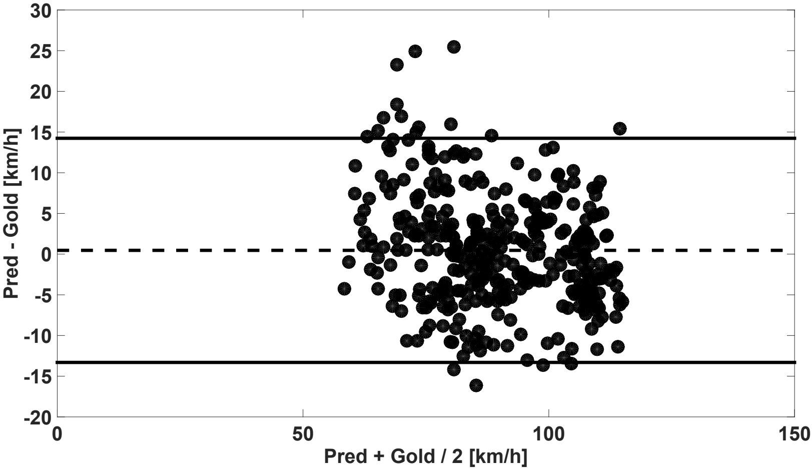


Figure 4.14.: Bland Altman plot using constant, linear, and interaction terms (configuration b). 'Pred' and 'Gold' denote the predicted and ground truth ball speed values, respectively.

compared to 360 instances of study C (section 4.2.3). Study C also included dynamic exercises compared to the stationary full-instep kicks considered in [Schu 16].
All in all, the performance of the proposed ball speed estimation algorithm could be improved by 2.2 % compared to [Schu 16].

Hawk Eye System

High-speed cameras are usually used as ground truth [Zwic 12]. Nevertheless, using high-speed cameras limited the considered scenarios which could be analyzed. In this thesis, high-speed cameras could not be used as ground truth due to the selected dynamic exercises (study C, section 4.2.3). Thus, the Hawk Eye system had to be used for gold standard measurements. The applicability of the Hawk Eye system was proven in a Master's thesis [Jako 16]. Reproducible full-instep kicks were performed by a robotic leg (Roboleg) [Sche 95]. The Roboleg enabled to transfer momentum to a soccer ball to achieve launch speeds up to 144 m/s. The launch speed and launch angle could be configured by changing the pendulum speed and length. A MAE of 2.1 km/h was achieved between

Table 4.6.: Subject-dependent MAPE in descending order. Subject ID and BMI are further given for each subject. Multiple linear regression was performed with constant, linear, and interaction terms.

Subject ID	MAPE [%]	BMI [kg/m^2]
8	9.6	26.1
6	8.6	26.0
3	7.0	26.3
10	6.9	26.1
5	6.4	24.0
4	6.0	22.9
7	6.0	23.9
9	5.9	23.8
1	5.4	23.7
2	5.1	24.8

Hawk Eye and Roboleg. Thus, the Hawk Eye system was assumed to be an adequate ground truth system.

Comparison of Term Configurations

The three term configurations used in the MLR achieved similar mean MAPE ranging from 6.7 % (configuration b) to 7.0 % (configuration a). Nevertheless, the interquatile range differed. The highest interquartile range of 2.7 % was achieved by configuration a (Figure 4.13). This indicates that the subject-dependent results showed a higher variability using only constant and linear terms. A regression model with interaction and quadratic terms seemed to be more robust regarding the application to different subjects. The proposed IMU-based ball speed estimation system achieved a MAE of 5.5 km/h (configuration b), which is 2-3 times higher than the measured Hawk Eye accuracy. Thus, the proposed IMU-based technique cannot be used as ground truth system e.g. for biomechanical measurements in the field. Nevertheless, the system might be used to perform a rough categorization of the achieved ball speed in low, medium, and high level.

Ball Speed-Dependent Performance

The Bland-Altman plot (Figure 4.14) showed that no systematic error is included in the analysis. Figure 4.14 further indicates that the proposed ball speed estimation system performed better for higher ball speed values, since less instances are outside the 95 % limits of agreement borders than for lower ball speed values. The reason might be the selection of the features. For lower ball speeds, a saturation phase might not be present which would result in a feature value of 0. Thus, for lower ball speed values the resolution of the ball speed prediction might be limited. In order to further improve the proposed ball speed estimation system, especially for lower speed values, additional phases of the full-instep kick should be considered in the analysis, e.g. backswing, leg cocking, and follow through [Nuno 02, Brop 07]. Additional dependent factors for ball speed might also have a positive influence of the performance such as technique analysis [Lees 10], approach angle [Kell 04], support leg dynamics [Inou 14, Augu 17], and trunk kinematics [Nait 10, Full 15].

Subject-Dependent Performance

Table 4.6 indicates that subjects with a higher BMI achieved a lower performance compared to subjects with a lower BMI. Subjects 8, 6, 3, and 10 had BMI values above 25 kg/m^2 which is classified as overweight [Seid 97]. Thus, only considering the description of the full-instep kick by IMU-based measurements might not be enough for a robust ball speed estimation system. In future, additional information such as anthropometric data should be integrated in the analysis. The aim in future should also consider a balanced dataset regarding BMI to ensure a suitable training of the regression model.

Comparison to Literature

Research work about the estimation of the ball speed after a full-instep kick was usually evaluated on stationary ball scenarios [Zwic 12, Schu 16]. In this thesis, the aim was to evaluate an IMU-based ball speed estimation algorithm under more realistic scenarios. Thus, dynamic exercises were chosen in the data collection, which included a full-instep kick after a dribbling or passing sequence (study C, section 4.2.3).

In [Zwic 12], a one-axial $\pm$ 70 g accelerometer was used as input for a regression-based analysis. A biomechanical collision model was applied with the following components: moment of inertia of shank-foot segment, velocity of foot before ball impact, mass of ball, distance between knee and point of contact, and coefficient of restitution. The accelerometer was used to estimate the foot velocity. The evaluation was based on one subject and the proposed method achieved a MAPE of 2.81 %. Compared to [Zwic 12], a machine learning based approach was applied in this thesis using the information of low-range accelerometer and gyroscope axes. The proposed approach achieved a higher MAPE of 6.7 %. Although the performance of the proposed approach seemed to be lower, the system was tested on more subjects and a more complex environment. A biomechanically-driven model might better integrate expert knowledge, whereas a data-driven machine learning technique might be more robust under complex environments. Future research should fuse both strategies in order to provide a high accurate and robust system. Nevertheless, the advantage of the proposed machine learning approach was that the algorithm did not require information about segment length or mass of ball. Thus, no user input was required for prediction.

In [Schu 16], the sensor was positioned in a shoe cavity. The machine learning based approach used only information of the accelerometer and achieved a MAE of 7.7 $\pm$ 4.1 km/h. The method was evaluated on only 45 instances. The proposed method in this thesis extended the method described in [Schu 16] by including information of the gyroscope and achieved both a lower mean and standard deviation of the MAE: 5.5 $\pm$ 0.9 km/h. The reason for the lower standard deviation might be the higher amount of training data (360 instances). Although the position in the shoe might be more secure, it needs an additional effort in the shoe production process. The position of the sensor near the ankle attached by a strap (Figure 4.3) offers more flexibility regarding the shoe choice and no additional step has to be performed in the shoe production process.

Future Work

The proposed ball speed estimation algorithm was trained and tested on data of a low-range accelerometer. Figure 4.8a shows that the low-range accelerometer reached saturation during the full-instep kick. Future work should investigate the application of a high-range accelerometer for ball speed estimation. A first investigation was done in the Bachelor Thesis of Maike Stoeve [Stoe 14] (supervised by the author of this dissertation). The performance of a ball speed estimation algorithm was investigated using an uniaxial accelerometer with a high range of $\pm 250g$ and a triaxial accelerometer with a low range of $\pm 16g$. In order to provide a fair comparison, the corresponding axis of the low-range accelerometer was selected for the evaluation. The results showed that using the high-range accelerometer increased the MAE by about 27 %. Future research should investigate, if the performance of ball speed estimation algorithms could further be improved by using triaxial high-range accelerometers or a fusion of several accelerometers with different ranges.

4.5 Application: Automatic, Sensor-Driven Video Summary

4.5.1 Overview

The goal of video summaries e.g. in soccer is to substitute a long video stream with a few number of selected scenes. The scenes should contain interesting highlights such as goals, bookings, shots on goals, or penalties [Rehm 14]. A manual browsing through the whole video stream is cumbersome and time-consuming [DOra 10]. In contrast, automatic approaches often relied on visual, audio, as well as textual information and were therefore mostly applicable to TV broadcast material in professional sports [Rui 00, Ekin 03, Baba 04]. Systems and methods for recreational sports are missing. To close this gap, a novel, innovative, and low-cost solution for the generation of video summaries was developed. Each individual player was provided with a personal video summary containing highlight scenes. The proposed system automatically associated frames in a video sequence. The frame selection relied on information provided by low-cost inertial sensors attached to soccer players and corresponding HAR algorithms. The algorithms included e.g. the previously introduced kick classification routines (section 4.3) and the ball speed estimation approach (section 4.4). The kick classification algorithms can be used to identify interesting highlights, e.g. during a game. The ball speed estimation approach can be used to rank and order the kicking sequences according to the achieved ball speed. The system and a set of corresponding HAR algorithms were filed in a patent application [Kirk 16]. An extension of the video summary system was further published in [Schu 16]. The extension included sensor-driven video and audio effects, which were added to the final highlight video. The following sections briefly introduce the main system, which associates video frames with an event such as a soccer kick, and the previously mentioned extension of the system.

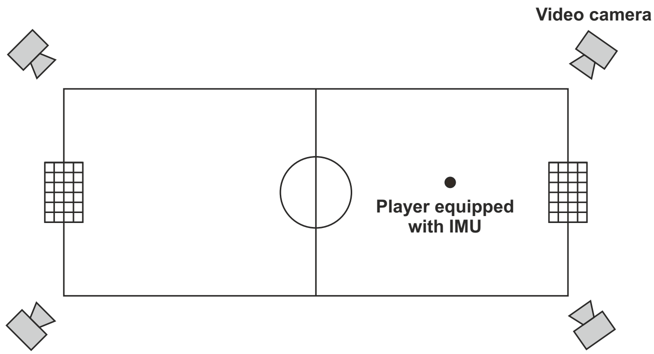


Figure 4.15.: Pitch setup of video summary system (adapted from [Kirk16]). Player on pitch equipped with IMU sensor and video cameras around pitch.

4.5.2 Video Frame Association with an Event

The system consisted of multiple video cameras, which were positioned around the pitch (Figure 4.15). The players on the pitch were further equipped with IMUs, which were synchronized with the video cameras. Inertial sensor data were used as input for HAR algorithms (Figure 4.16). The algorithms were used to identify soccer-specific activities which were defined as highlight examples. Soccer-specific activities might include full-instep and side-foot kicks, tricks, as well as fouls. The application of the kick classification routines (section 4.3) could be used to detect full-instep and side-foot kicks. The time stamp, at which a kick was predicted, and the corresponding kick label was stored in a list for further processing. If a full-instep kick was predicted, the corresponding ball speed was further determined, e.g. by the method described in section 4.4. Subsequent video frames of single highlight scenes were selected around the time stamp. In a post processing step, the order of highlight scenes were automatically set according to certain criteria. The full-instep kicks were ranked e.g. based on the achieved ball speed. Full-instep kicks with higher achieved ball speeds were shown earlier in the video summary.

4.5.3 Extension: Video and Audio Effects

The extension of the previously described sensor-driven video summary system added video and audio effects to highlight scenes containing full-instep kicks. An example is given in Figure 4.17. The phase-based kick classification algorithm described in section 4.3.3 was applied in order to identify a full-instep kick and different phases during the kick execution, e.g. heel strike, start of leg-acceleration phase, and ball contact. Each ball contact, which was identified as part of a full-instep kick, was considered for further processing. A video clip was generated starting 1 s before and 3 s after the ball contact. The video was further divided into three parts: (A) beginning of video clip to heel strike, (B) heel strike to ball contact, and (C) ball contact to end of video clip. In each part, video and audio effects were applied. During part (A), a fade-out video effect was performed. In part (B), difference images between two subsequent frames were computed and a slow motion effect were performed. In part (C), a sound was played starting at the ball contact. A text box was further displayed with the estimated ball speed.

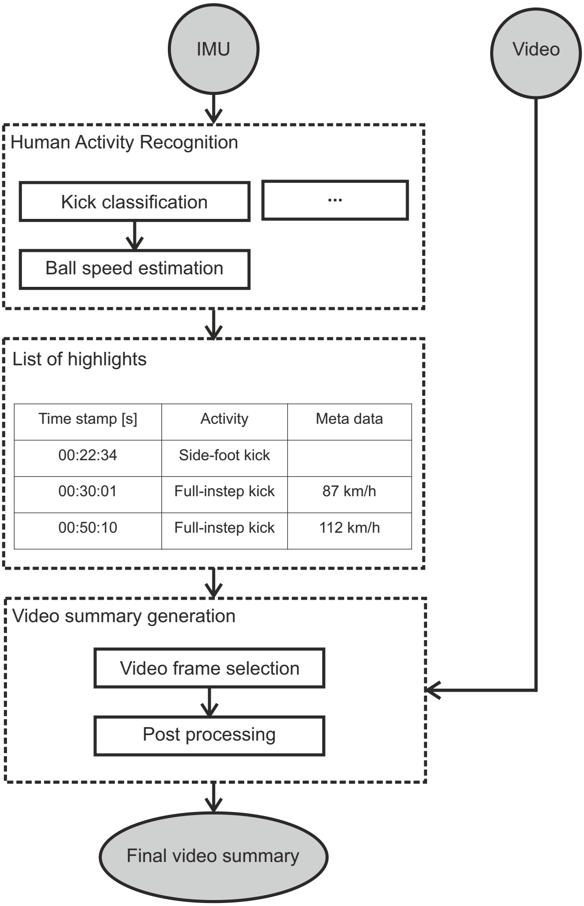
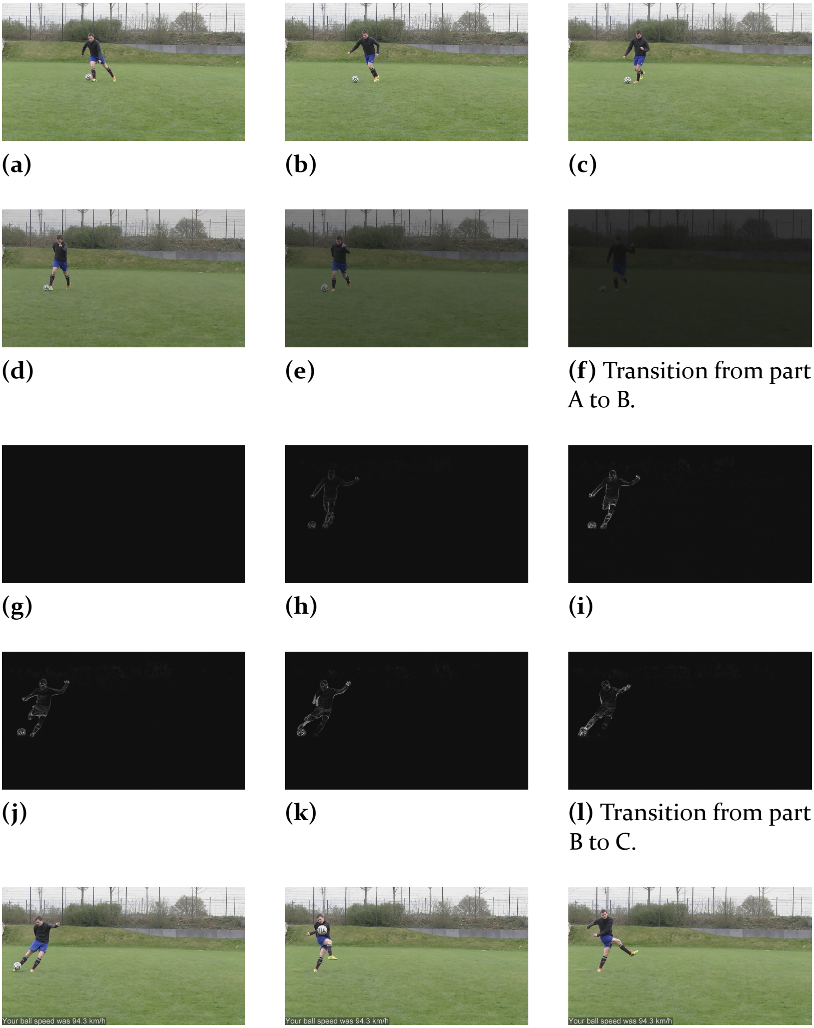


Figure 4.16.: Overview of pipeline for video summary application. IMU-based HAR algorithms detect desired soccer-specific activities such as kicks and store corresponding time stamps. If a full-instep kick is recognized, the ball speed is further determined. Based on the time stamps, the corresponding video frames around the activity are selected and post-processed.



(a) **(b)** **(c)**

(d) **(e)** **(f)** Transition from part A to B.

(g) **(h)** **(i)**

(j) **(k)** **(l)** Transition from part B to C.

(m) **(n)** **(o)**

Figure 4.17.: Extension of video summary system: integration of video and audio effects. Video was divided into three parts: (A) beginning of video clip to heel strike, (B) heel strike to ball contact, and (C) ball contact to end of video clip.

Chapter 5

Big Data Ready Human Activity Recognition

5.1 Introduction

In this chapter, one contribution regarding Big Data for HAR is presented (contribution 9). The whole contribution consisted of four parts. The first part was already presented in section 1.5.1 and included a literature review about the current state-of-the-art in wearable-based HAR. The remaining three parts were based on the findings of the literature review and provided an introduction of Big Data to the HAR research field. The three parts consisted of a HAR-related definition of Big Data, an extended ARC with additional Big Data components, and open issues as well as future research directions.

This chapter is structured as follows. Section 5.2 provides a general, application-independent background knowledge about Big Data introducing V-based descriptors [Ziko 12] and a Big Data value chain [Mill 13]. V-based descriptors such as volume, variety, velocity, veracity, and value are often used to define Big Data [Chen 14b, Gand 15]. The Big Data value chain comprises Big Data concepts and tools [Chen 14b]. Section 5.3 contains a HAR-related definition of Big Data. The definition of Big Data in the context of HAR should provide a better understanding of the future changes in data characteristics. In section 5.4, an extended ARC is proposed which integrated Big Data concepts and tools. The integration of Big Data components in the design of HAR algorithms is mandatory in order to provide Big Data ready HAR systems. The chapter concludes with HAR-related open issues and future research directions (section 5.5). The contribution was submitted [Schu 18].

5.2 Background Knowledge

5.2.1 Definition of Big Data

The authors in [Chen 14b] stated that Big Data is an abstract concept and the definitions vary among people from different domains.

According to [Lane 01, Ragh 14, Chen 14a, Sing 14], Big Data can be defined via the three V's volume, variety, and velocity. Volume refers to the huge amount of data which are generated and collected. The size of stored data is often in the range of petabytes [Ragh 14]. Various data types including semi-structured, unstructured, and structured data are considered (high variety) [Bano 16]. The data collection and analysis must be rapidly and timely conducted (high velocity) [Ragh 14]. In [Suth 16], the three V's were referred to the machine learning domain. Volume was defined as the number of instances of the data. Variety was defined as the number of classes. Velocity was defined as the ratio between the number of instances and the period of time within the data were collected.

In [Ziko 12], veracity was added to the list of descriptors which refers to the quality or trustworthiness of data. An example of untrustworthy data is spam available in social networks such as Twitter.

An additional descriptor of Big Data is value [Gant 11, Ziko 12]. Usually, greater value should be achieved from acquired datasets through insights from Big Data Analytics.

5.2.2 Big Data Value Chain

The authors in [Mill 13, Chen 14b, Gand 15] provided a Big Data value chain which described four steps how to extract value out of Big Data. The Big Data value chain consisted of data generation, data acquisition, data storage, and data analysis.

The first step of the Big Data value chain is data generation [Chen 14b]. The main sources of Big Data include enterprise data, IoT data, biomedical data, and data of scientific applications.

The second step of the Big Data value chain is data acquisition and consists of data collection, transmission, and preprocessing [Chen 14b]. Three commonly applied data collection methods include sensors, log files, and web crawlers. The collected data are transfered to a data storage infrastructure. Since the collected data include noise, redundancy, and inconsistency, methods have to be applied to preprocess the data [Chen 14b]. The preprocessing techniques are mandatory to reduce storage expense and improve the analysis accuracy.

The third step of the Big Data value chain is data storage. The storage platform should organize the acquired information in a suitable format for subsequent analysis and should provide scalable access interface to query as well as analyze a huge amount of data [Hu 14].

The fourth step of the Big Data value chain is data analysis and aims at extracting useful values and performing decision making [Chen 14b]. Depending on the type of input data, data analysis can be divided into text analytics, web analytics, network analytics, multimedia analytics, and mobile analytics [Chen 12a, Chen 14b]. Independent on the type of input data, Big Data analysis requires access to frameworks for writing and running distributed applications in order to perform scalable processing [Whit 09].

5.3 HAR-Related Definition of Big Data

The proposed definition of Big Data for HAR includes the five V-based descriptors volume, variety, velocity, veracity, and value (section 5.2). Each descriptor is related to HAR and examples are further provided.

5.3.1 Volume

Big Data volume in HAR can also be defined by the size of stored data [Sing 14]. The size of HAR datasets is mainly dependent on number of sensors, number of sensor axes, sampling rate, acquisition duration, and number of subjects. In the following paragraphs, the size of a current dataset is compared to a future dataset.

Current HAR dataset sizes reached a dimension of e.g. megabytes. The HAR algorithm proposed in [Schu 13b] was evaluated on DaLiAc including data of 19 subjects (Table 1.1. Each subject was equipped with four IMU sensors (triaxial accelerometer, triaxial gyroscope). The sampling rate was set to 204.8 Hz and it is assumed that each sample was stored as one single-precision float (4 Bytes). Each subject performed a pre-defined protocol and the algorithm was trained and tested on pre-segmented data. The duration of pre-segmented data was about 20 minutes for each subject. The resulting dataset size is computed as:

$$Size = 4\, nodes\; x\; 6\; axes\; x\; 204.8\; samples/s\; x\; 4\; bytes\; x\; 20\; minutes$$

$$x\; 60\; s\; x\; 19\; subjects \approx 450\; megabytes.$$

In the IoT era, a large group of people is connected with things anytime which increases the stored data [Verm 11]. An estimation of a possible dataset size is given by an example scenario. A HAR algorithm should be trained and tested based on data from 20,000 subjects. Each subject is equipped with three sensor nodes: smartphone, smart watch, and a pod in the shoe. Each sensor node comprises a triaxial accelerometer, triaxial gyroscope, and triaxial magnetometer. The sampling rate is set to 50 Hz and each sample is stored as one single-precision float (4 Bytes). Data are acquired 24 hours during a period of 12 years. The resulting dataset size is computed as:

$$Size = 3\; nodes\; x\; 9\; axes\; x\; 50\; samples/s\; x\; 4\; bytes\; x\; 24\; hours\; x\; 3600\; s$$

$$x\; 365\; days\; x\; 12\; years\; x\; 20,000\; subjects \approx 41\; petabytes.$$

The previous example should show that typical dataset sizes can be reached which are commonly considered in Big Data.

5.3.2 Variety

Big Data variety in HAR can be twofold. First, variety can refer to the data types such as semi-structured, unstructured, and structured data [Ragh 14]. Second, variety can refer to the data content and environment. Content and environment can be interpreted as classes in the context of machine learning [Suth 16]. Examples of variety in HAR are given in the following paragraphs.

Current datasets for HAR were mainly stored as a structured matrix [Lara 13, Bull 14]. Columns included e.g. time stamps, data of sensor axes, and activity labels. Devices which are connected to the Internet generate structured, semi-structured, and unstructured data including Inertial-Magnetic Measurement Units, temperature, blood sugar, blood pressure, respiration, electrocardiography, electromyography data, click-stream data, social media data, video as well as audio, and text.

Besides the data types, variety in HAR could be further defined regarding the data content and environment in which data were captured. The data might vary e.g. in terms of complexity levels of activities (single, concurrent, composite, interleaved, group activities), subject characteristics (age groups, social origin), seasons (winter, spring, summer, autumn), and locations (outdoors, indoors).

5.3.3 Velocity

Big Data velocity in HAR refers to the rate at which data are created and speed at which data should be analyzed similar to the definition given in [Chen 14a].

Current HAR algorithms mainly performed offline analysis (Table 1.2). First, devices recorded and stored sensor data. Second, HAR was performed on the stored data. The data velocity was rather low. The IoT enables the creation and processing of a huge amount of data in a small amount of time (high velocity).

5.3.4 Veracity

Big Data veracity in HAR also refers to the quality or trustworthiness of data [Ziko 12]. HAR datasets were mainly acquired under controlled conditions with a pre-defined protocol (Table 1.1). A high quality and trustworthiness is usually provided in terms of labeled data which can be used to train algorithms. Since IoT data are captured in rather non-controlled environments, the quality and reliability of the data will decrease in this case.

5.3.5 Value

Value in current HAR systems might be limited due to the low amount of available data. The final output of the algorithms was the information which activity was performed by a subject considering data acquired in a short amount of time. The value based on these insights might be that it could be checked if physical activity recommendations were met or not (chapter 1.1). IoT and the availability of a huge amount of diverse data could provide more insights in the physical activity behavior of humans and could increase the data value.

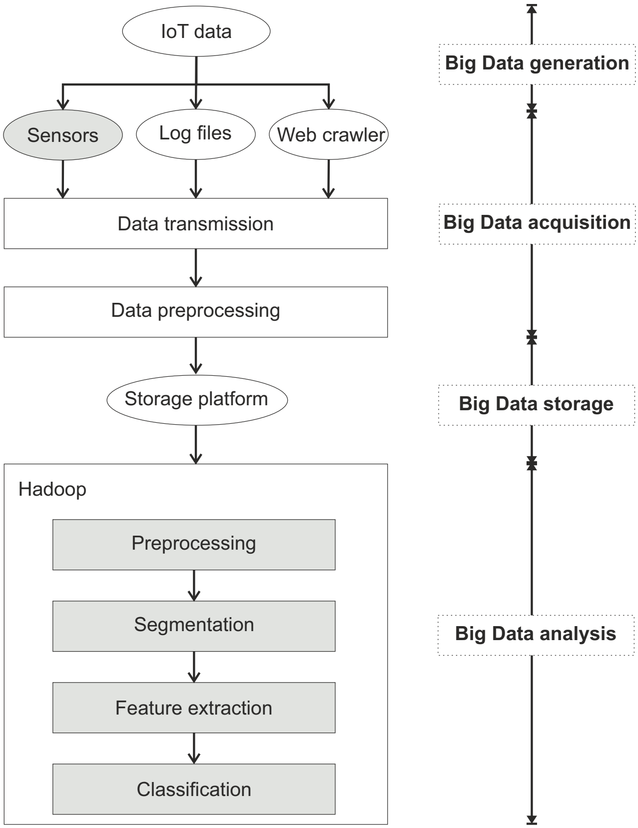


Figure 5.1.: Left: extended ARC, traditional ARC proposed by [Bull 14] is marked in grey. Right: Big Data value chain. Figure was adapted from [Schu 18].

5.4 Extended Activity Recognition Chain

In the following sections, components are identified which should be integrated in the traditional ARC (Figure 1.1) in order to be prepared for Big Data. The components were grouped according to the four steps of the Big Data value chain (section 5.2.2). The extended ARC is shown in Figure 5.1.

5.4.1 Big Data Generation

The traditional ARC does not include a specific data source in the design process (Figure 1.1). It is proposed to include IoT as main data source in the ARC in order to already connect HAR to Big Data at the beginning of the design process. IoT data for HAR might be available through smart wearables [Stei 15, Amft 17], smart cities [Jin 14], smart homes [De S 12] or smart cars [Walt 11].

5.4.2 Big Data Acquisition

The traditional ARC mainly contained the collection of a small amount of sensor data in the data acquisition step (Figure 1.1). It is proposed to add log files and web crawlers to the ARC (Figure 5.1). For HAR research, an example of a log file type might be a Twitter post. Web crawlers are tools which download and store content of web sites [Chen 14b]. Applications of log files and web crawlers are mentioned in section 5.5.

Compared to the traditional ARC, the future Big Data acquisition step should consider transmission of data to a common storage infrastructure as well. The transmission of data which were acquired by different research groups to a centralized platform enables a dataset fusion. This would address the limitations of current HAR systems regarding number of used sensors as well as sensor types, number of activities, and number of subjects (section 1.5.1). It is proposed to add a second data preprocessing step to the traditional ARC. Data preprocessing should contain two groups of methods. First, application-independent steps have to be applied such as data cleansing [Male 00], redundancy reduction [Chen 14a], privacy protection [Mach 12, Wu 14, Tsai 15], and data integration [Lenz 02]. Second, HAR-specific steps have to be applied, e.g. annotation of the data. Data annotation might be performed by crowdsourcing [Howe, Lesk 14, Wang 18]. Current HAR datasets are rather small due to the time-consuming data collection and annotation [Wang 18].

Crowdsourcing-based annotation would increase the amount of labeled data which can be used to train HAR systems.

5.4.3 Big Data Storage

The traditional ARC did not consider data storage, since data were mainly stored on local PCs, which required less effort (Figure 1.1). Due to the huge amount of diverse data in the future, it is proposed to add a storage platform to the ARC. A common centralized storage platform would reduce the search time for suitable input data which can be used to train HAR systems. The storage platform should provide file systems, databases, and programming models. Examples of the three components for HAR might be HDFS [Shva 10], HBase [Geor 11], and MapReduce [Dean 08], respectively. HDFS runs on commodity hardware and enables the storage of large files (hundreds of megabytes, gigabytes or terabytes). HBase is built on top of HDFS as a database. MapReduce allows the parallelization of large computations by using large-scale clusters of machines.

5.4.4 Big Data Analysis

Since currently available HAR algorithms were mainly evaluated on small data, frameworks for distributed data processing were not considered in the ARC (Figure 1.1). It is proposed to implement machine learning techniques for preprocessing, segmentation, feature extraction, and classification on a distributed framework. One example might be Hadoop, which is one of the most widely used framework for distributed data processing [Whit 09]. The previously mentioned HDFS file system, the HBase database, and the MapReduce programming model are part of Hadoop [Whit 09]. Approaches, which were comprised in Table 1.2, can often be implemented on Hadoop and re-used or adapted to Big Data. The applicability of MapReduce for machine learning was shown in [Chu 06]. Examples for MapReduce-based versions of machine learning techniques include SVM [Cata 13], Naive Bayes [Liu 13], C4.5 [Dai 14], Deep Belief Networks [Zhan 14], Expectation Maximization [Chen 13a], and Multilayer Perceptron [Zhan 16].

5.5 Open Issues and Future Research Directions

The previously mentioned V-based descriptors (section 5.3) and the proposed extended ARC (section 5.4) will change HAR research from a data, algorithm, and application perspective. The corresponding open issues and future research directions are discussed in the following sections.

5.5.1 Data Perspective

The focus of current research work in HAR on the data itself was limited (section 1.5.1). A limited amount of sensors were used and stored in a structured matrix. The high volume, variety, and low veracity of Big Data will change HAR research from a data perspective. Open issues and future research directions include the integration of new data sources in HAR research, a proper standardization of data, and an efficient data annotation.

New Data Sources

Major data sources in the IoT era include sensors, log files, and web crawlers (section 5.4). One major open issue is to integrate these data sources in HAR research.

A large variety of sensors might be used to measure among others humans' biological, physical, emotional, and cognitive state, which is important to analyse human's behavior [Tros 02]. Current HAR approaches mainly consider only a limited amount of sensor nodes and sensor types. A larger set of diverse sensors were often not yet available for research groups. Future research groups should investigate if these new sensors improve the performance of HAR algorithms. Log files such as Twitter posts might be used to correlate IMU-based activity profiles with psychological and social data of users [Abel 11, Eich 15]. Web crawlers might be used to select online material for HAR [Xu 14]. An automatic search might enable a faster provision of state-of-the-art HAR algorithms. Future research direction should include web crawlers which are optimized for HAR.

Standardization

Current provided publicly available datasets, listed in Table 1.1, differ in the used coordinate system, description of sensor placement, and used data format. Data were further stored on local PCs. As proposed in the extended ARC, captured data should be stored in a common database (Figure 5.1). Due to high data volume and variety, an open issue is a proper dataset standardization, which should be supported by relevant communities [Chen 12b]. A standardization should e.g. include the following items:

- **Common coordinate system**: proposed solution for a standardized coordinate system e.g. for IMUs is the usage of anatomical planes and axes for the different sensitivity axes of sensors (Figure 2.1). The advantage is a faster understanding of the orientation of the sensors placed, e.g. on the human body.
- **Description of sensor placement**: common textual description of exact sensor placement is often not mentioned in current research work, rather information like "placed at the chest" [Clel 13] or "at the wrists" [Ofli 13]. Future description of the study design should use e.g. anatomical directions and landmarks such as "near iliac crest/lateral side".
- **Data format**: data volume example which was provided in section 5.3 showed that in the future dataset sizes in the range of petabytes could be reached, which could be combined with important meta data such as subject characteristics, locations, and seasons (high data variety). Therefore, future researchers should investigate standardized data formats applicable for human activity data. Examples which are already applied in Big Data include Avro [1], Parquet [2], or ORC [3].

The provided aspects could be part of current standardization activities initiated e.g. by IEEE-SA [4] or VDI [5]. The standardization would leverage the interoperability of various devices and the collaboration of different research groups.

[1] https://avro.apache.org
[2] https://parquet.apache.org
[3] https://orc.apache.org
[4] https://standards.ieee.org
[5] http://www.vdi.eu/engineering/vdi-standards/

Data Annotation

The low veracity of Big Data requires smart and efficient data annotation tools. As pointed out in section 5.4, crowdsourcing could be used to annotate a huge amount of sensor data. Nevertheless, research work about crowdsourcing-based annotation is limited [Lase 13, Feng 14]. In [Feng 14], the challenge of stimulating smartphone users to join mobile crowdsourcing applications with smartphones was addressed. The proposed framework achieved individual rationality, truthfulness, and high computational efficiency. Nevertheless, open issues include e.g. the identification of misreports when offering support as well as the privacy protection of the individual users.

In [Lase 13], an active learning approach was applied to interactively respond to novel activities. If a new instance cannot be classified, crowd workers were asked to label the activity. The proposed system combined the input of multiple workers into a single ordered label set by a graph-based approach. Future research should include the investigation of probabilistic models in active learning from crowds [Rayk 10, Yan 11]. Probabilistic methods further enable both a selection of the best data sample and annotator [Yan 11].

5.5.2 Algorithm Perspective

Open issues and future directions from an algorithm perspective could be subdivided into two groups. The first group is connected to the research directions which were mentioned in section 1.5.1 including fusion strategies, classification architectures, user-specific systems, and learning techniques. The second group is connected to components in the extended ARC e.g. data transmission and data preprocessing as well as distributed frameworks (Figure 5.1).

Fusion Strategies

State-of-the-art research in HAR mainly focused on fusion techniques based on a limited amount of sensors, e.g. IMUs, GPS, and heart rate (Table 1.2). The fusion was performed at data level, feature level, or decision level. The high variety in Big Data requires novel fusion techniques combining classifier decisions which are performed at different ages of humans or at different seasons. If HAR algorithms are implemented on embedded devices, energy efficient implementations should be further considered [Caus 16]. As mentioned in section 5.4, Big Data analysis can be subdivided into text analysis, web analysis, network analysis, multimedia analysis, and mobile analysis. Future research work in HAR should further consider a combination of these analysis types in order to better understand humans' behavior.

Classification Architectures

Current algorithms for HAR were often either flat or hierarchical (section 1.5.1). In addition, the final classes were the activities itself. Due to the high variety in Big Data e.g. in terms of subject characteristics like social origins or age categories, a single activity-based architecture type might not be sufficient. An open issue includes novel architectures addressing classifiers for different social origins, age categories, and activities. One possible solution could be to implement first a hierarchical architecture classifying different social origins followed by an origin-dependent age category classifier. In the last step, a flat architecture could be used to recognize single activities.

User-Specific Systems and New Learning Techniques

Since a huge amount of data might be available for each single human, the ARC might be implemented as a user-dependent and personalized system. In order to provide an adaptable system, lifelong learning strategies might be beneficial. According to [Chen 16], lifelong learning is a continuous learning process. The learner can use past knowledge to support the learning of a new task. The learning might never end, become more and more knowledgeable, and improve the system performance. Preprocessing methods for lifelong learning might aim at reducing the redundancy of the gathered information and keeping only necessary attributes for lifelong learning approaches. In order to further improve the accuracy of HAR algorithms, a user relative coordinate system might be beneficial instead of a device or world coordinate system [Jahn 17].

Data Transmission and Preprocessing

The traditional ARC mainly considered algorithmic concepts from a machine learning point of view (Figure 1.1). The high volume, variety, and velocity of Big Data further requires algorithmic concepts for data transmission and data preprocessing such as inconsistency detection. These algorithmic concepts are often application-independent and should be considered in future HAR systems.

Distributed Frameworks

Only limited research about Hadoop-based algorithms for HAR exists so far. The authors in [Pani 12] acquired accelerometer data on a smartphone and sent the data to a cloud. HAR approaches were applied in the cloud using the Hadoop framework. Although approaches were implemented using a distributed framework, the algorithm was evaluated on about 500 MB of data. In [Bano 16], a cloud-based digital health and wellness framework was developed. Although an activity recognizer was implemented using a distributed framework, the corresponding algorithm was evaluated on ten subjects, who performed single activities such as walking, stretching, or sweeping. A HAR technique was developed and implemented in [Han 12] which was part of a Hadoop-based platform [Fahi 14]. The proposed technique was tested on ten subjects. The mentioned examples showed that distributed frameworks were often developed, but the implemented techniques were trained and tested on a limited amount of data. An open issue includes the performance analysis using Big Data.

Although Hadoop is most widely used for distributed data processing, the framework has different disadvantages [Agne 14]. For iterative algorithms, temporary data of each iteration has to be stored and loaded after each iteration. In interactive querying, a new set of MapReduce jobs has to be performed for each query not considering query history. Thus, for iterative machine learning tasks, an alternative distributed programming framework like Spark has to be used [Zaha 10]. Future researchers should compare different frameworks and choose the best one for a certain application.

5.5.3 Application Perspective

Current HAR approaches have mainly estimated the single activity that was performed. The applications which were mentioned in section 1.5.1 provided a limited value to human activity behavior analysis. One example was the comparison of the current physical activity level to predefined recommendations. The availability of Big Data and the proposed extended ARC (Figure 5.1) might increase the value which is inside the captured data and might enable the development of a more detailed analysis of humans' behavior.

Big Data analysis could be applied to identify factors which have a relationship to the participation in physical activity. The authors in

[Tros 02] classified the factors which are associated with physical activity as demographic, emotional, cognitive, psychological, social, physical, cultural, and environment. The identification of relevant factors can be used to design effective intervention programs which aim at increasing the participation in regular physical activity.

A longitudinal analysis which aggregates the information captured from child to elderly might further enable the recognition of unknown diseases or the automated detection of deviant behavior [Ludw 12].

The previously mentioned application examples require a common understanding of Big Data and knowledge about Big Data tools in order to solve the upcoming challenges. The content of this chapter can be seen as a starting point for future research in Big Data ready HAR.

Chapter 6

Summary, Discussion, and Conclusion

Human Activity Recognition (HAR) deals with the automatic recognition of physical activities and plays a major role in health and sports (see section 1.1). Nine contributions were made in the field of HAR and were presented in the previous chapters. In chapter 3 and 4, algorithms were presented, which were able to infer daily and soccer-specific activities from IMU data using machine learning techniques. Chapter 5 provided an introduction of Big Data to the HAR research field. In this chapter, an overall summary, discussion, and conclusion are given based on the findings of the previous chapters.

This chapter is structured as follows. First, the nine contributions of this thesis are summarized and discussed. The contributions were divided into three groups corresponding to the three chapters 3, 4, and 5. Second, a general outlook of future research in HAR is given. Third, conclusions are drawn from the presented findings in this thesis.

6.1 Summary and Discussion of Contributions

6.1.1 Recognition of Daily Activities

As pointed out in section 1.1, activity monitors were recommended in order to provide information about meeting physical activity recommendations, investigate causes of physical activity behavior, and assess the effect of intervention programmes. Activity monitors often include HAR systems, which determine activity type and activity intensity using IMU data as input for machine learning approaches. The literature review showed that a large number of research papers was available dealing with the recognition of daily activities based on these two parameters (section 1.5). Nevertheless, various limitations were found. Current HAR methods were often not flexible enough to react to environmental changes such as adding sensors or activities, were based on a single decision of a classifier, evaluated new methods on isolated small datasets, and compared novel approaches to state-of-the-art approaches based on the achieved performance not considering the different used input data. In this thesis, five contributions were presented, which addressed the mentioned limitations.

Contribution 1:
Hierarchical Classification System

In section 3.3, a hierarchical classification algorithm for HAR was developed (Figure 3.4). One local subsystem was trained per parent node recognizing only a subset of activities. An individual ARC was implemented for each subsystem consisting of preprocessing, segmentation, feature extraction, and classification (Figure 1.1). The subsystems were optimized for the corresponding activity subset resulting in a subsystem-specific feature set and classifier. The proposed hierarchical classification system achieved a balanced accuracy of 89.1 %.

Contribution 2:
Decision Level Fusion

An algorithm for HAR was presented in section 3.4, which combined the information of four sensor nodes at decision level (Figure 3.10). An individual ARC was implemented for each sensor node consisting of preprocessing, segmentation, feature extraction, and classification (Figure 1.1). The decision level fusion was performed by majority voting and highest rank method. The proposed algorithm achieved a balanced accuracy of 85.7 %.

Contribution 3:
Comparison of Algorithms Using a Common Evaluation Framework

In section 3.5, a common evaluation framework was introduced, which enabled a comparison of various algorithms for HAR (Figure 3.12). The proposed framework included the selection of a benchmark dataset, algorithm-specific data preparation, execution of re-implemented ARCs, and definition of experiments as well as evaluation. As an example, the framework was applied to two algorithms developed in this work (section 3.3 and 3.4) as well as to five re-implemented state-of-the-art algorithms [Bao 04, Ravi 05, Park 06, Pree 09, Liu 12]. The two proposed algorithms outperformed the state-of-the-art techniques in terms of balanced accuracy, but required more sensors.

Contribution 4:
Database Fusion Strategy

A novel database fusion algorithm was introduced in section 3.6. Sensor data of various publicly available databases were selected, prepared, and merged. The applicability of the proposed technique was shown for an example algorithm for HAR (Figure 3.15). The number of subjects, activities, and sensors could be increased from nine to 33, five to eight, and one to four, respectively. The proposed activity recognition algorithm achieved a balanced accuracy of 87.2 %.

Contribution 5:
Energy Expenditure Estimation

In section 3.7, a system was presented, which estimated the energy expenditure in MET based on IMU data from hip and shoe sensors. Preprocessing, segmentation, feature extraction, and regression were applied to data of four sensor sources separately. The individual predictions were fused at decision level (Figure 3.18). Due to a good trade-off between a low median RMSE of 0.42 MET and a low interquartile range of 0.12 MET, the combination of accelerometers attached to hip and shoe was proposed.

Implications

Three HAR systems were developed in order to infer the type of daily activities from IMU data, which provided a high flexibility in terms of adding new sensors and/or activities as well as combined multiple decisions of individual systems (contributions 1, 2, and 4). All algorithms implemented the ARC introduced in [Bull 14]. The approaches were similar in the single components of the ARC. A sliding window was applied to accelerometer and gyroscope data. State-of-the-art features were extracted (Table 2.1) and often reduced by a feature selection routine. The features were used as input for various classifiers. Accuracy values above 85.7 % showed the general applicability of the proposed approaches to activity recognition challenges. The proposed feature set for the classification of the activity type was further used for the determination of the activity intensity. In addition, the decision level fusion proposed in section 3.6 was adapted to the energy expenditure estimation (contribution 5). A RMSE of 0.42 MET showed that some selected components were applicable for both determining activity type and activity intensity.

Two common advantages of the mentioned algorithms were found. First, the techniques provided a high flexibility to react to environmental changes. Single trained subsystems in the hierarchical as well as in the decision level fusion approaches could be replaced or adapted without affecting the other subsystems. A retraining of the complete HAR system is not necessary. If the developed algorithms are implemented on deployed activity trackers, which were e.g. mentioned in section 1.6, parts of the HAR algorithm can be updated in terms of adding new sensors or activities without replacing the complete system. The second common advantage of the algorithms was that the activity type or intensity determination was not dependent on a single system decision. In the hierarchical approach, multiple decisions were performed along the path of the hierarchy. In the decision level fusion approaches, multiple decisions were made by majority voting or averaging estimations.

A disadvantage of the proposed techniques included e.g. the number of required sensors. All methods considered multiple sensor nodes as input, which increased the cost of the overall system. Nevertheless, the hierarchical and decision level fusion approaches can be adapted, if e.g. only one sensor is available. The hierarchical approach could be trained with data of only one sensor node. The feature selection routine might determine more features in order to capture the information of the desired activities. The decision level fusion technique described in section 3.4 was not applicable to only one sensor, but the extension introduced in section 3.6 supported a single sensor setup, at least if multiple sensor types were available. Information acquired by accelerometer and gyroscope could be fused e.g. with magnetometer data, which was often integrated in current sensor nodes (Table 1.1). A further step could be the combination of hierarchical and decision level fusion approaches. The hierarchical structure could be implemented for one sensor type separately, which includes the decision level fusion along the path in the hierarchy, and a final decision level fusion across all sensor types.

The algorithms listed in Table 1.2 mainly performed flat classification based on feature level fusion (Figure 1.2b and 1.3a). Integrating new activities or sensors might often be cumbersome and would require a retraining of the complete HAR system. As already mentioned, the proposed HAR systems might be more flexible to react to environmental changes. Nevertheless, various components of state-of-the-art approaches should be integrated in future work. Probabilistic methods in the decision level fusion using e.g. the posterior probability as weighting factors should be investigated, which showed a better classification performance [Chow 18]. Implementation of the proposed HAR approaches on embedded devices might further require a reduction of the sampling rate. The method introduced in [Khan 16] could be applied to determine the optimal sampling rate based on a similarity analysis. Currently, the developed algorithms were designed as user-independent systems [Bull 14]. Since the determined activity information is mainly person-dependent, components for a user-specific system as proposed in [Van 00, Alle 06, Rogg 13] might be beneficial to increase the overall system performance. Although the implemented algorithms in this thesis achieved accuracy values above 85.7 %, the study design included mainly single activities performed in a rather controlled environment. Authors in [Wu 07b, Hu 08, Blan 10, Gord 11, Hu 09] further investigated the application of HAR to more complex scenarios such as concurrent, interleaved, composite, multi-user group activities, and abnormal activities. The performance of the methods provided in this thesis might drop under these complex scenarios and additional components should be integrated.

The developed evaluation framework and database fusion strategy were rather new concepts (contribution 3 and 4). The evaluation framework could be one step towards a fair quantitative assessment of the best HAR algorithm for a specific application. A comparison of different methods is not performed by investigating the achieved performance based on different input datasets. The effect of different novel components of an ARC on the classification performance can be investigated. Combining different components of the methods listed in Table 1.2 might result in HAR systems, which are highly robust, flexible, and person-specific and provide accurate activity profiles. One step in the proposed framework included the selection of a common benchmark database. In order to increase the number of instances, the database fusion strategy could be further integrated in the framework (Figure 3.14). Fusing already available datasets can be seen as low-cost solution for providing HAR systems with more data without the need of an additional data collection session, which is usually time-consuming.

All in all, the developed HAR systems in this thesis showed suitable performance to infer daily activities. Implementing these systems e.g. in future activity trackers enables the provision of detailed and person-specific activity profiles. Due to the rather generic components of the algorithms and the high flexibility of the proposed methods, the set of activities and sensor settings could be optimized for each human individually. The person-specific activity profiles provide information about meeting physical activity recommendations [Warb 06], give insides in the causes of physical activity behavior [Sall 00], and show the effect of intervention programmes [Kahn 02].

6.1.2 Recognition of Soccer Activities

Understanding activity patterns in soccer enables coaches to prescribe and implement training programs that replicate physical demands during competition. The performance of young talents during various activities such as shooting, dribbling, and sprinting is further determined in order to identify talents. Thus, knowledge of the performed activity is beneficial in sports. The literature review showed that IMUs can be applied to recognize soccer-specific activities such as dribbling, sprinting or kicking (section 1.5.3). Nevertheless, various limitations were found. The authors often achieved a low accuracy regarding kicks, the algorithms were evaluated on rather controlled scenarios such as pre-defined exercises or pre-selected game scenes, and a distinction of the kick intensity was only based on stationary kicks. In this thesis, three contributions were presented, which addressed the mentioned limitations and showed the integration of HAR algorithms in a novel and innovative application.

Contribution 6:
Kick Classification

In section 4.3, two kick classification routines were introduced (Figure 4.7). The algorithms were able to distinguish between full-instep and side-foot kicks. The first algorithm was based on a hierarchical classification architecture (Figure 1.3b). The second algorithm exploited expert knowledge about the kick execution. The proposed methods achieved a balanced accuracy above 94 % on a set of pre-defined exercises. A balanced accuracy of 85.0 % was further achieved considering data of a complete 45 minute game.

Contribution 7:
Ball Speed Estimation

An IMU-based ball speed estimation technique was presented in section 4.4. The algorithm segmented different phases during a full-instep kick followed by feature extraction (Figure 4.12). MLR was further applied to predict the achieved ball speed. The proposed system achieved a MAE of 5.5 ± 0.9 km/h during dynamic exercises.

Contribution 8:
Video Summary Generation

In section 4.5, a sensor-driven, automatic, and low-cost solution for video summary generation was described. Players were equipped with inertial sensors and simultaneously captured by video cameras (Figure 4.15). HAR algorithms such as the previously mentioned kick classification routines and the ball speed estimation were applied to identify player-specific highlights in the sensor data (Figure 4.16). The corresponding video frames were selected and integrated in a final video summary. The ball speed estimation could be used to rank full-instep kicks based on the achieved ball speed. An extension of the system added video and audio effects to single highlight scenes.

Implications

Two HAR algorithms were developed in order to infer soccer-specific activities from IMU data, which were optimized for two of the main kick types in soccer (full-instep and side-foot kick). The algorithms implemented the ARC introduced in [Bull 14]. Both approaches relied on peak detection routines, which reduced the number of classification tasks by about 50 %. The proposed peak detection might further be applicable for other soccer-specific activities which include a ball contact, e.g. long passes or dribbling sequences. The proposed approach outperformed an existing approach considering a pre-defined set of exercises [Ahma 15]. Evaluating the developed algorithm on game data showed a lower performance compared to literature [Mitc 13], but the evaluation in this thesis was on more realistic data, since complete game data was considered.

One challenge in the classification of soccer-specific activities is the fact that the target activities are rather rare compared to the duration of the data acquisition. Although high sensitivities of kicks above 90 % were achieved, the precision was often low, especially in complex scenarios such as games. This shows that the proposed systems predicted more instances as kicks than actually available in the dataset. Nevertheless, an investigation of the false positive instances showed that often ball contacts were identified as kick candidates. Future research should focus on the integration of more ball contact activities in the proposed system. After eliminating the false positive instances, which included ball contacts, only 10 instances remained in average for each player compared to the acquired game duration of 45 minutes. Thus, the proposed algorithm can be seen as a good starting point for a more advanced system. One processing step on top of the current system could include concepts of process mining [Aals 11]. A ball contact classification system could generate an event log and a process model is structured capturing the behavior seen in the mentioned log. Another improvement could be an unsupervised online learning component in order to model the NULL class and try to further eliminate false positives.

The ball speed estimation algorithm achieved an accuracy of 5.5 km/h which was 2-3 times higher than the accuracy of the Hawk Eye system. Thus, the proposed system cannot replace expensive camera-based systems, but the system might be used to determine a rough categorization of kick speed in low, medium, and high intensity. Although the proposed approach achieved lower performance compared to current existing literature [Zwic 12], the method was evaluated on more subjects and on more realistic scenarios including dynamic training exercises. In order to further improve the accuracy, additional aspects such as kicking technique, approach angle, and support leg dynamics should be investigated.

All in all, the developed HAR systems in this thesis showed suitable performance to infer full-instep and side-foot kicks. Due to the low precision, the proposed system might not be applicable to create detailed match statistics about the number of performed kicks and replace current camera-based ground truth systems in professional sports. Nevertheless, amateur teams might benefit from such a low-cost solution. Coaches and scouts from local recreational teams could quantitatively compare the activity patterns of players. Besides comparing activity patterns, a novel and innovative application was further introduced in this thesis, which included the developed HAR systems (contribution 8). A video summary system provided each player with individual highlight reels. Besides a ranked order of kicks further highlights might include tricks or fouls.

6.1.3 Big Data Ready Human Activity Recognition

Wearable sensors are one of the biggest drivers of the IoT, which offers the generation of Big Data. Although Big Data offers new possibilities in HAR research, the concept of Big Data is not considered in the design of HAR systems so far. To close this gap, one contribution was presented in this thesis which introduced the abstract concept of Big Data to the HAR research field.

Contribution 9:
Introduction of Big Data to HAR Research

The contribution regarding Big Data for HAR was fourfold. First, a survey about state-of-the-art in wearable-based HAR was presented (section 1.5.1). It was pointed out that although Big Data could have a positive impact on HAR research, the current version of the ARC did not consider Big Data components. Therefore, the integration of Big Data concepts and tools in the ARC is mandatory as well as a general introduction of Big Data to the HAR community. Second, a HAR-related definition of Big Data was given using the five V-based descriptors volume, variety, velocity, veracity, and value (section 5.3). Third, extensions of the traditional ARC were introduced (section 5.4). Examples included a centralized storage platform and the usage of distributed data analysis frameworks such as Hadoop. Fourth, open issues and future research directions were comprised including standardization techniques for datasets and novel fusion strategies (section 5.5).

Implications

The connection between Big Data and HAR might be one important milestone in activity recognition research. In the past, research work focused on the assessment of sensors for activity recognition challenges, basic algorithms e.g. for data fusion as well as hierarchical architectures or user-dependent systems, and the evaluation of new approaches on small isolated datasets (section 1.5.1, Table 1.1 and 1.2). In the future, new research fields could be defined such as lifelong learning and the application of HAR algorithms to complex activity recognition. Research work might benefit from the availability of Big Data. The proposed definition of the term Big Data in HAR and the proposed extensions of the traditional ARC should be seen as one step towards Big Data ready HAR. The suitability of the definition and the extensions has to be proven and possible changes might be required. The proposal for a storage platform enables the provision of a large-scale benchmark database which would have a high impact on future investigations in activity recognition and is a mandatory step. Nevertheless, it requires high and time consuming effort in defining a structured data gathering and a collaboration of various entities. As a first step, the database could be setup in an iterative manner, to test new proposed standards. Data coming from different activity trackers, e.g. which were mentioned in section 1.6, could be further integrated in this new database.

All in all, Big Data ready HAR systems could have a high impact in the future assessment of the physical state of humans. Since more and more sensors are connected to the internet, detailed and person-specific information can be gathered to provide individual activity profiles. The causes of physical activity behavior, as mentioned in section 1.1, could further be investigated over a long period of time.

6.2 Outlook

In the future, the proposed algorithms and concepts, which were part of the contributions 1 to 9, could be further combined to a general architecture for future HAR. The architecture consists of a common database, an algorithm library, and an evaluation framework (Figure 6.1). Details of the three parts are given in the following sections. In addition, a future medical application of the proposed video summary tool is given.

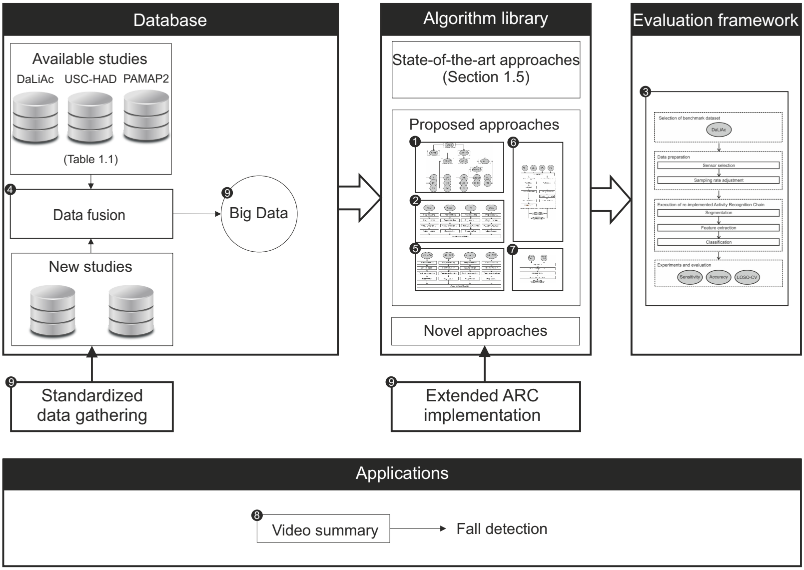


Figure 6.1.: Outlook. Combination and extensions of proposed algorithms and concepts, which were part of contributions 1 to 9. Contribution numbers are indicated by black circles.

6.2.1 Database

A standardized data gathering could be used to include new studies in a common database. Standardization includes a common coordinate system, description of sensor placement, and data format (contribution 9). Data should further be annotated by crowdsourcing tools. Study data might be collected by the devices mentioned in section 1.6. The effort in the standardized data gathering steps varies with respect to the application area. Labeling walking periods in daily life might be easier than labeling every ball contact in specific sports types. The common database further might include already available studies listed in Table 1.1, e.g. DaLiAc, USC-HAD, and PAMAP2. Datasets which were acquired in various studies could be combined with the proposed fusion strategy (contribution 4). The fusion approach could further be extended by a component which selects appropriate studies for a specific application.

All studies together result in Big Data, which are available for algorithm development as well as algorithm evaluation (contribution 9).

6.2.2 Algorithm Library

Algorithms, which were part of the contributions 1, 2, 5, 6, and 7, could be integrated in an open access algorithm library. Such an algorithm library would provide a variety of different implementations of the ARC introduced in [Bull 14]. Single components of the implementations could be reused in other methods, which might result in a faster development of robust HAR systems in the future. The library should include state-of-the-art approaches mentioned in section 1.5 as well as novel approaches.

Novel approaches should implement components of the proposed extended ARC (contribution 9). Machine learning tasks should e.g. be executed on Big Data frameworks such as Hadoop. Novel HAR approaches should further contain generic, self-configurable, and modular components, which could be applied to varying sets of sensors as well as activities and could be replaced by other components. Examples for these generic and modular components were introduced in this thesis. Algorithms, which were described in section 3.3 and 4.3, included an automatic feature selection or peak detection routines, which were optimized for a given activity set. The proposed peak detection algorithm could further be applied to other applications, which require the recognition of peak-rich parts in sensor signals, such as fall or step detection.

The proposed hierarchical system shown in Figure 3.4 contained one SPORTS subsystem, which could be replaced in future by the kick classification algorithms presented in section 4.3. With such a system, soccer players could be analyzed in sports as well as in other life situations. Future systems should further provide solutions for other sports types such as basketball or swimming. The proposed energy expenditure estimation system (Figure 3.18) could be added as a subsequent module after the WALK subsystem (Figure 3.4) in order to provide information about both activity type and intensity. Similar energy expenditure estimation modules could be developed and integrated for the HOUSE or BICYCLE subsystem (Figure 3.4).

6.2.3 Evaluation Framework

The proposed approach for a common evaluation framework (contribution 3) could be used to compare a novel approach to already existing algorithms in the library. The framework can identify the best approach for a specific application based on a quantitative and fair assessment using a common benchmark database. The result of the comparison could be used to further improve novel approaches and to support the identification of possible bottlenecks. A more advanced extension of the evaluation framework might be to automatically combine parts of different algorithms, which might perform better than each individual technique. Parts of algorithms include e.g. median filtering or extracting a certain feature set. The automatic kick segmentation technique, which was described in section 4.3, could be further applied to other applications as a data preparation step for certain experiments. The technique could e.g. be used to provide a set of segmented steps for gait analysis.

6.2.4 Applications

The proposed video summary tool (contribution 8) could be further applied in medical scenarios such as fall detection for elderly.

Inertial sensors were often used to detect falls which enables an intelligent surveillance system for elderly [Muba 13]. The proposed video summary tool could be adapted to the fall detection scenario. Highlight videos which include periods of falls could be shown to physicians. Video frames before and after a detected fall could be used to investigate causes of falls or to assess the subsequent treatment of relatives.

6.3 Conclusion

In this work, various contributions regarding IMU-based HAR in daily life and sports were made. Seven HAR algorithms were developed which were able to infer either daily or soccer-specific activities based on activity type or intensity. The approaches implemented a traditional Activity Recognition Chain (ARC) and were optimized for different scenarios. The soccer-specific HAR algorithms were further integrated in an automatic, sensor-driven video summary application. The system was able to provide each individual player with a personal video summary containing highlight scenes.

The implemented ARCs were part of higher level system architectures. Generic hierarchical systems were developed, implemented, and evaluated on daily and soccer-specific activities. Various activities were ordered along a path of a class hierarchy. The grouping of activities and subsequent classification systems allowed better flexibility in terms of changing sensor configuration or the considered activity set. Decision level fusion techniques were further developed which addressed the influence of sensor degradation during run-time and the corresponding drop in performance of HAR systems.

Besides various implementations of the ARC, two novel HAR algorithms were further proposed. First, a database fusion strategy was developed and investigated which aimed at increasing the amount of instances which could be used to train and test machine learning approaches. The technique did not require an additional time-consuming data collection session. Second, a common evaluation framework was presented which could be used to provide a fair comparison of various HAR algorithms.

Big Data concepts and tools for HAR systems were further introduced. In detail, a HAR-related definition of Big Data was provided, various extensions of the traditional ARC were proposed, and open issues were argued.

The contributions, which were presented in this thesis, enable a better assessment of humans' behavior in health and sports. Future multi-sensor HAR systems, which implement the proposed extended ARC and are trained on Big Data, will provide a robust, holistic, and long-term analysis of humans' physical state. A detailed monitoring of humans' physical state enables (a) an early detection of sedentary behavior which supports a healthier society and (b) a quantitative athlete assessment which might provide a fast identification of future talents.

The findings of this thesis can be transferred to other research fields as well. Clinical gait analysis can e.g. benefit from the application of Big Data techniques to a huge amount of sensor data from Parkinson patients. The step pattern could be analyzed over a long period of time. A smart multi-sensor system can be implemented which estimates the severity of the disease and recommends a suitable therapy plan. The proposed algorithms can be used to provide flexible and personalized systems which are optimized for single patients. These systems could improve the quality of life of each individual patient.

Chapter A

Data Collection

A.1 Description of Soccer Training Exercises

(a) Walking. **(b)** Running. **(c)** Side-stepping.

Figure A.1.: Walking, **running**, and **side-stepping** (exercise IDs 1-3, Table 4.1) were performed between two cones without the ball. The distance between the cones was about 16 m. Walking was performed once and included both a single-support phase and a double-support phase, when one foot and both feet are on the ground, respectively [Bart 07]. Running was performed three times and included a single-support phase and an additional recovery phase, when both feet are off the ground. In the first two trials, the player had to run with two self-selected increasing speed levels in forward direction. In the third trial, the player had to run with one self-selected speed level in backward direction. Side-stepping was performed twice, from left to right and from right to left.

(a) Distance: 4 m **(b)** Distance: 9 m **(c)** Distance: 15 m

Figure A.2.: Eight subsequent side-foot kicks to a teammate were performed considering the three pre-defined distances 4 m, 9 m, and 15 m (exercise ID 4, Table 4.1). The three distances were chosen to vary the intensity of the passes. For each distance, the pass sequence was separately performed with left and right leg.

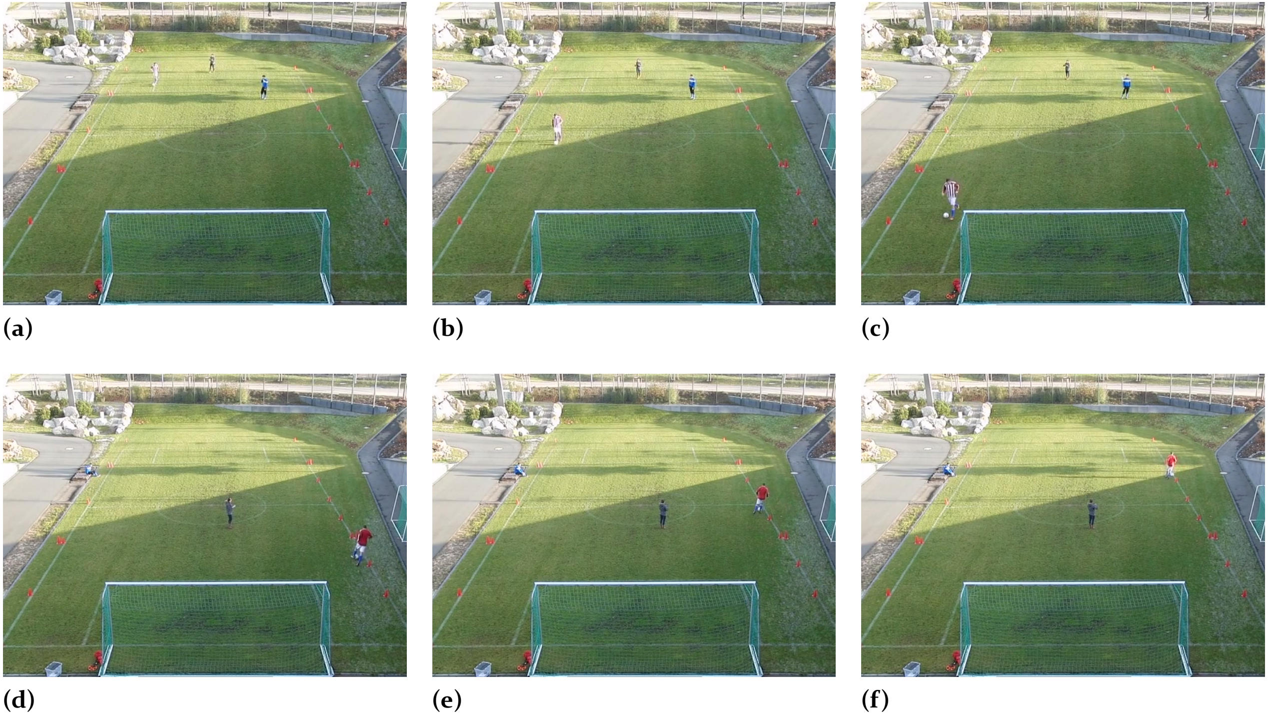


Figure A.3.: Dribbling (A.3a - A.3c) and **slalom** (A.3d - A.3f) were performed three times (exercise IDs 5 and 6, Table 4.1). Both exercises were performed with the preferred leg on a distance of 32 m.

(a) (b) (c)

Figure A.4.: Three **full-instep kicks on a stationary ball** (exercise ID 7, Table 4.1). Exercise was performed with left and right leg.

(a) Slalom **(b)** Side-foot kick **(c)** Jogging

Figure A.5.: Slalom between cones was performed with a subsequent **side-foot kick** to a teammate (exercise ID 8, Table 4.1). Player had to **jog** back to the starting position.

(a) **(b)** **(c)**

Figure A.6.: Player had to **control** the ball with the **thigh** with a subsequent **side-foot kick** to a teammate (exercise ID 9, Table 4.1).

(a) (b) (c)

Figure A.7.: Player had to **control** the ball with the **chest** with a subsequent **side-foot kick** to a teammate (exercise ID 10, Table 4.1).

(a) Dribbling **(b)** Side-foot kick **(c)** Dribbling

Figure A.8.: Player had to perform a **dribbling** with a subsequent **side-foot kick** to a teammate (exercise ID 11, Table 4.1). Player received the ball and had to **dribble** back to the starting position.

(a) Dribbling **(b)** Side-foot kick **(c)** Full-instep kick

Figure A.9.: Player had to perform a **dribbling** with a subsequent **side-foot kick** to a teammate (exercise ID 12, Table 4.1). Player received the ball and had to perform a **full-instep kick**.

(a) Dribbling **(b)** Tackling **(c)** Full-instep kick

Figure A.10.: Player had to perform a **dribbling** followed by a **tackling** against an opponent (exercise ID 13, Table 4.1). Player had to finish exercise with **full-instep kick**.

(a) (b) (c)

Figure A.11.: Interactive player test was a pass-exercise (exercise ID 14, Table 4.1). Four walls were placed around the player. Behind each wall, a flash light was positioned. During the exercise, one flash light was randomized activated and the player had to pass towards the corresponding wall. 15 passes had to be performed.

Abbreviations

ARC Activity Recognition Chain

AS Ascending stairs

BC100 Bicycling (100 watt)

BC50 Bicycling (50 watt)

BMI Body-Mass-Index

CART Classification and Regression Trees

DS Descending stairs

FN False Negatives

FPR False Positive Rate

FP False Positives

HAR Human Activity Recognition

IMU Inertial Measurement Unit

IoT Internet of Things

k-NN k - Nearest - Neighbor

LOSO-CV Leave-One-Subject-Out Cross-Validation

LY Lying

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MET Metabolic Equivalent of Task

MLR Multiple Linear Regression

NB Naive Bayes

RF Random Forest

RJ Rope jumping

RMSE Root Mean Square Error

ROC Receiver Operating Characteristic

RU Running

SEE Standard Error of Estimate

SI Sitting

SMA Signal Magnitude Area

SMV Signal Magnitude Vector

ST Standing

SVM Support Vector Machine

SW Sweeping

TPR True Positive Rate

VC Vacuuming

WD Washing dishes

WK Walking

List of Figures

List of Tables

Bibliography

[Aals 11] W. M. P. van der Aalst. *Process Mining*. Springer Berlin Heidelberg, 2011.

[Abel 11] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. "Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web". In: *Proc ESWC '11*, pp. 375–389, 2011.

[Agga 11] J. K. Aggarwal and M. S. Ryoo. "Human Activity Analysis: A Review". *ACM Comput Surv*, Vol. 43, No. 3, pp. 16:1–16:43, 2011.

[Agga 13] C. C. Aggarwal. *Managing and Mining Sensor Data*. Springer, New York, NY, USA, 1st Ed., 2013.

[Agne 14] V. S. Agneeswaran. *Big Data Analytics Beyond Hadoop: Real-Time Applications with Storm, Spark, and More Hadoop Alternatives*. Pearson Education, Upper Saddle River, NJ, USA, 2014.

[Ahma 15] A. Ahmadi, E. Mitchell, C. Richter, F. Destelle, M. Gowing, N. E. O'Connor, and K. Moran. "Toward Automatic Activity Classification and Movement Assessment During a Sports Training Session". *IEEE Internet Things J*, Vol. 2, No. 1, pp. 23–32, 2015.

[Ahn 16] J. G. Ahn, J. K. Park, and J. T. Kim. "Investigating Activity Recognition using 9-Axis Sensors and Filters in Wearable Devices". *IJECE*, Vol. 10, No. 9, pp. 1213–1216, 2016.

[Ains 00] B. E. Ainsworth, W. L. Haskell, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, W. L. O'Brien, D. R. J. Bassett, K. H.

Schmitz, P. O. Emplaincourt, D. R. J. Jacobs, and A. S. Leon. "Compendium of Physical Activities: An Update of Activity Codes and MET Intensities.". *Med Sci Sports Exerc*, Vol. 32 (supplement 9), pp. S498–S504, 2000.

[Ains 11] B. E. Ainsworth, W. L. Haskell, H. S. D., N. Meckes, D. R. J. Bassett, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon. "2011 Compendium of Physical Activities: A Second Update of Codes and MET Values". *Med Sci Sports Exerc*, Vol. 43, No. 8, pp. 1575–1581, 2011.

[Ains 93] B. E. Ainsworth, W. L. Haskell, A. S. Leon, D. R. J. Jacobs, H. J. Montoye, J. F. Sallis, and R. S. J. Paffenbarger. "Compendium of Physical Activities: Classification of Energy Costs of Human Physical Activities". *Med Sci Sports Exerc*, Vol. 25, No. 1, pp. 71–80, 1993.

[Alle 06] F. R. Allen, E. Ambikairajah, N. H. Lovell, and B. G. Celler. "Classification of a Known Sequence of Motions and Postures from Accelerometry Data using Adapted Gaussian Mixture Models". *Physiol Meas*, Vol. 27, No. 10, pp. 935–951, 2006.

[Alma 17] B. Almaslukh, J. AlMuhtadi, and A. Artoli. "An Effective Deep Autoencoder Approach for Online Smartphone-Based Human Activity Recognition". *IJCSNS*, Vol. 17, No. 4, pp. 160–165, 2017.

[Alsh 15] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan. "Deep Activity Recognition Models with Triaxial Accelerometers". arXiv:1511.04664v1 [cs.LG], 2015.

[Alti 15] M. Altini, J. Penders, R. Vullers, and O. Amft. "Estimating Energy Expenditure Using Body-Worn Accelerometers: A Comparison of Methods, Sensors Number and Positioning". *IEEE J Biomed Health Inform*, Vol. 19, No. 1, pp. 219–226, 2015.

[Altu 10] K. Altun, B. Barshan, and O. Tunçel. "Comparative Study on Classifying Human Activities with Miniature Inertial and Magnetic Sensors". *Pattern Recogn*, Vol. 43, No. 10, pp. 3605–3620, 2010.

[Amft 17] O. Amft and K. Van Laerhoven. "What Will We Wear After Smartphones ?". *IEEE Pervasive Comput*, Vol. 16, No. 4, pp. 80–85, 2017.

[Amin 02] K. Aminian, B. Najafi, C. Büla, P.-F. Leyvraz, and P. Robert. "Spatio-Temporal Parameters of Gait Measured by an Ambulatory System using Miniature Gyroscopes". *J Biomech*, Vol. 35, No. 5, pp. 689–699, 2002.

[Amin 04] K. Aminian and B. Najafi. "Capturing Human Motion Using Body-Fixed Sensors: Outdoor Measurement and Clinical Applications". *Comp Anim Virtual Worlds*, Vol. 15, No. 2, pp. 79–94, 2004.

[Amin 99] K. Aminian, P. Robert, E. E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon. "Physical Activity Monitoring based on Accelerometry: Validation and Comparison with Video Observation". *Med Biol Eng Comput*, Vol. 37, No. 3, pp. 304–308, 1999.

[Angu 13] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. "A Public Domain Dataset for Human Activity Recognition Using Smartphones". In: *Proc ESANN '13*, pp. 437–442, 2013.

[Augu 17] S. Augustus, P. Mundy, and N. Smith. "Support Leg Action Can Contribute to Maximal Instep Soccer Kick Performance: An Intervention Study". *J Sports Sci*, Vol. 35, No. 1, pp. 89–98, 2017.

[Avci 10] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga. "Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey". In: *ARCS '10, Workshop*, pp. 167–176, 2010.

[Baba 04] N. Babaguchi, Y. Kawai, T. Ogura, and T. Kitahashi. "Personalized Abstraction of Broadcasted American Football Video by Highlight Selection". *IEEE Trans Multimedia*, Vol. 6, No. 4, pp. 575–586, 2004.

[Babb 13] R. Babbar, I. Partalas, E. Gaussier, and M.-R. Amini. "On Flat versus Hierarchical Classification in Large-Scale Taxonomies". In: *Proc NIPS '13*, pp. 1824–1832, 2013.

[Bai 16] L. Bai, C. Efstratiou, and C. Siang Ang. "weSport: Utilising Wrist-Band Sensing to Detect Player Activities in Basketball Games". In: *Proc IEEE PerCom Workshops '16*, pp. 1–6, 2016.

[Bano 13] O. Banos, M. Damas, H. Pomares, F. Rojas, B. Delgado-Marquez, and O. Valenzuela. "Human Activity Recognition Based on a Sensor Weighting Hierarchical Classifier". *Soft Comput*, Vol. 17, No. 2, pp. 333–343, 2013.

[Bano 14] O. Banos, M. A. Toth, M. Damas, H. Pomares, and I. Rojas. "Dealing with the Effects of Sensor Displacement in Wearable Activity Recognition". *Sensors*, Vol. 14, No. 6, pp. 9995–10023, 2014.

[Bano 16] O. Banos, B. A. M., W. Ali Khan, M. Afzal, M. Hussain, B. H. Kang, and S. Lee. "The Mining Minds Digital Health and Wellness Framework". *Biomed Eng Online*, Vol. 15 (supplement 1), pp. S165–S186, 2016.

[Bao 04] L. Bao and S. S. Intille. "Activity Recognition from User-Annotated Acceleration Data". In: A. Ferscha and F. Mattern, Eds., *Pervasive Computing*, pp. 1–17, Springer Berlin Heidelberg, 2004.

[Barb 01] N. Barbour and G. Schmidt. "Inertial Sensor Technology Trends". *IEEE Sensors J*, Vol. 1, No. 4, pp. 332–339, 2001.

[Barb 08] J. C. Barbero-Alvarez, V. M. Soto, V. Barbero-Alvarez, and J. Granda-Vera. "Match Analysis and Heart Rate of Futsal Players During Competition". *J Sports Sci*, Vol. 26, No. 1, pp. 63–73, 2008.

[Bare 04] N. C. Barengo, G. Hu, T. A. Lakka, H. Pekkarinen, A. Nissinen, and J. Tuomilehto. "Low Physical Activity as a Predictor for Total and Cardiovascular Disease Mortality in Middle-aged Men and Women in Finland". *Eur Heart J*, Vol. 25, No. 24, pp. 2204–2211, 2004.

[Barf 02] W. R. Barfield, D. T. Kirkendall, and B. Yu. "Kinematic Instep Kicking Differences Between Elite Female and Male Soccer Players". *J Sports Sci Med*, Vol. 1, No. 3, pp. 72–79, 2002.

[Barr 08] S. Barris and C. Button. "A Review of Vision-Based Motion Analysis in Sport". *Sports Med*, Vol. 38, No. 12, pp. 1025–1043, 2008.

[Bart 07] R. Bartlett. *Introduction to Sports Biomechanics: Analyzing Human Movement Patterns*. Routledge, London, UK, 2nd Ed., 2007.

[Baum 12] A. E. Bauman, R. S. Reis, J. F. Sallis, J. C. Wells, R. J. F. Loos, and B. W. Martin. "Correlates of Physical Activity: Why Are Some People Physically Active and Others Not?". *Lancet*, Vol. 380, No. 9838, pp. 258–271, 2012.

[Beng 09] Y. Bengio. "Learning Deep Architectures for AI". *Found Trends Mach Learn*, Vol. 2, No. 1, pp. 1–127, 2009.

[Berc 10] M. Berchtold, M. Budde, D. Gordon, H. R. Schmidtke, and M. Beigl. "ActiServ: Activity Recognition Service for Mobile Phones". In: *Proc ISWC '10*, pp. 1–8, 2010.

[Bhat 80] A. Bhattacharya, E. P. McCutcheon, E. Shvartz, and J. E. Greenleaf. "Body Acceleration Distribution and O2 Uptake in Humans During Running and Jumping". *J Appl Physiol Respir Environ Exerc Physiol*, Vol. 49, No. 5, pp. 881–887, 1980.

[Bish 06] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, NY, 1st Ed., 2006.

[Blan 10] U. Blanke and B. Schiele. "Remember and Transfer What You Have Learned - Recognizing Composite Activities based on Activity Spotting". In: *Proc ISWC '10*, pp. 1–8, 2010.

[Blan 14] P. Blank, P. Kugler, H. Schlarb, and B. M. Eskofier. "A Wearable Sensor System for Sports and Fitness Applications". In: *Book of Abstracts of the 19th Annual Congress of the European College of Sport Science*, p. 703, 2014.

[Blan 15] P. Blank, J. Hoßbach, D. Schuldhaus, and B. M. Eskofier. "Sensor-based Stroke Detection and Stroke Type Classification in Table Tennis". In: *Proc ISWC '15*, pp. 93–100, 2015.

[Blan 86] J. M. Bland and D. G. Altman. "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement". *Lancet*, Vol. 1, No. 8476, pp. 307–310, 1986.

[Blan 99] J. M. Bland and D. G. Altman. "Measuring Agreement in Method Comparison Studies". *Stat Methods Med Res*, Vol. 8, No. 2, pp. 135–160, 1999.

[Blum 98] A. Blum and T. Mitchell. "Combining Labeled and Unlabeled Data with Co-Training". In: *Proc COLT '98*, pp. 92–100, 1998.

[Bono 09a] A. G. Bonomi, A. H. C. Goris, B. Yin, and K. R. Westerterp. "Detection of Type, Duration, and Intensity of Physical Activity Using an Accelerometer". *Med Sci Sports Exerc*, Vol. 41, No. 9, pp. 1770–1777, 2009.

[Bono 09b] A. G. Bonomi, G. Plasqui, A. H. C. Goris, and K. R. Westerterp. "Improving Assessment of Daily Energy Expenditure by Identifying Types of Physical Activity with a Single Accelerometer". *J Appl Physiol*, Vol. 107, No. 3, pp. 655–661, 2009.

[Bout 94] C. V. Bouten, K. R. Westerterp, M. Verduin, and J. D. Janssen. "Assessment of Energy Expenditure for Physical Activity Using a Triaxial Accelerometer". *Med Sci Sport Exer*, Vol. 26, No. 12, pp. 1516–1523, 1994.

[Bout 97a] C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. "A Triaxial Accelerometer and Portable Data Processing Unit for the Assessment of Daily Physical Activity". *IEEE Trans Biomed Eng*, Vol. 44, No. 3, pp. 136–147, 1997.

[Bout 97b] C. V. C. Bouten, A. A. H. J. Sauren, M. Verduin, and J. D. Janssen. "Effects of Placement and Orientation of Body-fixed Accelerometers on the Assessment of Energy Expenditure During Walking". *Med Biol Eng Comput*, Vol. 35, No. 1, pp. 50–56, 1997.

[Boyd 13] L. J. Boyd, K. Ball, and R. J. Aughey. "Quantifying External Load in Australian Football Matches and Training Using Accelerometers". *Int J Sports Physiol Perform*, Vol. 8, No. 1, pp. 44–51, 2013.

[Brei 01] L. Breiman. "Random Forests". *Mach Learn*, Vol. 45, No. 1, pp. 5–32, 2001.

[Brei 84] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and Regression Trees*. Taylor & Francis, London, UK, 1984.

[Brig 11] C. M. N. Brigante, N. Abbate, A. Basile, A. Carmelo Faulisi, and S. Sessa. "Towards Miniaturization of a MEMS-Based Wearable Motion Capture System". *IEEE Trans. Ind. Electron.*, Vol. 58, No. 8, pp. 3234–3241, 2011.

[Brod 10] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. "The Balanced Accuracy and its Posterior Distribution". In: *Proc ICPR '10*, pp. 3121–3124, 2010.

[Brop 07] R. H. Brophy, S. I. Backus, B. S. Pansy, and R. J. Lyman, S. Williams. "Lower Extremity Muscle Activation and Alignment During the Soccer Instep and Side-foot Kicks". *J Orthop Sports Phys Ther*, Vol. 37, No. 5, pp. 260–268, 2007.

[Bull 09] F. C. Bull, T. S. Maslin, and T. Armstrong. "Global Physical Activity Questionnaire (GPAQ): Nine Country Reliability and Validity Study". *J Phys Act Health*, Vol. 6, No. 6, pp. 790–804, 2009.

[Bull 14] A. Bulling, U. Blanke, and B. Schiele. "A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors". *ACM Comput Serv*, Vol. 46, No. 3, pp. 33:1–33:33, 2014.

[Bull 99] T. Bull Andersen, H. C. Dörge, and F. I. Thomsen. "Collisions in Soccer Kicking". *Sports Eng*, Vol. 2, No. 2, pp. 121–125, 1999.

[Burg 98] C. J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition". *Data Min Knowl Discov*, Vol. 2, No. 2, pp. 121–167, 1998.

[Burn 10] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca. "SHIMMER^{TM} - A Wireless Sensor Platform for Noninvasive Biomedical Research". *IEEE Sensors J*, Vol. 10, No. 9, pp. 1527–1534, 2010.

[Cala 11] A. Calatroni, D. Roggen, and G. Tröster. "Automatic Transfer of Activity Recognition Capabilities between Body-Worn Motion Sensors: Training Newcomers to Recognize Locomtion". In: *Proc INSS '11*, pp. 1–8, 2011.

[Cao 09] H. Cao, V. Leung, C. Chow, and H. Chan. "Enabling Technologies for Wireless Body Area Networks: A Survey and Outlook". *IEEE Commun Mag*, Vol. 47, No. 12, pp. 84–93, 2009.

[Capp 82] A. Cappozzo. "Low Frequency Self-Generated Vibration During Ambulation in Normal Men". *J Biomech*, Vol. 15, No. 8, pp. 599–609, 1982.

[Carl 05] C. Carling, A. M. Williams, and T. Reilly. *Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance*. Routledge, London, UK, 1st Ed., 2005.

[Carl 08] C. Carling, J. Bloomfield, L. Nelsen, and T. Reilly. "The Role of Motion Analysis in Elite Soccer: Contemporary Performance Measurement Techniques and Work Rate Data". *Sports Med*, Vol. 38, No. 10, pp. 839–862, 2008.

[Casp 85] C. J. Caspersen, K. E. Powell, and C. G. M. "Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research.". *Public Health Rep*, Vol. 100, No. 2, pp. 126–131, 1985.

[Cast 14] J. Castellano, D. Alvarez-Pastor, and P. S. Bradley. "Evaluation of Research Using Computerised Tracking Systems (Amisco® and Prozone®) to Analyze Physical Performance in Elite Soccer: A Systematic Review". *Sports Med*, Vol. 44, No. 5, pp. 701–712, 2014.

[Cata 13] F. O. Catak and M. E. Balaban. "CloudSVM: Training an SVM Classifier in Cloud Computing Systems". In: Q. Zu, B. Hu, and A. Elci, Eds., *Pervasive Computing and the Networked World - Joint International Conference, ICPCA/SWS 2012, Istanbul, Turkey, November 28-30, 2012, Revised Selected Papers*, pp. 57–68, Springer Berlin Heidelberg, 2013.

[Caus 16] M. Causo, S. Benatti, A. Frappé, A. Cathelin, E. Farella, A. Kaiser, L. Benini, and J. M. Rabaey. "Sampling Modulation: An Energy Efficient Novel Feature Extraction for Biosignal Processing". In: *Proc BioCas '16*, pp. 348–351, 2016.

[Chan 12] M. Chan, D. Esteve, J.-Y. Fourniols, C. Escriba, and E. Campo. "Smart Wearable Systems: Current Status and Future Challenges". *Artif Intell Med*, Vol. 56, No. 3, pp. 137–156, 2012.

[Chap 10] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, Cambridge, MA, 1st Ed., 2010.

[Chav 13] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millan, and D. Roggen. "The Opportunity Challenge: A Benchmark Database for On-Body Sensor-based Activity Recognition". *Pattern Recogn Lett*, Vol. 34, No. 15, pp. 2033–2042, 2013.

[Chaw 02] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. "SMOTE: Synthetic Minority Over-sampling Technique". *J Artif Intell Res*, Vol. 16, No. 1, pp. 321–357, 2002.

[Chen 05] K. Y. Chen and D. R. J. Bassett. "The Technology of Accelerometry-based Activity Monitors: Current and Future". *Med Sci Sports Exerc*, Vol. 37 (supplement 9), pp. S490–S500, 2005.

[Chen 12a] H. Chen and R. H. L. Chiang. "Business Intelligence and Analytics: From Big Data to Big Impact". *MIS Quart*, Vol. 36, No. 4, pp. 1165–1188, 2012.

[Chen 12b] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu. "Sensor-Based Activity Recognition". *IEEE Trans Syst, Man, Cybern C, Appl Rev*, Vol. 42, No. 6, pp. 790–808, 2012.

[Chen 13a] C. Chen, Z. Liu, W.-H. Lin, S. Li, and K. Wang. "Distributed Modeling in a MapReduce Framework for Data-Driven Traffic Flow Forecasting". *IEEE Trans Intell Transp Syst*, Vol. 14, No. 1, pp. 22–33, 2013.

[Chen 13b] S. Chen, J. Lach, O. Amft, M. Altini, and J. Penders. "Unsupervised Activity Clustering to Estimate Energy Expenditure with a Single Body Sensor". In: *Proc BSN '13*, pp. 1–6, 2013.

[Chen 14a] M. Chen, S. Mao, and Y. Liu. "Big Data: A Survey". *Mobile Netw Appl*, Vol. 19, No. 2, pp. 171–209, 2014.

[Chen 14b] M. Chen, S. Mao, Y. Zhang, and V. C. M. Leung. "Big Data - Related Technologies, Challenges and Future Prospects". Springer International Publishing, Cham, Switzerland, 2014.

[Chen 15] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong. "Data Mining for the Internet of Things: Literature Review and Challenges". *Int J Distrib Sens N*, Vol. 2015, pp. 12:1–12:14, 2015.

[Chen 16] Z. Chen and B. Liu. *Lifelong Machine Learning - Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool, Williston, VT, 1st Ed., 2016.

[Chen 97] K. Y. Chen and M. Sun. "Improving Energy Expenditure Estimation by Using a Trixial Accelerometer". *J Appl Physiol*, Vol. 83, No. 6, pp. 2112–2122, 1997.

[Chow 18] A. Chowdhury, D. Tjondronegoro, V. Chandran, and S. Trost. "Physical Activity Recognition using Posterior-Adapted Class-Based Fusion of Multi-Accelerometers Data". *IEEE J Biomed Health Inform*, Vol. 22, No. 3, pp. 678–685, 2018.

[Chu 06] C.-T. Chu, S. K. Kim, Y.-A. Lin, Y. Y. Yu, G. Bradski, A. Y. Ng, and K. Olukotun. "Map-Reduce for Machine Learning on Multicore". In: *Proc NIPS '06*, pp. 281–288, 2006.

[Clel 13] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, and D. Finlay. "Optimal Placement of Accelerometer for the Detection of Everyday Activities". *Sensors*, Vol. 13, No. 7, pp. 9183–9200, 2013.

[Cohn 94] D. Cohn, L. Atlas, and R. Ladner. "Improving Generalization with Active Learning". *Mach Learn*, Vol. 15, No. 2, pp. 201–221, 1994.

[Conn 11] D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, M. Walsh, and C. O'Mathuna. "Multi-Sensor Classification of Tennis Strokes". In: *Proc IEEE Sensors '11*, pp. 1437–1440, 2011.

[Cons 08] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay. "Activity Sensing in the Wild: A Field Trial of UbiFit Garden". In: *Proc CHI '08*, pp. 1797–1806, 2008.

[Cook 13] D. Cook, K. D. Feuz, and N. C. Krishnan. "Transfer Learning for Activity Recognition: A Survey". *Knowl Inf Syst*, Vol. 36, No. 3, pp. 537–556, 2013.

[Corm 14] S. J. Cormack, R. L. Smith, M. M. Mooney, W. B. Young, and B. J. O'Brien. "Accelerometer Load as a Measure of Activity Profile in Different Standards of Netball Match Play". *Int J Sports Physiol Perform*, Vol. 9, No. 2, pp. 283–291, 2014.

[Cort 95] C. Cortes and V. Vapnik. "Support-Vector Networks". *Mach Learn*, Vol. 20, No. 3, pp. 273–297, 1995.

[Cove 67] T. M. Cover and P. E. Hart. "Nearest Neighbor Pattern Classification". *IEEE Trans. Inf. Theory*, Vol. 13, No. 1, pp. 21–27, 1967.

[Crai 03] C. L. Craig, A. L. Marshall, M. Sjöström, A. E. Bauman, M. L. Booth, B. E. Ainsworth, M. Pratt, U. Ekelund, A. Yngve, J. F. Sallis, and P. Oja. "International Physical Activity Questionnaire: 12-Country Reliability and Validity". *Med Sci Sports Exerc*, Vol. 35, No. 8, pp. 1381–1395, 2003.

[Crou 06a] S. E. Crouter, J. R. Churilla, and B. J. D. R. "Estimating Energy Expenditure Using Accelerometers". *Eur J Appl Physiol*, Vol. 98, No. 6, pp. 601–612, 2006.

[Crou 06b] S. E. Crouter, K. G. Clowers, and B. J. D. R. "A Novel Method for Using Accelerometer Data to Predict Energy Expenditure". *J Appl Physiol*, Vol. 100, No. 4, pp. 1324–1331, 2006.

[Crou 10] S. E. Crouter, E. Kuffel, J. D. Haas, E. A. Frongillo, and B. J. D. R. "Refined Two-Regression Model for the ActiGraph Accelerometer". *Med Sci Sports Exerc*, Vol. 42, No. 5, pp. 1029–1037, 2010.

[Dai 14] W. Dai and W. Ji. "A MapReduce Implementation of C4.5 Decision Tree Algorithm". *IJDTA*, Vol. 7, No. 1, pp. 49–60, 2014.

[De S 12] L. C. De Silva, C. Morikawa, and I. M. Petra. "State of the Art of Smart Homes". *Eng Appl Artif Intell*, Vol. 25, No. 7, pp. 1313–1321, 2012.

[Dean 08] J. Dean and S. Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters". *Commun ACM*, Vol. 51, No. 1, pp. 107–113, 2008.

[Deng 14] L. Deng and D. Yu. "Deep Learning". *Foundations and Trends in Signal Processing*, Vol. 7, No. 3-4, pp. 197–387, 2014.

[Devo 12] J. L. Devore and K. N. Berk. *Modern Mathematical Statistics with Applications*. Springer Science+Business Media, LLC, New York, NY, 2nd Ed., 2012.

[Di S 06] V. Di Salvo, A. Collins, B. McNeill, and M. Cardinale. "Validation of Prozone ®: A New Video-Based Performance Analysis System". *Int J Perform Anal Sport*, Vol. 6, No. 1, pp. 108–119, 2006.

[Diet 02] T. G. Dietterich. "Machine Learning for Sequential Data: A Review". In: T. Caelli, A. Amin, R. P. W. Duin, D. de Ridder, and M. Kamel, Eds., *Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshops SSPR 2002 and SPR 2002 Windsor, Ontario, Canada, August 6–9, 2002 Proceedings*, pp. 15–30, Springer Berlin Heidelberg, 2002.

[Dobk 11] B. H. Dobkin and A. Dorsch. "The Promise of mHealth: Daily Activity Monitoring and Outcome Assessment by Wearable Sensors". *Neurorehabil Neural Repair*, Vol. 25, No. 9, pp. 788–798, 2011.

[DOra 10] T. D'Orazio and M. Leo. "A Review of Vision-Based Systems for Soccer Video Analyis". *Pattern Recogn*, Vol. 43, No. 8, pp. 2911–2926, 2010.

[Dorn 13] S. Dorn. *Implementation and Evaluation of Energy Expenditure Estimation Algorithms for Multiple Sclerosis Patients*. Bachelor's thesis, Digital Sports Group, Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2013.

[Dors 14] E. Dorschky. *Sensor-Based Event Classification in Sports Applications*. Master's thesis, Digital Sports Group, Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2014.

[Dors 15] E. Dorschky, D. Schuldhaus, H. Koerger, and B. M. Eskofier. "A Framework for Early Event Detection for Wearable Systems". In: *Proc ISWC '15*, pp. 109–112, 2015.

[Duda 00] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2nd Ed., 2000.

[Duff 10] R. Duffield, M. Reid, J. Baker, and W. Spratford. "Accuracy and Reliability of GPS Devices for Measurement of Movement Patterns in Confined Spaces for Court-Based Sports". *J Sci Med Sport*, Vol. 13, No. 5, pp. 523–525, 2010.

[Duga 05] S. A. Dugan and K. P. Bhat. "Biomechanics and Analysis of Running Gait". *Phys Med Rehabil Clin N Am*, Vol. 16, No. 3, pp. 603–621, 2005.

[Dör 02] H. C. Dörge, T. Bull Andersen, H. Sørensen, and E. B. Simonsen. "Biomechanical Differences in Soccer Kicking with the Preferred and the Non-preferred Leg". *J Sports Sci*, Vol. 20, No. 4, pp. 293–299, 2002.

[Dör 99] H. C. Dörge, T. Bull Andersen, H. Sørensen, E. B. Simonsen, H. Aagaard, P. Dyhre-Poulsen, and K. Klausen. "EMG Activity of the Iliopsoas Muscle and Leg Kinetics during the Soccer Place Kick". *Scand J Med Sci Sports*, Vol. 9, No. 4, pp. 195––200, 1999.

[Eich 15] J. C. Eichstaedt, H. A. Schwartz, M. L. Kern, G. Park, D. R. Labarthe, R. M. Merchant, S. Jha, M. Agrawal, L. A. Dziurzynski, M. Sap, C. Weeg, L. H. Larson, E. E. Ungar, and M. E. P. Seligman. "Psychological Language on Twitter Predicts County-Level Heart Disease Mortality". *Psychol Sci*, Vol. 26, No. 2, pp. 159–169, 2015.

[Ekin 03] A. Ekin, A. M. Tekalp., and R. Mehrotra. "Automatic Soccer Video Analysis and Summarization". *IEEE Trans Image Process*, Vol. 12, No. 7, pp. 796–807, 2003.

[Elli 14] K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall. "A Random Forest Classifier for the Prediction of Energy Expenditure and Type of Physical Activity from Wrist and Hip Accelerometers". *Physiol Meas*, Vol. 35, No. 11, pp. 2191–2203, 2014.

[Erme 08] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and K. I. "Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions". *IEEE Trans Inf Technol Biomed*, Vol. 12, No. 1, pp. 20–26, 2008.

[Fahi 14] M. Fahim, M. Idris, R. Ali, C. Nugent, B. Khang, E.-N. Huh, and S. Lee. "ATHENA: A Personalized Platform to Promote an Active Lifestyle and Wellbeing Based on Physical, Mental and Social Health Primitives". *Sensors*, Vol. 14, No. 5, pp. 9313–9329, 2014.

[Fahr 97] J. Fahrenberg, F. Foerster, M. Smeja, and W. Müller. "Assessment of Posture and Motion by Multichannel Piezoresistive Accelerometer Recordings". *Psychophysiology*, Vol. 34, No. 5, pp. 607–612, 1997.

[Fati 13] I. Fatima, M. Fahim, Y.-K. Lee, and S. Lee. "A Unified Framework for Activity Recognition-Based Behavior Analysis and Action Prediction in Smart Homes". *Sensors*, Vol. 13, No. 2, pp. 2682–2699, 2013.

[Fawc 06] T. Fawcett. "An Introduction to ROC Analysis". *Pattern Recogn Lett*, Vol. 27, No. 8, pp. 861–874, 2006.

[Feng 14] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos. "TRAC: Truthful Auction for Location-Aware Collaborative Sensing in Mobile Crowdsourcing". In: *Proc IEEE INFOCOM '14*, pp. 1231–1239, 2014.

[Ferr 95] F. Ferraris, U. Grimaldi, and M. Parvis. "Procedure for Effortless In-Field Calibration of Three-Axis Rate Gyros and Accelerometers". *Sensor Mater*, Vol. 7, No. 5, pp. 311–330, 1995.

[Fitz 15] E. C. Fitzhugh. "Methods to Measure Physical Activity Behaviors in Health Education Research". *AJHE*, Vol. 46, No. 1, pp. 1–6, 2015.

[Fleu 10] A. Fleury, M. Vacher, and N. Noury. "SVM-Based Multi-Modal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms and First Experimental Results". *IEEE Trans Inf Technol Biomed*, Vol. 14, No. 2, pp. 274–283, 2010.

[Foer 99] F. Foerster, M. Smeja, and J. Fahrenberg. "Detection of Posture and Motion by Accelerometry: A Validation Study in Ambulatory Monitoring". *Comp Hum Behav*, Vol. 15, No. 5, pp. 571–583, 1999.

[Fran 91] I. M. Franks and G. Miller. "Training Coaches to Observe and Remember". *J Sports Sci*, Vol. 9, No. 3, pp. 285–297, 1991.

[Fren 10] W. G. P. Frencken, K. A. P. M. Lemmink, and N. J. Delleman. "Soccer-Specific Accuracy and Validity of the Local Position Measurement (LPM) System". *J Sci Med Sport*, Vol. 13, No. 6, pp. 641–645, 2010.

[Frig 89] M. Frigge, D. C. Hoaglin, and B. Iglewicz. "Some Implementations of the Boxplots". *Am Stat*, Vol. 43, No. 1, pp. 50–54, 1989.

[Full 15] A. M. Fullenkamp, B. M. Campbell, M. Laurent, and A. P. Lane. "The Contribution of Trunk Axial Kinematics to Poststrike Ball Velocity During Maximal Instep Soccer Kicking". *J Appl Biomech*, Vol. 31, No. 5, pp. 370–376, 2015.

[Full 17] E. Fullerton, B. Heller, and M. Munoz-Organero. "Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers". *IEEE Sensors J*, Vol. 17, No. 16, pp. 5290–5297, 2017.

[Fult 09] S. K. Fulton, D. B. Pyne, and B. Burkett. "Validity and Reliability of Kick Count and Rate in Freestyle Using Inertial Sensor Technology". *J Sports Sci*, Vol. 27, No. 10, pp. 1051–1058, 2009.

[Gabb 15] T. J. Gabbett. "Relationship Between Accelerometer Load, Collisions, and Repeated High-Intensity Effort Activity in Rugby League Players". *J Strength Cond Res*, Vol. 29, No. 12, pp. 3424–3431, 2015.

[Gand 15] A. Gandomi and M. Haider. "Beyond the Hype: Big Data Concepts, Methods, and Analytics". *IJIM*, Vol. 35, No. 2, pp. 137–144, 2015.

[Gant 11] J. Gantz and D. Reinsel. "Extracting Value from Chaos". *IDC IView*, No. 1142, pp. 1–12, 2011.

[Garc 06] J. Garcia-Aymerich, P. Lange, M. Benet, P. Schnohr, and J. M. Antó. "Regular Physical Activity Reduces Hospital Admission and Mortality in Chronic Obstructive Pulmonary Disease: A Population Based Cohort Study". *Thorax*, Vol. 61, No. 9, pp. 772–778, 2006.

[Gemp 98] F. Gemperle, C. Kasabach, J. Stivoric, M. Bauer, and R. Martin. "Design for Wearability". In: *Proc ISWC '98*, pp. 116–122, 1998.

[Geor 11] L. George. *HBase: The Definitive Guide*. O'Reilly Media, Inc., Sebastopol, CA, 2011.

[Godf 08] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin. "Direct Measurement of Human Movement by Accelerometry". *Med Eng Phys*, Vol. 30, No. 10, pp. 1364–1386, 2008.

[Good 14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative Adversarial Nets". In: *NIPS '14*, pp. 2672–2680, Curran Associates, Inc., 2014.

[Gord 11] D. Gordon, J.-H. Hanne, M. Berchtold, T. Miyaki, and M. Beigl. "Recognizing Group Activities using Wearable Sensors". In: A. Puiatti and T. Gu, Eds., *Mobile*, pp. 350–361, Springer Berlin Heidelberg, 2011.

[Groh 14] B. H. Groh, N. Weeger, F. Warschun, and B. M. Eskofier. "Simplified Orientation Determination in Ski Jumping using Inertial Sensor Data". In: *Proc DGON ISS '14*, pp. 1–11, 2014.

[Groh 16] B. H. Groh, M. Fleckensten, and B. M. Eskofier. "Wearable Trick Classification in Freestyle Snowboarding". In: *Proc BSN '16*, pp. 89–93, 2016.

[Grü 11] T. von der Grün, N. Franke, D. Wolf, N. Witt, and A. Eidloth. "A Real-Time Tracking System for Football Match and Training Analysis". In: A. Heuberger, G. Elst, and R. Hanke, Eds., *Microelectronic Systems: Circuits, Systems and Applications*, pp. 199–212, Springer Berlin Heidelberg, 2011.

[Gupt 13] B. C. Gupta and I. Guttman. *Statistics and Probability with Applications for Engineers and Scientists*. John Wiley & Sons, Inc., Hoboken, NJ, 1st Ed., 2013.

[Guyo 03] I. Guyon and A. Elisseeff. "An Introduction to Variable and Feature Selection". *JMLR*, Vol. 3, pp. 1157–1182, 2003.

[Habe 13] P. Haber. *Lungenfunktion und Spiroergometrie: Interpretation und Befunderstellung unter Einschluss der arteriellen Blutgasanalyse*. Springer, Wien, Austria, 3rd Ed., 2013.

[Hach 12] H. Hachiya, M. Sugiyama, and N. Ueda. "Importance-Weighted Least-Squares Probabilistic Classifier for Covariate Shift Adaptation with Application to Human Activity Recognition". *Neurocomputing*, Vol. 80, pp. 93–101, 2012.

[Hall 00] M. A. Hall. "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning". In: *Proc ICML '00*, pp. 359–366, 2000.

[Hall 12] P. C. Hallal, L. B. Andersen, F. C. Bull, R. Guthold, W. Haskell, and U. Ekelund. "Global Physical Activity Levels: Surveillance Progress, Pitfalls, and Prospects". *Lancet*, Vol. 380, No. 9838, pp. 247–257, 2012.

[Hall 97] D. Hall and J. Llinas. "An Introduction to Multisensor Data Fusion". *Proc IEEE*, Vol. 85, No. 1, pp. 6–23, 1997.

[Hall 99a] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1999.

[Hall 99b] M. A. Hall and L. A. Smith. "Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper". In: *Proc FLAIRS '99*, pp. 235–239, 1999.

[Han 12] M. Han, L. T. Vinh, Y.-K. Lee, and S. Lee. "Comprehensive Context Recognizer Based on Multimodal Sensors in a Smartphone". *Sensors*, Vol. 12, No. 9, pp. 12588–12605, 2012.

[Hard 15] M. Hardegger, B. Ledergerber, S. Mutter, C. Vogt, J. Seiter, A. Calatroni, and G. Tröster. "Sensor Technology for Ice Hockey and Skating". In: *Proc IEEE BSN '15*, pp. 1–6, 2015.

[Hast 09] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Springer, New York, NY, 2nd Ed., 2009.

[Hatt 10] Y. Hattori, S. Inoue, T. Masaki, G. Hirakawa, and O. Sudo. "Gathering Large Scale Human Activity Information Using Mobile Sensor Devices". In: *Proc BWCCA '10*, pp. 708–713, 2010.

[He 09] H. He and E. A. Garcia. "Learning from Imbalanced Data". *IEEE Trans Knowl Data Eng*, Vol. 21, No. 9, pp. 1263–1284, 2009.

[Hend 00] D. Hendelman, K. Miller, C. Baggett, E. Debold, and P. Freedson. "Validity of Accelerometry for the Assessment of Moderate Intensity Physical Activity in the Field". *Med Sci Sports Exerc*, Vol. 32 (supplement 9), pp. S442–S449, 2000.

[High 12] D. G. Higham, D. B. Pyne, J. M. Anson, and A. Eddy. "Movement Patterns in Rugby Sevens: Effects of Tournament Level, Fatigue and Substitute Players". *J Sci Med Sport*, Vol. 15, No. 3, pp. 277–282, 2012.

[Hill 14] A. P. Hills, N. Mokhtar, and N. M. Byrne. "Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures". *Front Nutr*, Vol. 1, No. 5, pp. 1–16, 2014.

[Hint 06] G. E. Hinton and R. R. Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks". *Science*, Vol. 313, No. 5786, pp. 504–507, 2006.

[Ho 94] T. K. Ho, J. J. Hull, and S. N. Srihari. "Decision Combination in Multiple Classifier Systems". *IEEE Trans Pattern Anal Mach Intell*, Vol. 16, No. 1, pp. 66–75, 1994.

[Hoss 17] H. M. S. Hossain, M. A. A. H. Khan, and N. Roy. “SoccerMate: A Personal Soccer Attribute Profiler using Wearables”. In: *Proc IEEE PerCom Workshops ’17*, pp. 164–169, 2017.

[Howe] J. Howe. “Crowdsourcing - Why the Power of the Crowd is Driving the Future of Business”. Accessed on October 15th 2018, http://www.crowdsourcing.com/.

[Hu 01] F. B. Hu, M. F. Leitzmann, M. J. Stampfer, G. A. Colditz, W. C. Willett, and E. B. Rimm. “Physical Activity and Television Watching in Relation to Risk for Type 2 Diabetes Mellitus in Men”. *Arch Intern Med*, Vol. 161, No. 12, pp. 1542–1548, 2001.

[Hu 08] D. H. Hu and Q. Yang. “CIGAR: Concurrent and Interleaving Goal and Activity Recognition”. In: *Proc AAAI ’08*, pp. 1363–1368, 2008.

[Hu 09] D. H. Hu, X.-X. Zhang, J. Yin, V. W. Zheng, and Q. Yang. “Abnormal Activity Recognition Based on HDP-HMM Models”. In: *Proc IJCAI ’09*, pp. 1715–1720, 2009.

[Hu 14] H. Hu, Y. Wen, T.-S. Chua, and X. Li. “Toward Scalable Systems for Big Data Analytics: A Technology Tutorial”. *IEEE Access*, Vol. 2, pp. 652–687, 2014.

[Hugh 04] M. Hughes and I. M. Franks. *Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport*. Routledge, London, UK, 2nd Ed., 2004.

[Hugh 15] M. Hughes and I. M. Franks. *Essentials of Performance Analysis in Sport*. Routledge, London, UK, 2nd Ed., 2015.

[Hynd 06] R. J. Hyndman and A. B. Koehler. “Another Look at Measures of Forecast Accuracy”. *Int J Forecast*, Vol. 22, No. 4, pp. 679–688, 2006.

[Inou 14] K. Inoue, H. Nunome, T. Sterzing, H. Shinkai, and Y. Ikegami. “Dynamics of the Support Leg in Soccer Instep Kicking”. *J Sports Sci*, Vol. 32, No. 11, pp. 1023–1032, 2014.

[Inti 06] S. S. Intille, K. Larson, E. M. Tapia, J. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson. “Using a Live-in Laboratory for Ubiquitous Computing Research”. In: *Proc PERVASIVE ’06*, pp. 349–365, 2006.

[Ishi 12] H. Ishii, T. Yanagiya, H. Naito, S. Katamoto, and T. Maruyama. "Theoretical Study of Factors Affecting Ball Velocity in Instep Soccer Kicking". *J Appl Biomech*, Vol. 28, No. 3, pp. 258–270, 2012.

[Jahn 17] A. Jahn, M. Bachmann, P. Wenzel, and K. David. "Focus on the User: A User Relative Coordinate System for Activity Detection". In: *Proc CONTEXT '17*, pp. 582–595, 2017.

[Jako 16] C. Jakob. *Sensor Fusion for High-Intensity Event Analyis in Sports Under Realistic Conditions*. Master's thesis, Digital Sports Group, Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2016.

[Jame 13] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning - with Applications in R*. Springer, New York, 1st Ed., 2013.

[Jens 13] U. Jensen, F. Prade, and B. M. Eskofier. "Classification of Kinematic Swimming Data with Emphasis on Resource Consumption". In: *Proc IEEE BSN '13*, pp. 1–5, 2013.

[Jens 16] U. Jensen, P. Blank, P. Kugler, and B. M. Eskofier. "Unobtrusive and Energy-Efficient Swimming Exercise Tracking Using On-Node Processing". *IEEE Sensors J*, Vol. 16, No. 10, pp. 3972–3980, 2016.

[Jett 90] M. Jetté, K. Sidney, and G. Blümchen. "Metabolic Equivalents (METS) in Exercise Testing, Exercise Prescription, and Evaluation of Functional Capacity". *Clin Cardiol*, Vol. 13, No. 8, pp. 555–565, 1990.

[Jin 14] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami. "An Information Framework for Creating a Smart City Through Internet of Things". *IEEE Internet Things J*, Vol. 1, No. 2, pp. 112–121, 2014.

[John 95] G. H. John and P. Langley. "Estimating Continuous Distributions in Bayesian Classifiers". In: *Proc UAI '95*, pp. 338–345, 1995.

[Jone 13] T. B. Jones and N. G. Nenadic. *Electromechanics and MEMS*. Cambridge University Press, Cambridge, UK, 1st Ed., 2013.

[Jone 14] N. Jones. "The Learning Machines". *Nature*, Vol. 505, No. 7482, pp. 146–148, 2014.

[Kahn 02] E. B. Kahn, L. T. Ramsey, R. C. Brownson, G. W. Heath, E. H. Howze, K. E. Powell, E. J. Stone, M. W. Rajab, and P. Corso. "The Effectiveness of Interventions to Increase Physical Activity. A Systematic Review". *Am J Prev Med*, Vol. 22 (supplement 4), pp. S73–S107, 2002.

[Kara 06] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. "Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring". *IEEE Trans Inf Technol Biomed*, Vol. 10, No. 1, pp. 156–167, 2006.

[Kast 10] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse. "Activity Recognition using Semi-Markov Models on Real World Smart Home Datasets". *J Ambient Intell Smart Environ*, Vol. 2, No. 3, pp. 311–325, 2010.

[Kaut 15a] T. Kautz and B. M. Eskofier. "A Robust Kalman Framework with Resampling and Optimal Smoothing". *Sensors*, Vol. 15, No. 3, pp. 4975–4995, 2015.

[Kaut 15b] T. Kautz, B. H. Groh, and B. M. Eskofier. "Sensor Fusion for Multi-Player Activity Recognition in Game Sports". In: *KDD Workshop on Large-Scale Sports Analytics '15*, pp. 1–4, 2015.

[Kaut 17] T. Kautz, B. M. Eskofier, and F. Pasluosta. "Generic Performance Measure for Multiclass-Classifiers". *Pattern Recogn*, Vol. 68, pp. 111–125, 2017.

[Ke 13] S.-R. Ke, H. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi. "A Review on Video-Based Human Activity Recognition". *Computers*, Vol. 2, No. 2, pp. 88–131, 2013.

[Kell 04] E. Kellis, A. Katis, and I. Gissis. "Knee Biomechanics of the Support Leg in Soccer Kicks from Three Angles of Approach". *Med Sci Sports Exerc*, Vol. 36, No. 6, pp. 1017–1028, 2004.

[Kell 07] E. Kellis and A. Katis. "Biomechanical Characteristics and Determinants of Instep Soccer Kick". *J Sports Sci Med*, Vol. 6, No. 2, pp. 154––165, 2007.

[Kell 12] D. Kelly, G. F. Coughlan, B. S. Green, and B. Caulfield. "Automatic Detection of Collisions in Elite Level Rugy Union Using a Wearable Sensing Device". *Sports Eng*, Vol. 15, No. 2, pp. 81–92, 2012.

[Kemp 11] V. Kempe. *Inertial MEMS: Principles and Practice*. Cambridge University Press, Cambridge, UK, 1st Ed., 2011.

[Kenn 11] W. L. Kenney, J. Wilmore, and D. Costill. *Physiology of Sport and Exercise*. Human Kinetics, Champaign, IL, 5th Ed., 2011.

[Khan 16] A. Khan, N. Hammerla, S. Mellor, and T. Plötz. "Optimising Sampling Rates for Accelerometer-Based Human Activity Recognition". *Pattern Recogn Lett*, Vol. 73, pp. 33–40, 2016.

[Kirk 16] R. F. Kirk, C. Zwick, D. Schuldhaus, B. M. Eskofier, H. Koerger, and M. P. Stöve. "Method and Device for Associating Frames in a Video of an Activity of a Person with an Event". US Patent Application US 20160314818 A1, adidas AG, October 27, 2016.

[Kitt 98] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. "On Combining Classifiers". *IEEE Trans Pattern Anal Mach Intell*, Vol. 20, No. 3, pp. 273–297, 1998.

[Knee 16] P. Knees and M. Schedl. "Music Similarity and Retrieval: An Introduction to Audio-and Web-based Strategies". Springer-Verlag Berling Heidelberg, 2016.

[Kohl 12] H. W. Kohl 3rd, C. L. Craig, E. V. Lambert, S. Inoue, J. R. Alkandari, G. Leetongin, and S. Kahlmeier. "The Pandemic of Physical Inactivity: Global Action for Public Health". *Lancet*, Vol. 380, No. 9838, pp. 294–305, 2012.

[Kunz 07] M. Kunz. "Big Count: 265 Million Playing Football". In: *FIFA Magazine*, pp. 10–15, 2007.

[Kuo 06] S. M. Kuo, B. H. Lee, and W. Tian. *Real-Time Digital Signal Processing: Implementations and Applications*. John Wiley & Sons, Ltd, West Sussex, England, 2nd Ed., 2006.

[Kwap 10] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. "Activity Recognition using Cell Phone Accelerometers". *ACM SIGKDD Explorations Newsletter*, Vol. 12, No. 2, pp. 74–82, 2010.

[Lach 68] P. A. Lachenbruch and M. R. Mickey. "Estimation of Error Rates in Discriminant Analysis". *Technometrics*, Vol. 10, No. 1, pp. 1–11, 1968.

[Lamo 01] M. J. Lamonte and B. E. Ainsworth. "Quantifying Energy Expenditure and Physical Activity in the Context of Dose Response". *Med Sci Sport Exer*, Vol. 33 (supplement 6), pp. S370–S378, 2001.

[Lane 01] D. Laney. "3D Data Management: Controlling Data Volume, Velocity, and Varity". Tech. Rep., META Group Research Note, February, 6, 2001.

[Lara 13] O. D. Lara and M. A. Labrador. "A Survey on Human Activity Recognition using Wearable Sensors". *IEEE Commun Surv Tutor*, Vol. 15, No. 3, pp. 1192–1209, 2013.

[Lase 13] W. S. Lasecki, Y. C. Song, H. Kautz, and J. P. Bigham. "Real-Time Crowd Labeling for Deployable Activity Recognition". In: *Proc CSCW '13*, pp. 1203–1212, 2013.

[LeCu 98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-Based Learning Applied to Document Recognition". *Proc IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[Lee 14] Y.-S. Lee and S.-B. Cho. "Activity Recognition with Android Phone using Mixture-of-Experts Co-trained with Labeled and Unlabeled Data". *Neurocomputing*, Vol. 126, pp. 106–115, 2014.

[Leen 03] N. Y. Leenders, T. E. Nelson, and W. M. Sherman. "Ability of Different Physical Activity Monitors to Detect Movement During Treadmill Walking". *Int J Sports Med*, Vol. 24, No. 1, pp. 43–50, 2003.

[Lees 10] A. Lees, T. Asai, T. B. Andersen, H. Nunome, and T. Sterzing. "The Biomechanics of Kicking in Soccer: A Review". *J Sports Sci*, Vol. 28, No. 8, pp. 805–817, 2010.

[Lees 98] A. Lees and L. Nolan. "The Biomechanics of Soccer: A Review". *J Sports Sci*, Vol. 16, No. 3, pp. 211–234, 1998.

[Lenz 02] M. Lenzerini. "Data Integration: A Theoretical Perspective". In: *Proc ACM PODS '02*, pp. 233–246, 2002.

[Lesk 14] J. Leskovec, A. Rajaraman, and J. D. Ullman. *Mining of Massive Datasets*. Cambridge University Press, Cambridge, UK, 2nd Ed., 2014.

[Leva 98] J. Levanon and J. Dapena. “Comparison of the Kinematics of the Full-Instep and Pass Kicks in Soccer”. *Med Sci Sports Exerc*, Vol. 30, No. 6, pp. 917-–927, 1998.

[Lido 05] R. Lidor, B. Falk, M. Arnon, C. Y., G. Segal, and Y. Lander. “Measurement of Talent in Team Handball: the Questionable Use of Motor and Physical Tests”. *J Strength Cond Res*, Vol. 19, No. 2, pp. 318–325, 2005.

[Liu 12] S. Liu, R. X. Gao, D. John, J. W. Staudenmayer, and P. S. Freedson. “Multisensor Data Fusion for Physical Activity Assessment”. *IEEE Trans Biomed Eng*, Vol. 59, No. 3, pp. 687–696, 2012.

[Liu 13] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen. “Scalable Sentiment Classification for Big Data Analysis Using Naive Bayes Classifier”. In: *Proc IEEE BigData ’13*, pp. 99–104, 2013.

[Loga 07] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille. “A Long-Term Evaluation of Sensing Modalities for Activity Recognition”. In: *Proc UbiComp ’07*, pp. 483–500, 2007.

[Long 10] B. Longstaff, S. Reddy, and D. Estrin. “Improving Activity Classification for Health Applications on Mobile Devices using Active and Semi-Supervised Learning”. In: *Proc Pervasive Health ’10*, pp. 1–7, 2010.

[Ludw 12] W. Ludwig, K.-H. Wolf, C. Duwenkamp, N. Gusew, N. Hellrung, M. Marschollek, M. Wagner, and R. Haux. “Health-enabling technologies for the elderly - An overview of services based on a literature review”. *Comput Meth Prog Bio*, Vol. 106, No. 2, pp. 70–78, 2012.

[Lyle 02] J. Lyle. *Sports Coaching Concepts: A Framework for Coaches’ Behaviour*. Routledge, London, UK, 1st Ed., 2002.

[Löl 10] H. Löllgen, E. Erdmann, and A. Gitt. *Ergometrie: Belastungsuntersuchungen in Klinik und Praxis*. Springer, Heidelberg, Germany, 3rd Ed., 2010.

[Mach 12] A. Machanavajjhala and J. P. Reiter. "Big Privacy: Protecting Confidentiality in Big Data". *XRDS*, Vol. 19, No. 1, pp. 20–23, 2012.

[Mack 16] K. A. Mackintosh, A. H. Montoye, K. A. Pfeiffer, and M. A. McNarry. "Investigating Optimal Accelerometer Placement for Energy Expenditure Prediction in Children Using a Machine Learning Approach". *Physiol Meas*, Vol. 37, No. 10, pp. 1728–1740, 2016.

[MacL 09] H. MacLeod, J. Morris, A. Nevill, and C. Sunderland. "The Validity of a Non-Differential Global Positioning System for Assessing Player Movement Patterns in Field Hockey". *J Sports Sci*, Vol. 27, No. 2, pp. 121–128, 2009.

[Male 00] J. I. Maletic and A. Marcus. "Data Cleansing: Beyond Integrity Analysis". In: *Proc IQ '00*, pp. 1–10, 2000.

[Mang 10] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury. "A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification". *IETE Tech Rev*, Vol. 27, No. 4, pp. 293–307, 2010.

[Mant 01] J. Mäntyjärvi, J. Himberg, and T. Seppänen. "Recognizing Human Motion with Multiple Acceleration Sensors". In: *Proc IEEE Sys Man Cybern '01*, pp. 747–752, 2001.

[Mars 12] F. Marsland, K. Lyons, J. Anson, G. Waddington, C. Macintosh, and D. Chapman. "Identification of Cross-Country Skiing Movement Patterns Using Micro-Sensors". *Sensors*, Vol. 12, No. 4, pp. 5047–5066, 2012.

[Math 03] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler. "Detection of Daily Physical Activities Using a Triaxial Accelerometer". *Med Biol Eng Comput*, Vol. 41, No. 3, pp. 296–301, 2003.

[Math 04a] M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. Coster. "Classification of Basic Daily Movements Using a Triaxial Accelerometer". *Med Biol Eng Comput*, Vol. 42, No. 5, pp. 679–687, 2004.

[Math 04b] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler. "Accelerometry: Providing an Integrated, Practical Method for Long-Term, Ambulatory Monitoring of Human Movement". *Physiol Meas*, Vol. 25, No. 2, pp. R1–R20, 2004.

[McAr 07] W. D. McArdle, F. I. Katch, and V. L. Katch. *Exercise Physiology: Energy, Nutrition, and Human Performance*. Lippincott Williams and Wilkins, Baltimore, MD, 6th Ed., 2007.

[McGi 13] P. M. McGinnis. *Biomechanics of Sport and Exercise*. Human Kinetics, Champaign, IL, 3rd Ed., 2013.

[McGr 12] D. McGrath, B. R. Greene, K. J. O'Donovan, and B. Caulfield. "Gyroscope-based Assessment of Temporal Gait Parameters during Treadmill Walking and Running". *Sports Eng*, Vol. 15, No. 4, pp. 207–213, 2012.

[McIn 95] S. E. McInnes, J. S. Carlson, C. J. Jones, and M. J. McKenna. "The Physiological Load Imposed on Basketball Players During Competition". *J Sports Sci*, Vol. 13, No. 5, pp. 387–397, 1995.

[McKe 00] T. L. McKenzie, S. J. Marshall, J. F. Sallis, and T. L. Conway. "Leisure-Time Physical Activity in School Environments: An Observational Study Using SOPLAY". *Prev Med*, Vol. 30, No. 1, pp. 70–77, 2000.

[McKe 06] T. L. McKenzie, D. A. Cohen, A. Sehgal, S. Williamson, and D. Golinelli. "System for Observing Play and Recreation in Communities SOPARC: Reliability and Feasibility Measures". *J Phys Act Health*, Vol. 3 (supplement 1), pp. S208–S222, 2006.

[McNa 15] D. J. McNamara, T. Gabbett, P. Chapman, G. Naughton, and P. Farhart. "The Validity of Microsensors to Automatically Detect Bowling Events and Counts in Cricket Fast Bowlers". *Int J Sports Physiol Perform*, Vol. 10, No. 1, pp. 71–75, 2015.

[Meye 06] J. A. Meyerhardt, E. I. Giovannucci, M. D. Holmes, A. T. Chan, J. A. Chan, G. A. Colditz, and C. S. Fuchs. "Physical Activity and Survival After Colorectal Cancer Diagnosis". *J Clin Oncol*, Vol. 24, No. 22, pp. 3527–3534, 2006.

[Micu 17] D. Micucci, M. Mobilio, and P. Napoletano. "UniMiB SHAR: A New Dataset for Human Activity Recognition using Acceleration Data from Smartphones". *arXiv preprint arXiv:1611.07688*, 2017.

[Mill 13] H. G. Miller and P. Mork. "From Data to Decisions: A Value Chain for Big Data". *IT Prof*, Vol. 15, No. 1, pp. 57–59, 2013.

[Milu 08] E. Miluzzo, N. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. "Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application". In: *Proc ACM SenSys '08*, pp. 337–350, 2008.

[Mitc 13] E. Mitchell, D. Monaghan, and N. E. O'Connor. "Classification of Sporting Activities Using Smartphone Accelerometers". *Sensors*, Vol. 13, No. 4, pp. 5317–5337, 2013.

[Mont 10] P. G. Montgomery, D. B. Pyne, and C. L. Minahan. "The Physical and Physiological Demands of Basketball Training and Competition". *Int J Sports Physiol Perform*, Vol. 5, No. 1, pp. 75–86, 2010.

[Mont 12] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Regression Analysis*. Wiley-Blackwell, Hoboken, NJ, 5th Ed., 2012.

[Mont 16] A. H. Montoye, B. Dong, S. Biswas, and K. A. Pfeiffer. "Validation of a Wireless Accelerometer Network for Energy Expenditure Measurement". *J Sports Sci*, Vol. 34, No. 21, pp. 2130–2139, 2016.

[Mont 17] A. H. K. Montoye, M. Begum, Z. Henning, and K. A. Pfeiffer. "Comparison of Linear and Non-Linear Models for Predicting Energy Expenditure from Raw Accelerometer Data". *Physiol Meas*, Vol. 38, No. 2, pp. 343–357, 2017.

[Motl 09] R. W. Motl, E. M. Snook, S. Agiovlasitis, and Y. Suh. "Calibration of Accelerometer Output for Ambulatory Adults With Multiple Sclerosis". *Arch Phys Med Rehabil*, Vol. 90, No. 10, pp. 1778–1784, 2009.

[Mova 14] S. Movassaghi, M. Abolhasan, J. Lipman, D. Smith, and A. Jamalipour. “Wireless Body Area Networks: A Survey”. *IEEE Commun Surveys Tuts*, Vol. 16, No. 3, pp. 1658–1686, 2014.

[MPU 11] *MPU-6000 and MPU-6050 Product Specification*. PS-MPU-6000A-00. InvenSense, Inc., October 2011. Rev. 3.1.

[MPU 13] *MPU-9150 Product Specification*. PS-MPU-9150A-00. InvenSense, Inc., September 2013. Rev. 4.3.

[Muba 13] M. Mubashir, L. Shao, and L. Seed. “A Survey on Fall Detection: Principles and Approaches”. *Neurocomputing*, Vol. 100, pp. 144–152, 2013.

[Nait 10] K. Naito, Y. Fukui, and T. Maruyama. “Multijoint Kinetic Chain Analysis of Knee Extension During the Soccer Instep Kick”. *Hum Mov Sci*, Vol. 29, No. 2, pp. 259–276, 2010.

[Nam 14] C. N. K. Nam, H. J. Kang, and Y. S. Suh. “Golf Swing Motion Tracking Using Inertial Sensors and a Stereo Camera”. *IEEE Trans Instrum Meas*, Vol. 63, No. 4, pp. 943–952, 2014.

[Nef 15] T. Nef, P. Urwyler, M. Büchler, I. Tarnanas, R. Stucki, D. Cazzoli, R. Müri, and M. U. “Evaluation of Three State-of-the-Art Classifiers for Recognition of Activities of Daily Living from Smart Home Ambient Data”. *Sensors*, Vol. 15, No. 5, pp. 11725–11740, 2015.

[Niem 90] H. Niemann. *Pattern Analysis and Understanding*. *Springer Series in Information Sciences*, Springer-Verlag, Heidelberg, Germany, 2nd Ed., 1990.

[Nuno 02] H. Nunome, T. Asai, Y. Ikegami, and S. Sakurai. “Three-Dimensional Kinetic Analysis of Side-Foot and Instep Soccer Kicks”. *Med Sci Sports Exerc*, Vol. 34, No. 12, pp. 2028–2036, 2002.

[Nuno 06] H. Nunome, M. Lake, A. Georgakis, and L. K. Stergioulas. “Impact Phase Kinematics of Instep Kicking in Soccer”. *J Sports Sci*, Vol. 24, No. 1, pp. 11––22, 2006.

[Ofli 13] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy. “Berkeley MHAD: A Comprehensive Multimodal Human Action Database”. In: *Proc WACV ’13*, pp. 53–60, 2013.

[Ogil 07] D. Ogilvie, C. E. Foster, H. Rothnie, N. Cavill, V. Hamilton, C. F. Fitzsimons, and N. Mutrie. "Interventions to Promote Walking: Systematic Review". *BMJ*, Vol. 334, No. 7605, p. 1204, 2007.

[Ordo 16] F. J. Ordóñez and D. Roggen. "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition". *Sensors*, Vol. 16, No. 1, pp. 115:1–115:25, 2016.

[Ounp 94] S. Ounpuu. "The Biomechanics of Walking and Running". *Clin Sport Med*, Vol. 13, No. 4, pp. 843–863, 1994.

[Pall 99] R. Pallás-Areny and J. G. Webster. *Analog Signal Processing*. John Wiley & Sons, Hoboken, NJ, 1st Ed., 1999.

[Pan 10] S. J. Pan and Q. Yang. "A Survey on Transfer Learning". *IEEE Trans Knowl Data Eng*, Vol. 22, No. 10, pp. 1345–1359, 2010.

[Pani 12] C. Paniagua, H. Flores, and S. N. Srirama. "Mobile Sensor Data Classification for Human Activity Recognition using MapReduce on Cloud". In: *Proc MobiWIS '12*, pp. 585–592, 2012.

[Park 06] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and K. I. "Activity Classification Using Realistic Data From Wearable Sensors". *IEEE Trans Inf Technol Biomed*, Vol. 10, No. 1, pp. 119–128, 2006.

[Park 12] J. Park, A. Patel, D. Curtis, S. Teller, and J. Ledlie. "Online Pose Classification and Walking Speed Estimation using Handheld Devices". In: *Proc UbiComp '12*, pp. 113–122, 2012.

[Plas 05] G. Plasqui, A. M. Joosen, A. D. Kester, A. H. Goris, and K. R. Westerterp. "Measuring Free-Living Energy Expenditure and Physical Activity with Triaxial Accelerometry". *Obes Res*, Vol. 13, No. 8, pp. 1363–1369, 2005.

[Plas 07] G. Plasqui and K. R. Westerterp. "Physical Activity Assessment with Accelerometers: An Evaluation against Doubly Labeled Water". *Obesity*, Vol. 15, No. 10, pp. 2371–2379, 2007.

[Pobe 06] D. M. Pober, J. Staudenmayer, C. Raphael, and P. S. Freedson. "Development of Novel Techniques to Classify Physical Activity Mode Using Accelerometers". *Med Sci Sports Exerc*, Vol. 38, No. 9, pp. 1626–1634, 2006.

[Popp 10] R. Poppe. "A Survey on Vision-Based Human Action Recognition". *Image Vis Comput*, Vol. 28, No. 6, pp. 976–990, 2010.

[Pree 09] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard. "A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data". *IEEE Trans Biomed Eng*, Vol. 56, No. 3, pp. 871–879, 2009.

[Pric 16] K. Price, S. R. Bird, N. Lythgo, I. S. Raj, J. Y. L. Wong, and C. Lynch. "Validation of the Fitbit One, Garmin Vivofit and Jawbone UP Activity Tracker in Estimation of Energy Expenditure During Treadmill Walking and Running". *J Med Eng Technol*, Vol. 41, No. 3, pp. 208–215, 2016.

[Ragh 14] W. Raghupathi and V. Raghupathi. "Big Data Analytics in Healthcare: Promise and Potential". *Health Inf Sci Syst*, Vol. 2, No. 1, p. 3, 2014.

[Ravi 05] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. "Activity Recognition from Accelerometer Data". In: *Proc IAAI '05*, pp. 1541–1546, 2005.

[Rayk 10] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy. "Learning From Crowds". *JMLR*, Vol. 11, pp. 1297–1322, 2010.

[Redd 10] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. "Using Mobile Phones to Determine Transportation Modes". *TOSN*, Vol. 6, No. 2, pp. 13:1–13:27, 2010.

[Redm 85] D. P. Redmond and F. W. Hegge. "Observations on the Design and Specification of a Wrist-Worn Human Activity Monitoring System". *Behav Res Methods Instrum Comput*, Vol. 17, No. 6, pp. 659–669, 1985.

[Rehm 14] A. Rehman and T. Saba. "Features Extraction for Soccer Video Semantic Analysis: Current Achievements and Remaining Issues". *Artif Intell Rev*, Vol. 41, No. 3, pp. 451–461, 2014.

[Reil 00] T. Reilly, A. M. Williams, A. Nevill, and A. Franks. "A Multidisciplinary Approach to Talent Identification in Soccer". *J Sports Sci*, Vol. 18, No. 9, pp. 695–702, 2000.

[Reil 03] T. Reilly and A. M. Williams. *Science and Soccer*. Routledge, London, UK, 2nd Ed., 2003.

[Reis 12] A. Reiss and D. Stricker. "Creating and Benchmarking a New Dataset for Physical Activity Monitoring". In: *Proc PETRA '12*, no pagination, 2012.

[Rich 91] E. Rich, K. Knight, and S. B. Nair. *Artificial Intelligence*. Tata McGraw-Hill, New Delhi, India, 3rd Ed., 1991.

[Roet 06] D. Roetenberg. *Inertial and Magnetic Sensing of Human Motion*. PhD thesis, University of Twente, Enschede, Netherlands, 2006.

[Rogg 13] D. Roggen, K. Förster, A. Calatroni, and G. Tröster. "The adARC Pattern Analysis Architecture for Adaptive Human Activity Recognition Systems". *J Ambient Intell Human Comput*, Vol. 4, No. 2, pp. 169–186, 2013.

[Roth 07] M. P. Rothney, M. Neumann, A. Béziat, and K. Y. Chen. "An Artificial Neural Network Model of Energy Expenditure Using Nonintegrated Acceleration Signals". *J Appl Physiol*, Vol. 103, No. 4, pp. 1419–1427, 2007.

[Rui 00] Y. Rui, A. Gupta, and A. Acero. "Automatically Extracting Highlights for TV Baseball Programs". In: *Proc MULTIMEDIA '00*, pp. 105–115, 2000.

[Sall 00] J. F. Sallis, N. Owen, and M. J. Fotheringham. "Behavioral Epidemiology: A Systematic Framework to Classify Phases of Research on Health Promotion and Disease Prevention". *Ann Behav Med*, Vol. 22, No. 4, pp. 294–298, 2000.

[Sanc 07] N. Sánchez-Maroño, A. Alonso-Betanzos, and M. Tombilla-Sanromán. "Filter Methods for Feature Selection - A Comparative Study". In: *Proc IDEAL '07*, pp. 178–187, 2007.

[Sapo 08] T. S. Saponas, J. Lester, J. Froehlich, J. Fogarty, and J. Landay. "iLearn on the iPhone: Real-Time Human Activity Classification on Commodity Mobile Phones". Tech. Rep., University of Washington CSE Tech Report UW-CSE-08-04-02, 2008.

[Sazo 11] N. Sazonova, R. C. Browning, and E. Sazonov. "Accurate Prediction of Energy Expenditure Using a Shoe-Based Activity Monitor". *Med Sci Sports Exerc*, Vol. 43, No. 7, pp. 1312–1321, 2011.

[Sche 95] H. Schempf, C. Kraeuter, and M. Blackwell. "Roboleg: A Robotic Soccer-Ball Kicking Leg". In: *Proc ICRA '95*, pp. 1314–1318, 1995.

[Schn 06] P. Schnohr, P. Lange, H. Scharling, and J. S. Jensen. "Long-term Physical Activity in Leisure Time and Mortality from Coronary Heart Disease, Stroke, Respiratory Diseases, and Cancer. The Copenhagen City Heart Study". *Eur J Cardiovasc Prev Rehabil*, Vol. 13, No. 2, pp. 173–179, 2006.

[Scho 17] P. M. Scholl and K. Van Laerhoven. "On the Statistical Properties of Body-Worn Inertial Motion Sensor Data for Identifying Sensor Modality". In: *Proc ISWC '17*, pp. 50–53, 2017.

[Schu 01] Y. Schutz, R. L. Weinsier, and G. R. Hunter. "Assessment of Free-Living Physical Activity in Humans: An Overview of Currently Available and Proposed New Measures". *Obes Res*, Vol. 9, No. 6, pp. 368–379, 2001.

[Schu 13a] D. Schuldhaus, H. Leutheuser, and B. M. Eskofier. "Classification of Daily Life Activities by Decision Level Fusion of Inertial Sensor Data". In: *Proc BodyNets '13*, pp. 77–82, 2013.

[Schu 13b] D. Schuldhaus, H. Leutheuser, and B. M. Eskofier. "Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-art Algorithms Using a Benchmark Dataset". *PLoS One*, Vol. 8, No. 10, e75196, 2013.

[Schu 14a] D. Schuldhaus, S. Dorn, H. Leutheuser, A. Tallner, J. Klucken, and B. M. Eskofier. "An Adaptable Inertial Sensor Fusion-Based Approach for Energy Expenditure Estimation". In: *Proc ICBME '14*, pp. 124–127, 2014.

[Schu 14b] D. Schuldhaus, H. Leutheuser, and B. M. Eskofier. "Towards Big Data for Activity Recognition: A Novel Database Fusion Strategy". In: *Proc BodyNets '14*, pp. 97–103, 2014.

[Schu 15] D. Schuldhaus, C. Zwick, H. Koerger, E. Dorschky, B. Kirk, and B. M. Eskofier. "Inertial Sensor-Based Approach for Shot/Pass Classification During a Soccer Match". In: *KDD Workshop on Large-Scale Sports Analytics '15*, pp. 1–4, 2015.

[Schu 16] D. Schuldhaus, C. Jakob, C. Zwick, H. Koerger, and B. M. Eskofier. "Your Personal Movie Producer: Generating Highlight Videos in Soccer Using Wearables". In: *Proc ISWC '16*, pp. 80–83, 2016.

[Schu 18] D. Schuldhaus, B. M. Eskofier, K. Van Laerhoven, D. Roggen, and A. V. Vasilakos. "Big Data Analytics for Human Activity Recognition: Current State, Proposed Approach, and Open Issues". *Submitted to ACM Comput Surv on 2018/09/26*, pp. 1–23, 2018.

[Schw 12] L. A. Schwarz, D. Mateus, and N. Navab. "Recognizing Multiple Human Activities and Tracking Full-Body Pose in Unconstrained Environments". *Pattern Recogn*, Vol. 45, No. 1, pp. 11–23, 2012.

[Seid 16] T. Seidl, M. Völker, N. Witt, D. Poimann, T. Czyz, N. Franke, and M. Lochmann. "Evaluating the Indoor Football Tracking Accuracy of a Radio-Based Real-Time Locating System". In: P. Chung, A. Soltoggio, C. Dawson, Q. Meng, and M. Pain, Eds., *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, pp. 217–224, Springer, Cham, 2016.

[Seid 97] J. C. Seidell and K. M. Flegal. "Assessing Obesity: Classification and Epidemiology". *Br Med Bull*, Vol. 53, No. 2, pp. 238–252, 1997.

[Shep 99] R. J. Shephard. "Biology and Medicine of Soccer: An Update". *J Sports Sci*, Vol. 17, No. 10, pp. 757–786, 1999.

[Sher 01] D. Sherry and P. Hawkins. "Video Processing Systems for Ball Tracking in Ball Games". Patent number WO 01/41884 A1, June 2001.

[Shin 09] H. Shinkai, H. Nunome, M. Isokawa, and Y. Ikegami. "Ball Impact Dynamics of Instep Soccer Kicking". *Med Sci Sports Exerc*, Vol. 41, No. 4, pp. 889–897, 2009.

[Shva 10] K. Shvachko, H. Kuang, S. Radia, and R. Chansler. "The Hadoop Distributed File System". In: *Proc MSST '10*, pp. 1–10, 2010.

[Siir 12] P. Siirtola and J. Röning. "Recognizing Human Activities User-independently on Smartphones Based on Accelerometer Data". *IJIMAI*, Vol. 1, No. 5, pp. 38–45, 2012.

[Sill 11] C. N. Silla Jr. and A. A. Freitas. "A Survey of Hierarchical Classification Across Different Application Domains". *Data Min Knowl Disc*, Vol. 22, No. 1-2, pp. 31–72, 2011.

[Sing 14] D. Singh and C. K. Reddy. "A Survey on Platforms for Big Data Analytics". *J Big Data*, Vol. 2, No. 1, 2014.

[Soko 09] M. Sokolova and G. Lapalme. "A Systematic Analysis of Performance Measures for Classification Tasks". *Inf Process Manage*, Vol. 45, No. 4, pp. 427–437, 2009.

[Stau 09] J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, and P. Freedson. "An Artificial Neural Network to Estimate Physical Activity Energy Expenditure and Identify Physical Activity Type from an Accelerometer". *J Appl Physiol*, Vol. 107, No. 4, pp. 1300–1307, 2009.

[Stei 15] S. R. Steinhubl, E. D. Muse, and E. J. Topol. "The Emerging Field of Mobile Health". *Sci Transl Med*, Vol. 7, No. 283, 283rv3, 2015.

[Stik 08] M. Stikic, K. Van Laerhoven, and B. Schiele. "Exploring Semi-Supervised and Active Learning for Activity Recognition". In: *Proc ISWC '08*, pp. 81–88, 2008.

[Stoe 14] M. Stoeve. *Comparison of Inertial Sensors for High-Intensity Event Analysis in Sports Applications*. Bachelor's thesis, Digital Sports Group, Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2014.

[Stra 13] S. J. Strath, L. A. Kaminsky, E. U. Ainsworth, B. E., P. S. Freedson, R. A. Gary, C. R. Richardson, D. T. Smith, and A. M. Swartz. "Guide to the Assessment of Physical Activity: Clinical and Research Applications: A Scientific Statement from the American Heart Association". *Circulation*, Vol. 128, No. 20, pp. 2259–2279, 2013.

[Su 05] S. Su, L. Wang, B. Celler, E. Ambikairajah, and A. Savkin. "Estimation of Walking Energy Expenditure by Using Support Vector Regression". In: *Proc IEEE Eng Med Biol Soc '05*, pp. 3526–3529, 2005.

[Sun 12] Z. Sun and G. Fox. "Study on Parallel SVM Based on MapReduce". In: *Proc PDPTA '12*, pp. 16–19, 2012.

[Suth 16] S. Suthaharan. *Machine Learning Models and Algorithms for Big Data Classification - Thinking with Examples for Effective Learning*. Springer Science+Business Media, New York, NY, 2016.

[Swan 12] M. Swan. "Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0". *J Sens Actuator Netw*, Vol. 1, No. 3, pp. 217–253, 2012.

[Teha 06] B. Tehard, C. M. Friedenreich, J.-M. Oppert, and F. Clavel-Chapelon. "Effect of Physical Activity on Women at Increased Risk of Breast Cancer: Results from the E2N Cohort Study". *Cancer Epidemiol Biomarkers Prev*, Vol. 15, No. 1, pp. 57–64, 2006.

[Theo 09] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, San Diego, CA, 4th Ed., 2009.

[Tros 02] S. G. Trost, N. Owen, A. E. Bauman, J. F. Sallis, and W. Brown. "Correlates of Adults' Participation in Physical Activity: Review and Update". *Med Sci Sports Exerc*, Vol. 34, No. 12, pp. 1996–2001, 2002.

[Tsai 15] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos. "Big Data Analytics: A Survey". *Journal of Big Data*, Vol. 2, No. 21, pp. 1–32, 2015.

[Twom 10] N. Twomey, S. Faul, and W. P. Marnane. "Comparison of Accelerometer-based Eenergy Expenditure Estimation Algorithms". In: *Proc PervasiveHealth '10*, pp. 1–8, 2010.

[Uste 13] Y. E. Ustev, O. D. Incel, and C. Ersoy. "User, Device and Orientation Independent Human Activity Recognition on Mobile Phones: Challenges and a Proposal". In: *Proc ACM UbiComp '13*, pp. 1427–1436, 2013.

[Vala 06] E. M. Valanou, C. Bamia, and A. Trichopoulou. "Methodology of Physical-Activity and Energy-Expenditure Assessment: A Review". *J Public Health*, Vol. 14, No. 2, pp. 58–65, 2006.

[Van 00] K. Van Laerhoven and O. Cakmakci. "What Shall We Teach Our Pants?". In: *Proc IEEE ISWC '00*, pp. 77–83, 2000.

[Vath 10] H. Vathsangam, B. Emken, E. Schroeder, D. Spruijt-Metz, and G. S. Sukhatme. "Energy Estimation of Treadmill Walking Using On-Body Accelerometers and Gyroscopes". In: *Proc IEEE Eng Med Biol Soc '10*, pp. 6497–6501, 2010.

[Velt 93] P. H. Veltink, H. B. J. Bussmann, F. Koelma, H. M. Franken, W. L. J. Martens, and R. C. van Lummel. "The Feasibility of Posture and Movement Detection by Accelerometry". In: *Proc IEEE Eng Med Biol Soc '93*, pp. 1230–1231, 1993.

[Velt 96] P. H. Veltink, H. B. Bussmann, W. de Vries, W. L. Martens, and R. C. van Lummel. "Detection of Static and Dynamic Activities using Uniaxial Accelerometers". *IEEE Trans Rehabil Eng*, Vol. 4, No. 4, pp. 375–385, 1996.

[Verm 11] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I. S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer, and P. Doody. "Internet of Things Strategic Research Roadmap". *Internet ThingsGlob Technol Soc Trends*, Vol. 1, pp. 9–52, 2011.

[Walk 16] E. J. Walker, A. J. McAinch, A. Sweeting, and R. J. Aughey. "Inertial Sensors to Estimate the Energy Expenditure of Team-Sport Athletes". *J Sci Med Sport*, Vol. 19, No. 2, pp. 177–181, 2016.

[Walt 11] M. Walter, B. Eilebrecht, T. Wartzek, and S. Leonhardt. "The Smart Car Seat: Personalized Monitoring of Vital Signs in Automotive Applications". *Pers Ubiquit Comput*, Vol. 15, No. 7, pp. 707–715, 2011.

[Wang 18] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu. "Deep Learning for Sensor-Based Activity Recognition: A Survey". *Pattern Recognit Lett*, Vol. (in press), 2018.

[Warb 06] D. E. R. Warburton, C. W. Nicol, and S. S. D. Bredin. "Health Benefits of Physical Activity: The Evidence". *CMAJ*, Vol. 174, No. 6, pp. 801–809, 2006.

[Warr 10] J. M. Warren, U. Ekelund, H. Besson, A. Mezzani, N. Geladas, and L. Vanhees. "Assessment of Physical Activity - A Review of Methodologies with Reference to Epidemiological Research: A Report of the Exercise Physiology Section of the European Association of Cardiovascular Prevention and Rehabilitation". *Eur J Cardiovasc Prev Rehabil*, Vol. 17, No. 2, pp. 127–139, 2010.

[Weis 05] S. Weisberg. *Applied Linear Regression*. John Wiley & Sons, Inc., Hoboken, NJ, 3rd Ed., 2005.

[Whit 09] T. White. *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., Sebastopol, CA, USA, 1st Ed., 2009.

[Will 00] A. M. Williams and T. Reilly. "Talent Identification and Development in Soccer". *J Sports Sci*, Vol. 18, No. 9, pp. 657–667, 2000.

[Witt 11] I. H. Witten, E. Frank, and M. A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, Burlington, MA, 3rd Ed., 2011.

[Worl 09] World Health Organization. "Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks". 2009.

[Worl 11] World Health Organization. "Global Status Report on Noncommunicable Diseases 2010". 2011.

[Wu 07a] J. K. Wu, L. Dong, and W. Xiao. "Real-Time Physical Activity Classification and Tracking using Wearable Sensors". In: *Proc IEEE ICICS '07*, pp. 1–6, 2007.

[Wu 07b] T. Wu, C. Lian, and J. Y. Hsu. "Joint Recognition of Multiple Concurrent Activities using Factorial Conditional Random Fields". In: *Proc AAAI '07*, pp. 82–87, 2007.

[Wu 14] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding. "Data Mining with Big Data". *IEEE Trans Knowl Data Eng*, Vol. 26, No. 1, pp. 97–107, 2014.

[Xu 14] S. Xu, H.-J. Yoon, and G. Tourassi. "A User-Oriented Web Crawler for Selectively Acquiring Online Content in e-Health Research". *Bioinformatics*, Vol. 30, No. 1, pp. 104–114, 2014.

[Yan 11] Y. Yan, R. Rosales, G. Fung, and J. G. Dy. "Active Learning from Crowds". In: *Proc ICML '11*, pp. 1161–1168, 2011.

[Yang 09] J. Yang. "Toward Physical Activity Diary: Motion Recognition Using Simple Acceleration Features with Mobile Phones". In: *Proc IMCE '09*, pp. 1–10, 2009.

[Yang 10] C.-C. Yang and Y.-L. Hsu. "A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring". *Sensors*, Vol. 10, No. 8, pp. 7772–7788, 2010.

[Yang 14] G.-Z. Yang. *Body Sensor Networks*. Springer, London, UK, 2nd Ed., 2014.

[Zaha 10] M. Zaharia, M. Chowdhury, M. J. Franklin, S. S., and I. Stoica. "Spark: Cluster Computing with Working Sets". In: *Proc HotCloud '10*, pp. 1–7, 2010.

[Zake 10] I. F. Zakeri, A. L. Adolph, M. R. Puyau, F. A. Vohra, and N. F. Butte. "Multivariate Adaptive Regression Splines Models for the Prediction of Energy Expenditure in Children and Adolescents". *J Appl Physiol*, Vol. 108, No. 1, pp. 128–136, 2010.

[Zapp 07] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster. "Activity Recognition from On-body Sensors by Classifier Fusion: Sensor Scalability and Robustness". In: *Proc ISSNIP '07*, pp. 281–286, 2007.

[Zern 78] R. F. Zernicke and E. M. Roberts. "Lower Extremity Forces and Torques During Systematic Variation of Non-Weight Bearing Motion". *Med Sci Sports*, Vol. 10, No. 1, pp. 21–26, 1978.

[Zhan 12] M. Zhang and A. A. Sawchuk. "USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors". In: *Proc UbiComp '12*, pp. 1036–1043, 2012.

[Zhan 14] K. Zhang and X.-W. Chen. "Large-Scale Deep Belief Nets With MapReduce". *IEEE Access*, Vol. 2, pp. 395–403, 2014.

[Zhan 16] Y. Zhang, T. Cao, S. Li, X. Tian, L. Yuan, H. Jia, and A. V. Vasilakos. "Parallel Processing Systems for Big Data: A Survey". *Proc IEEE, Special Issue on Big Data*, 4th Quarter 2016.

[Zhen 14] Y.-L. Zheng, X.-R. Ding, C. C. Poon, B. P. Lo, H. Zhang, X.-L. Zhou, G.-Z. Yang, N. Zhao, and Y.-T. Zhang. "Unobtrusive Sensing and Wearable Devices for Health Informatics". *IEEE Trans Biomed Eng*, Vol. 61, No. 5, pp. 1538–1554, 2014.

[Zhen 15] Y. Zheng. "Human Activity Recognition Based on the Hierarchical Feature Selection and Classification Framework". *JECE*, Vol. 2015, No. 140820, 2015.

[Ziko 11] P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill, New York, NY, 1st Ed., 2011.

[Ziko 12] P. Zikopoulos, D. deRoos, K. Parasuraman, T. Deutsch, D. Corrigan, and J. Giles. *Harness the Power of Big Data - The IBM Big Data Platform*. McGraw-Hill, New York, NY, 1st Ed., 2012.

[Zoub 09] N. Zouba, F. Bremond, and M. Thonnat. "Multisensor Fusion for Monitoring Elderly Activities at Home". In: *Proc AVSS '09*, pp. 98–103, 2009.

[Zwic 12] C. Zwick. *Development and Evaluation of a New Ball Speed Measurement System Using Inertial Sensing Technology*. Master's thesis, University of Applied Sciences Technikum Wien - Sports Equipment Technology, 2012.

[Zwic 16] C. Zwick, R. F. Kirk, C. Dibenedetto, E. A. Dorschky, B. M. Eskofier, D. Schuldhaus, P. G. Laitenberger, and I. J. Sabberton. "Shoes for Ball Sports". US Patent Application US 20160309834 A1, adidas AG, October 27, 2016.

Human Activity Recognition (HAR) deals with the automatic recognition of physical activities and plays a major role in the health and sports sector. Knowledge about the performed activities can be used to monitor compliance regarding physical activity recommendations, investigate the causes of physical activity behavior, implement sport-specific training programs, and replicate the physical demands during sport competition. Currently available tools for HAR often rely on questionnaires which involve problems in the reliability when recalling activities.

In this thesis, algorithms for HAR are introduced and evaluated which apply machine learning techniques to inertial sensor data. Daily as well as sport-specific activities are considered including sitting, washing dishes, climbing stairs, and kicking in soccer. Besides the development and implementation of algorithms, mandatory extensions regarding the design of HAR systems are further identified and future research directions are provided.

ISBN 978-3-96147-225-3

9 783961 472253

FAU UNIVERSITY PRESS 2019