Enhancing Wellbeing Using Artificial Intelligence Techniques

"Améliorer le bien-être grâce à l'utilisation de Techniques d'Intelligence Artificielle"

THESIS

submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland

by

Athanasios I. KYRITSIS

Under the direction of

Prof. Dimitri KONSTANTAS, co-supervisor
Dr. Michel DERIAZ, co-supervisor

in fulfillment of the requirements for the degree of

Docteur ès économie et management
mention *information systems*

Jury members:

Prof. Gilles FALQUET, Chair, University of Geneva
Dr. Michel DERIAZ, co-supervisor, University of Geneva
Prof. Dimitri KONSTANTAS, co-supervisor, University of Geneva
Prof. Katarzyna WAC, University of Geneva
Prof. Giovanna DI MARZO SERUGENDO, University of Geneva
Prof. Theodoros KOSTOULAS, Bournemouth University
Mr. Yacine BENMANSOUR, Director, DOSI, État de Genève

Thesis no 75
Geneva, December 19, 2019

La Faculté d'économie et de management, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par-là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 3 février 2020

Dean
Marcelo OLARREAGA

Impression d'après le manuscrit de l'auteur

UNIVERSITÉ DE GENÈVE

GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

LE DOYEN

A QUI DE DROIT

IMPRIMATUR

Je, soussigné, Professeur Marcelo OLARREAGA, Doyen de la Faculté d'Economie et de Management, confirme que **Monsieur Athanasios KYRITSIS** obtient l'imprimatur pour sa thèse N°75, suite à sa soutenance publique du 19 décembre 2019 pour le grade de docteur en systèmes d'information.

MO

Prof. Marcelo OLARREAGA
Doyen

Genève, le 3 février 2020
MO/GK/kra

Keep Ithaka always in your mind.
Arriving there is what you are destined for.
But do not hurry the journey at all.
Better if it lasts for years,
so you are old by the time you reach the island,
wealthy with all you have gained on the way,
not expecting Ithaka to make you rich.

Ithaka gave you the marvelous journey.
Without her you would not have set out.
She has nothing left to give you now.

And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you will have understood by then what these Ithakas mean.

Constantine P. Cavafy [1911]

Acknowledgments

I would like to start by thanking both my co-supervisors, Michel Deriaz and Dimitri Konstantas. Thank you, Michel, for showing trust in me and for giving me the opportunity to pursue a Ph.D. Thanks for your supervision, for allowing me to work on a variety of projects and to enhance my skills on topics interesting to me. Thank you, Dimitri, for always being there when I needed you. Thanks for allowing me to teach courses at the university, an activity that I particularly enjoyed, and for the multiple conversations we had. Our communication was always direct and pleasant. I would also like to thank the members of the jury: Gilles Falquet, Katarzyna Wac, Giovanna Di Marzo Serugendo, Theodoros Kostoulas, and Yacine Benmansour. Their time and their valuable feedback helped me improve this work.

Moreover, I would like to acknowledge all the funding agencies that co-supported my research. The State Secretariat for Education, Research and Innovation of the Swiss federal government and the European Union, in the frame of the EU AAL projects StayActive (aal-2013-6-126), SmartHeat (aal-2014-153), and EDLAH2 (aal-2015-022). The Commission for Technology and Innovation CTI (currently replaced by Innosuisse), of the Swiss federal government, in the frame of the CTI projects F2D (CTI grant 15876.2 PFES-ES) and Recover@home (CTI grant 18566.2 PFES-ES). Finally, the Swiss National Science Foundation in the frame of the FNS project From Lab to Life: Cognitive Aging Revisited (contract 100019_165572).

I would like to thank my colleagues in the TaM group of CUI of the University of Geneva for the collaboration we had in the office and the fun outside of it. Mainly I would like to thank Grigorios for being a good friend, for the guidance he provided me with his experience in the academic world and for his valuable comments regarding my work. With Abbass, we were always discussing ideas and implementations since the beginning of our parallel journeys towards a Ph.D. Kevin taught me how to be professional by being an exemplary project leader. Tiago helped me with my first steps as a junior researcher and developer in the team. Michaël helped me to become a better software engineer. Last but not least, Alberto, Carlos, and Mauricio, although never directly collaborated in a project, each one made this whole journey towards my Ph.D. more fun.

I would also like to thank people from Battelle with whom I have spent the last four and a half years in the same working environment. These include Allan, Alex, Arianna, Coralie, Marios, Michal, and Stephanos. Also, Marie-France and Daniel, who were always kind, pleasant, and willing to help.

People outside Battelle have also contributed towards my psychological stability all those years. Thank you, Biagio, Julia, Francesco, and Elpida for all the food and wine adventures. Gianni, Michu, Axel, Paschali, and Stefane for your ear tolerance in the low frequencies. Mano, Margarita, Niko, Ioanna, and George for all the trips and the travel stories we've shared. Nadia, Pano, and Ritsa for the good times we've had together. Gianna for the artistic and start-up discussions we've had. Spyro Friday for the advice on machine learning and for being a friend.

Finally, I would like to thank Evi for the love and compassion she has given me in the last eight years. My brother George, for continuously supporting me, for advising me during my first steps on machine learning and for always being around. My parents, Gianni and Vana, for their relentless love.

Geneva, December 18, 2019 A. K.

Abstract

The landscape of technology is rapidly evolving, and its pace of growth is not slowing down. During the last decades, and especially after the mass adoption of the internet, new technological advances have revolutionized every aspect of human life. We are living in the ubiquitous computing era, where connected devices form the internet of things and produce data faster than we can logically process. With the latest advances in mobile communications and with the widespread use of smartphones and connected sensors, various aspects of wellbeing can be monitored and improved. The goal of this thesis is to propose new algorithms, methodologies, and applications that can be used as components in health and wellbeing systems that support healthy aging, enhance human-machine interactions, and support postoperative rehabilitation with the use of modern connected devices and machine learning techniques. The core research question of the thesis is the following: "How can artificial intelligence techniques ameliorate human wellbeing by using data produced by modern smart devices?"

In this thesis, we initially investigate how modern technology and applications that support healthy aging can be attractive to be used by older adults. We conducted a research study in order to understand the needs and requirements of older adults, as well as the reasons behind the age technology gap. In this study, we provide useful insights to developers who are building user-centric applications and want to appeal to users of all age groups.

We are then presenting three components we developed that leverage the use of modern technologies to improve daily living. We are presenting a low-cost, easy to deploy and use, indoor localization system with room-level accuracy. The algorithm we are proposing takes into account signals from radio frequency beacon transmitters like Bluetooth beacons and the room geometry when inferring a position. We continue by presenting a mood and stress detection system that monitors non-invasive sensor data and smartphone usage patterns in order to estimate the psychological state of the users. We are then presenting methods of detecting abnormal behavior from activity and mobile app usage data. Unsupervised anomaly detection techniques were employed for detecting potentially problematic scenarios during the day and for triggering relevant actions.

In the frame of building context-aware applications, we have also contributed towards developing activity recognition systems using inertial sensors. Initially, we present all the parameters that have an impact on every stage of the development of an activity recognition system. We have analyzed the significance of the parameters in the design, implementation, testing, and evaluation phase of an activity recognition system with an experiment that included several activities to be identified. We put all this acquired knowledge into practice by building an activity recognition system optimized for a specific scenario. We created a gait recognition system targeting people who have undergone lower body orthopedic surgery. We collaborated with the physiotherapist team of Hirslanden Clinique La Colline, an orthopedic clinic in Geneva. We built a system that is meant to be used by the physiotherapists during the rehabilitation phase of a patient, in order to be able to monitor the evolving gait pattern of the patient during everyday life, and not only during the time-limited physiotherapy sessions.

The thesis contributions include techniques on how machines can learn about different human aspects from data. All the presented components in the thesis can be used to support wellbeing systems that enhance daily life.

Keywords: Abnormality detection, Active assisted living, Activity recognition, Ambient assisted living, Bluetooth, Bluetooth low energy, Deep learning, eHealth, Feature extraction, Health informatics, Indoor localization, Indoor positioning, Machine learning, mHealth, Mobile applications, Mobile devices, Neural networks, Pattern recognition, Postoperative rehabilitation, Room-level accuracy, RSSI, Senior citizens, Smart devices, Smartphones, Smartwatches, Stress detection, Survey, User requirements, Wearable computers, Wearable sensors.

Résumé

Le monde de la technologie évolue rapidement et son rythme de croissance ne ralentit pas. Au cours des dernières décennies, et surtout après l'adoption massive d'Internet, les nouvelles avancées technologiques ont révolutionné tous les aspects de la vie humaine. Nous vivons à l'ère de l'informatique omniprésente, où les appareils connectés forment l'Internet des objets et produisent des données plus rapidement que nous ne pouvons logiquement les traiter. Avec les dernières avancées dans les communications mobiles et avec l'utilisation généralisée des smartphones et des capteurs connectés, divers aspects du bien-être peuvent être observés et améliorés.
Le but de cette thèse est de proposer de nouveaux algorithmes, méthodologies et applications pouvant être utilisés comme composants dans des systèmes de santé et de bien-être qui soutiennent le vieillissement en bonne santé, améliorent les interactions homme-machine et soutiennent la réadaptation postopératoire avec l'utilisation d'appareils modernes connectés et de techniques de Machine Learning.
La question principale de recherche de la thèse est la suivante : "Comment les techniques d'intelligence artificielle peuvent-elles améliorer le bien-être humain en utilisant les données produites par les appareils intelligents modernes ?"
Dans cette thèse, nous étudions dans un premier temps comment les technologies et applications modernes qui soutiennent le vieillissement en bonne santé peuvent être attrayantes pour être utilisées par les personnes âgées. Nous avons mené une étude afin de comprendre les besoins et les exigences des personnes âgées, ainsi que les raisons du fossé technologique lié à l'âge. Dans cette étude, nous fournissons des informations utiles aux développeurs qui créent des applications centrées sur l'utilisateur et souhaitent cibler tous les groupes d'âge.
Nous présentons ensuite trois composants que nous avons développés qui tirent parti de l'utilisation des technologies modernes pour améliorer la vie quotidienne.

Résumé

Nous présentons un système de positionnement en intérieur à faible coût, facile à déployer et à utiliser avec une précision au niveau de la pièce. L'algorithme que nous proposons prend en compte les signaux de balises émettrices de fréquences radio comme les balises Bluetooth et la géométrie de la pièce lors de la déduction d'une position. Nous continuons en présentant un système de détection de l'humeur et du stress qui observe les données des capteurs non invasifs et les habitudes d'utilisation des smartphones afin d'estimer l'état psychologique des utilisateurs. Nous présentons ensuite des méthodes de détection des comportements anormaux à partir des données d'activité et d'utilisation des applications mobiles. Des techniques de détection d'anomalies non supervisées ont été utilisées pour détecter des scénarios potentiellement problématiques pendant la journée et pour déclencher des actions pertinentes.

Dans le cadre de la création d'applications sensibles au contexte, nous avons également contribué au développement de systèmes de reconnaissance d'activité utilisant des capteurs inertiels. Dans un premier temps, nous présentons tous les paramètres qui ont un impact à chaque étape du développement d'un système de reconnaissance d'activité. Nous avons analysé l'importance des paramètres dans la phase de conception, de mise en œuvre, de test et d'évaluation d'un système de reconnaissance d'activité en faisant une expérience comprenant plusieurs activités à identifier. Nous mettons en pratique toutes ces connaissances acquises en créant un système de reconnaissance d'activités optimisé pour un scénario spécifique. Nous avons créé un système de reconnaissance de la démarche ciblant les personnes qui ont subi une chirurgie orthopédique dans la partie basse du corps. Nous avons collaboré avec l'équipe de physiothérapeutes de Hirslanden Clinique La Colline, une clinique orthopédique à Genève. Nous avons construit un système qui est destiné à être utilisé par les physiothérapeutes pendant la phase de réadaptation d'un patient, afin de pouvoir suivre l'évolution de la démarche du patient au cours de la vie quotidienne, et pas uniquement pendant les séances de physiothérapie qui sont limitées dans le temps.

Les contributions de thèse comprennent des techniques sur la façon dont les machines peuvent en apprendre davantage sur les différents aspects humains à partir des données. Tous les éléments présentés dans la thèse peuvent être utilisés pour soutenir des systèmes de bien-être qui améliorent la vie quotidienne.

Mots-clés: Active assisted living, Ambient Assisted Living, Appareils mobiles, Applications mobiles, Apprentissage automatique, Apprentissage profond, Besoins des utilisateurs, Bluetooth, Bluetooth Low Energy, Capteurs portables, Dispositifs intelligents, Détection d'anomalie, Détection de stress, eHealth, Extraction de caractéristiques, Informatique de la santé, Localisation en intérieur, mHealth, Montres intelligentes, Personnes âgées, Positionnement intérieur, Précision au niveau de la pièce, Reconnaissance d'activité, Reconnaissance de formes, RSSI, Réseaux de neurones artificiels, Rééducation postopératoire, Smartphones, Sondage, Technologie mettable.

Contents

1 Introduction

1.1 Overview

Technology has been evolving rapidly over the past decades. In contrast to the relatively stationary early days of the internet, today, the vast majority of the human population is continuously connected to the digital world via smartphones or other mobile devices. There has been a plethora of newly developed connected devices that makes their way into the market each year, and many Internet of Things (IoT) platforms are capturing data about their users and their environments. Sensors of all kinds, monitoring different aspects can be found virtually everywhere, producing data on a large scale.

Technology, however, penetrates the lives of people in different ways and varying degrees. Theoretically, the acceptance of a technology depends on its perceived usefulness, its ease of use, the attitude, and the intention of the users to actually use it. There is, however, an age technology gap, and the aging population shift is expected to deteriorate it. Although older generations have started to embrace digital life and to adopt modern smart devices such as smartphones and tablets, they often lack awareness of the latest technological advances and services they can profit from. This is why it is essential for designers and developers to properly communicate and motivate technology usage to people of all ages.

The widespread use of smartphones has increased the need for location-based services. Almost all smartphones include Global Navigation Satellite System (GNSS) receivers like the Global Positioning System (GPS), and most people have already the experience of being positioned outdoors, via the use of mobile applications such as Google Maps. It is not the case, however, for indoor scenarios. Many technologies have been evaluated to be used for indoor localization, but it has yet to exist a generic and universal solution, as GPS is for outdoor localization. Existing indoor localization solutions are also oftentimes hard to deploy and of considerable cost.

Stress has become an integral part of modern society and is triggered by the challenges of a fast-paced modern life. Stress is psychological pain and can be detected by several reactions of the human body's physiological mechanisms. Measuring those responses, however, would require special instruments and would be intrusive to the users. At the same time, smartphone technology is continuously evolving, and in conjunction with smartphone embedded sensors produce lots of user-related data. It is possible, therefore, also by taking into account the increasing time people spend with their smartphones, to study human stress, mood and behavior in a non-intrusive manner by monitoring human-smartphone interaction during daily life.

Inertial sensors are nowadays omnipresent and can be found in all smartphones and smartwatches. Data recorded from inertial sensors can be used to monitor user movements and even to detect complex activities in the frame of designing context-aware applications. Developing an activity recognition system, however, is challenging and a very complex task because of the large number of variables that impact the design of such a system. Moreover, developing such a system not only for showcasing but to bring value to an industry adds another level of complexity to the whole research and development process.

In this thesis, we are researching components from which a system targeting to enhance human wellbeing can benefit. We are presenting new algorithms, methodologies, and applications that support healthy aging, enhance human-machine interactions, and support postoperative rehabilitation. We conducted a study where we are identifying the needs of older adults that are using modern applications and devices. We are proposing a low cost and easy to use and deploy indoor positioning system. We are proposing a mood and stress detection system running on a smartphone, as well as abnormal behavior detection techniques from mobile devices usage patterns. Lastly, we have conducted in-depth research on designing activity recognition systems using inertial sensors, and we have created such a system to assist physiotherapists to monitor patients during the postoperative rehabilitation phase.

1.2 Contributing Publications

The content of this thesis is based on six scientific publications. Parts of the publications were revised to either include more information, or cite the newest related works in the fields, or to ameliorate the flow of this manuscript. Those publications are:

1. *'User Requirement Analysis for the Design of a Gamified Ambient Assisted Living Application'*, Kyritsis, A.I., Nuss, J., Holding, L., Rogers, P., O'Connor, M., Kostopoulos, P., Suffield, M., Deriaz, M. and Konstantas, D., In International Conference on Computers Helping People with Special Needs (ICCHP 2018), Linz, Austria, July 2018.

2. *'A BLE-Based Probabilistic Room-Level Localization Method'*, Kyritsis, A.I., Kostopoulos, P., Deriaz, M. and Konstantas, D., In International Conference On Localization and GNSS (ICL-GNSS 2016), Barcelona, Spain, June 2016.

3. *'Anomaly Detection Techniques in Mobile App Usage Data among Older Adults'*, Kyritsis, A.I., Deriaz, M. and Konstantas, D., In IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom 2018), Ostrava, Czech Republic, September 2018.

4. *‘Considerations for the Design of an Activity Recognition System Using Inertial Sensors’*, Kyritsis, A.I., Deriaz, M. and Konstantas, D., In IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom 2018), Ostrava, Czech Republic, September 2018.

5. *‘Gait Recognition with Smart Devices Assisting Postoperative Rehabilitation in a Clinical Setting’*, Kyritsis, A.I., Willems, G., Deriaz, M. and Konstantas, D., In IEEE International Conference on Artificial Intelligence for Industries (ai4i 2018), Laguna Hills, CA, USA, September 2018.

   This publication was granted the IEEE Computer Society Best Paper Award.

6. *‘Gait Pattern Recognition Using a Smartwatch Assisting Postoperative Physiotherapy’*, Kyritsis, A.I., Willems, G., Deriaz, M. and Konstantas, D., In International Journal of Semantic Computing, Vol. 13, No. 2, p. 245–257, 2019.

1.3 Thesis Frame and Research Projects

Several European and Swiss research projects supported the works presented in this thesis. The results of our research were put into practice and contributed to the deliverables and the overall success of the projects. Those belonged to different funding schemes including the European AAL programme, and the Swiss national ones Innosuisse (formerly known as CTI) and FNS.

AAL

The Active and Assisted Living (AAL) is a European funding program [1] that aims to support healthy aging, create a better quality of life for older adults and support innovation and technologies for this cause. The AAL projects that co-funded this work are StayActive, SmartHeat, and EDLAH2. Parts of the works of this thesis were adjusted to meet the needs of the projects and were validated through those.

The StayActive [2] project's goal was to detect stress and prevent burnout conditions through the use of a smartphone. A fundamental requirement was to analyze data that a modern smartphone is able to capture so that our solution would be completely non-invasive. Our approach was to explore how the various parameters that were captured by the smartphone could translate to the behavior of the user, that would then translate to the user's stress level. The way of analyzing those data to infer stress and mood that is presented in this thesis was used in the context of this project.

SmartHeat [3] was a smart energy saving heating system whose target is to heat every room of a house independently according to the users' habits. The overall system would monitor the users' presence in the house via connected Internet of Things (IoT) sensors and would then automatically and optimally heat the house, room by room, in order to optimize the wellbeing and the comfort of the users. One way to provide the contextual information regarding the occupancy of each room is via the use of a Bluetooth beacon-based indoor localization system with room-level accuracy.

The Enhanced Daily Living and Health 2 (EDLAH2) [4] project aimed to create a holistic tablet application that would support autonomous healthy aging as long as possible. The tablet application would introduce to the older adults many new components and technologies and would be used in conjunction with a connected bracelet. The functionality that the app encapsulates includes video games, health monitoring, activity tracking, social interaction, and web browsing. The application would also silently monitor the behavior of the user and would trigger alarms to the family and professional caregivers in case an abnormality is detected.

Innosuisse

Innosuisse (formerly known as CTI) is the Swiss Innovation Promotion Agency [5]. Innosuisse's role is to promote science-based innovation in the interests of industry and society in Switzerland by promoting partnerships between academia and the market. Innosuisse supported this work through the Innosuisse projects F2D and Recover@home.

F2D [6] was about creating a fall detection system running on a smartwatch. The user would wear a popular and non-stigmatizing smartwatch, and unlike traditional systems, a home base station would not be required. In case of a fall, the system would trigger an alarm to professional caregivers based on specific contextual information, including the location of the user either indoors or outdoors.

The Recover@home [7] project is a tool to support the work of physiotherapists during the postoperative rehabilitation of patients that have undergone a lower body operation. Each patient would wear a smartwatch throughout the recovery process, and using the inertial sensors different gait patterns would be recognized. The physiotherapist would be able to track the evolution of the gait pattern of each user continuously, even outside the time-limited physiotherapy sessions.

FNS

The last source of support for this work comes from the Swiss National Science Foundation (FNS) [8]. FNS supports scientific research in various academic disciplines in order to support the spread of knowledge in society, the economy and politics, and demonstrate the value of research. Parts of the work presented in this thesis were developed to support the research for the From Lab to Life: Cognitive Aging Revisited project.

The From Lab to Life: Cognitive Aging Revisited [9] project aims at exploring the functioning of prospective memory by testing whether and in what terms cognitive decline measured in the laboratory transfers to everyday life. A smart sensor network system is used to collect information about the everyday behavior of the user. Memory performance is also measured with a smartphone using experience sampling methods.

1.4 Thesis Structure and Contributions

The core research question of the thesis is the following: "How can artificial intelligence techniques ameliorate human wellbeing by using data produced by modern smart devices?" The author of this thesis is proposing new algorithms, methodologies, and applications that support healthy aging, enhance human-machine interactions, and support postoperative rehabilitation. Modern smart devices are omnipresent, and machine learning backed solutions are infiltrating every sector of society, enhancing processes and revolutionizing entire industries. The thesis is divided into chapters, each adding a component to a more general wellbeing system by answering a different research question. An overview of all the components of a wellbeing system that were researched can be seen in Figure 1.1.

Initially, the author is identifying the user requirements and the needs of older adults that are using modern applications and devices. Older adults have traditionally been more resistant in adopting current trends of technology. It is necessary, therefore, for every developer of such applications to firstly understand the older users' needs in order also to appeal to this group of our society.

Before presenting any application or methodology, the author is presenting a user requirements study for the use case of designing a modern tablet application in Chapter 2. The research question of this chapter is, "What is the current view of people over 55 towards mobile technology, and how can it become more appealing?" A set of questions divided into several thematic areas was developed with the aid of end-user experts, and older adults, mostly over 60 years old, living either on their own or in care homes answered it. The goals of the survey were to:

1. understand the current rate at which older adults are using technology, in the form of computers and the internet,

Figure 1.1: Overview of the researched components of a wellbeing system.

2. evaluate the willingness of older adults to use new smart devices including smartphones, tablets and connected devices for health and physical activity monitoring,
3. study the motivation behind using newly developed technologies and the needs that they should cover, and
4. estimate how willing older adults would be to play video games and to socialize with new means of communication.

To summarize the results of the study, there is still a lack of awareness among our target group about what can be achieved using modern technological advances. The study group, however, expressed that they would be willing to try new devices and new kinds of technologies, provided that there is proper guidance, explanations, and motivation to do so. With this questionnaire study, the author is providing an insightful view of how the elderly feel towards several aspects of technology. This work has been published in [10]. The contributions of the author in this study were the involvement in the definition of the questionnaire, the data analysis, and the presentation of the results.

Subsequently, the author of this thesis is presenting new and state-of-the-art applications that utilize modern smart devices. Smartphones have become ubiquitous during the last years, and more devices than ever are connected online, forming the IoT. Those devices can be used to monitor various aspects of daily living, and new systems that are aiming at improving the quality of life are constantly being developed. The author is presenting three such components, an indoor positioning system, a stress and mood detection system, and an abnormal behavior detection system.

The first component is an indoor positioning system that uses Bluetooth Low Energy beacons and is presented in Chapter 3. The research question of this chapter is, "How can room-level indoor localization be effectively achieved by using one Bluetooth beacon per room?" The goals of the developed component were to:

1. localize users indoors using Bluetooth beacons,
2. maximize room-level accuracy,
3. minimize the cost by minimizing the additional hardware needed, that is the number of beacons, and
4. require minimum set-up time.

The author has introduced an algorithm that takes into account the geometry of the rooms to be localized, in terms of the surface area and the heights of them. Only a single beacon is attached to the ceiling in the center of every room. A threshold-based approach is classifying every RSSI reading from every beacon and is calculating probabilities for every room. After deploying in two different locations, the author has evaluated the performance of the system, and achieved an improvement of room estimation accuracy, especially in the boundary locations of the rooms. The proposed component has been published in [11] and can be deployed and used by any application requiring room-level indoor localization. The contributions of the author in this study were the conception and the definition of the localization algorithm, the development of the Android application for data acquisition, the testing in two different scenarios, and the data analysis.

In Chapter 4, the author is proposing a stress and mood detection component running on a typical smartphone. The research question of this chapter is, "How can smartphone data be translated to mood and stress states?" The goals of the component were to:

1. use only information that can be captured by a smartphone, either through the use of it or through its sensors, and
2. infer mood and stress levels from the noticed historical patterns.

The author is presenting a method to fuse heterogeneous data coming from different sensors. Several machine learning methods were tested, and a lift of predictive performance was achieved for all psychological states of interest. The contributions of the author in this study were the development of parts of the Android application for data acquisition, the organization of the experiment, and the data analysis.

Moreover, the author is presenting an abnormal behavior detection component running on a tablet in Chapter 5. The research question of this chapter is, "How can behavior abnormalities be detected from the usage patterns of a tablet launcher application?" The component's goals were to:

1. record the usage of several aspects of the tablet launcher application,
2. monitor the daily activity of users, in terms of steps and resting time, via a connected wearable device, and
3. require no direct input from the users.

The author proposes several ways to fuse data and create relevant datasets. Different anomaly detection techniques were applied to the datasets to detect anomalies. The proposed abnormal behavior detection component can be used in remote monitoring applications where alarms need to be triggered and sent to family members or professional caregivers. This work has been published in [12]. The contributions of the author in this study were the development of parts of the Android application for data acquisition, and the data analysis.

Lastly, the author of this thesis is presenting contributions to applied artificial intelligence for Activity Recognition (AR). Knowing user activities can be a piece of valuable information for a wide array of modern applications, both in academia and in industry. The author has approached activity recognition with the use of inertial sensor data. The contributions in this domain include a general introduction to developing activity recognition systems using inertial sensors and a more specific system developed to be used by physiotherapists during the postoperative rehabilitation phase of a patient to monitor the evolving gait pattern of the patient.

The author is presenting and discussing in detail research challenges in the development of activity recognition systems using inertial sensors in Chapter 6. The research question of this chapter is, "What are the main considerations in the design of an inertial sensor-based activity recognition system, and what are the tradeoffs?" The goals of this work were to:

1. identify factors that affect the design of an activity recognition system,
2. group considerations and parameters that impact every AR system during the design, implementation, testing and evaluating phase,
3. exemplify the role of the parameters as mentioned above with an experiment, and
4. provide a concise reference for future endeavors in the field of human activity recognition.

Before creating an activity recognition system, it is crucial to define the activities of interest, as those should be diverse enough for the system to be able to distinguish. In a continuous data collection scenario, the irrelevant activities form the majority of the dataset and ignoring them forms another challenge by itself. The dataset that is available to train the AR system may bring more challenges on the table as it reflects the way the experiment was designed. The way data are acquired, preprocessed and segmented, and the way features may be engineered can wildly vary across different applications, according to the needs of the developed system. There are also application-specific challenges, and miscellaneous tradeoffs should be considered. The author is showcasing the importance of various parameters by training multiple machine learning models on data captured by an experiment. The author is then evaluating the predictive performance of all those models and discusses how accuracy fluctuates for different design decisions. This work has been published in [13] and can be used as a reference in the design of an inertial sensor based activity recognition system. The contributions of the author in this study were the classification of the presented considerations for the design of an activity recognition system, the definition and the execution of the experiment, and the data analysis.

The author continues by applying all the considerations as discussed above into developing an activity recognition component for a real-world application in Chapter 7. The research question of this chapter is, "How can an activity recognition system assist physiotherapists to monitor patients during the rehabilitation phase of someone that has undergone a lower body operation?" The goals of the component were to:

1. assist physiotherapists with their postoperative rehabilitation protocol,
2. adapt the predicted activities to the ones needed by the application, and
3. deliver a fully functional system comprising of a smartwatch application, a smartphone application, and a server-side application.

The author has developed the component after defining the requirements of it with professional physiotherapists of the Hirslanden Clinique La Colline, an orthopedic clinic in Geneva. During the experimental phase, different gait patterns of multiple patients were recorded, soon after the patients had undergone a lower body operation. Several machine learning models, including neural networks, were tested in order to find an optimal way to use the recorded data and to maximize the predictive accuracy. The author is describing the whole data flow that enables the component to operate and assist the work of physiotherapists during each patient's rehabilitation phase, even outside the time-limited physiotherapy sessions. This work has been published in [14] and an extended version in [15]. The contributions of the author in this study were the definition of the architecture of the online learning system, and the data analysis.

The author of this thesis is presenting a general summary of all the previous parts of this work in Chapter 8. All the innovative algorithms and methodologies presented in this thesis are discussed once more, highlighting their contributions to the current state-of-art. The author also discusses future insights in Chapter 9 for all the different applications proposed throughout this thesis.

The most important contributions of the thesis can be summarized as:

1. the proposal of a BLE based room-level indoor localization algorithm that improves accuracy in boundary locations,

2. a method to fuse heterogeneous data coming from different sensors of a smartphone to infer mood and stress levels,
3. an abnormal behavior detection component running on a tablet for remote monitoring applications,
4. the definition and classification of the different considerations in the design of an inertial sensor-based activity recognition system, and
5. the design of an online learning, gait recognition system to monitor patients during the rehabilitation phase.

2 User Requirement Analysis for the Design of an Ambient Assisted Living Application

2.1 Chapter Abstract

Most countries of the world are heading towards an aging society. At the same time, newer technologies are constantly being created, while the advances in networks and wireless communications allow other technologies like the internet of things, mobile and cloud computing to become ubiquitous. This leads to a problem that we are identifying and confronting, that is making the use of modern technology more accessible for older adults since it is in principle more easily perceivable by younger people. This chapter presents a questionnaire study that took place during the design of a gamified mobile application that targets people of older age. In total, 133 older adults answered the questionnaire consisting of 41 questions, providing an insightful view of their attitude towards modern technology, their health, physical activity tracking, playing games and social interaction using technology. The results of the questionnaire provide useful insights to researchers and developers who target this age group for their human-centric applications and services.

This work has been published as: *'User Requirement Analysis for the Design of a Gamified Ambient Assisted Living Application'*, Kyritsis, A.I., Nuss, J., Holding, L., Rogers, P., O'Connor, M., Kostopoulos, P., Suffield, M., Deriaz, M. and Konstantas, D., In International Conference on Computers Helping People with Special Needs (ICCHP 2018), Linz, Austria, July 2018.

2.2 Introduction and Related Work

We are heading towards an aging society, and the median age of the population has been rising during the last decades, due to declining fertility rates and rising of life expectancy [16]. Although the European population is currently the most aged around the world and is projected to remain so, every other continent has been experiencing the same demographic transition [17]. This shift is likely to be of major significance during the coming decades, transforming the age pyramid and leading to a much older population structure.

There is, however, an age technology gap and the aging population shift is expected to deteriorate it. Older adults often lack awareness of many technologies and are less likely to use them [18]. Elderlies are also less confident with new technologies in general [19]. They have different concerns and needs than younger technology users and may not be attracted by the latest technological advances [20].

Defining old age is not a straightforward task with a universal answer, but usually differs according to the context. Although most developed countries currently accept the age to 65 years as the cutoff to refer to the older population, this does not adapt well with African countries. The United Nations (UN) has agreed on denoting 60+ years people like older adults, while the World Health Organization (WHO) has set this cutoff to the age of 50 for studying Africa [21]. For our project, we are targeting people that are at least 55 years old.

Due to the aging society, there has been a rapid surge in the developed Ambient Assisted Living (AAL) tools and applications. These promote independent and safe living and aim to reduce the cost of healthcare as well as the caregiver burden [22]. AAL has emerged into a multi-disciplinary field that promotes the advancements of communication and information technologies to be used by older adults. Over the last years, there has been a plethora of AAL systems, platforms, frameworks, standards, and technologies [23]. They each usually have a different goal and aim to fulfill different needs of the end users [24]. However, all those technologies undeniably target to contribute to the overall wellbeing of older adults [25].

There have been many studies about the perception of technology by people and in our case, by older ones. In principle, the acceptance of a technology depends on its perceived usefulness, its ease of use, the attitude and the intention of the users to actually use it, as well as the actual usage of the system [26]. There has been an implicit assumption that information technologies are of great use through all sectors of society, but older adults prove to be more ambivalent towards using such technologies in their day-to-day lives [27].

The use of gamification that is the application of game design elements in non-game contexts proves to be a way to promote the use of technology, encourage specific behavior, improve the user experience and reduce the getting used time. There have been works that apply gamification techniques to improve older peoples' aspects of wellbeing, either by engaging the elderly in telemedicine [28], or by supporting their efforts to maintain good physical activity routines [29].

Designing an application for older adults is a demanding task. In principle, they encounter several constraints when dealing with computer-based technologies [30]. Their attitude towards technology and their learning rate differ when compared to younger adults, but with proper encouragement and explanations, elders prove to be equally effective [31] and may form and participate in online communities [32]. Specific design methodologies should be followed when developing applications that target older adults [33]. With our work, we are pursuing to comprehend how the elderly think about several aspects of technology, what are their needs and how their wellbeing can be enhanced when developing a platform that satisfies their requirements.

Our work was partially conducted to support the development of the EDLAH2 (Enhanced Daily Living and Health 2) [4] project. EDLAH2 is a European AAL project that plans to make the usage of smart technology easy and to promote wellbeing and health among older people. The goal of the project was to create an easy-to-use gamified tablet application that includes games, social, and health tracking features. The content and the features of the app can be remotely managed via a web interface. The questionnaire was developed to properly guide the development of this project and no pre-validated questionnaires were used.

The functionality of the app consists of an easy to manage photo library that includes photos sent by family members, integrated video/audio communication with Skype, an e-mail system, a web browser with an easy way to set bookmarks, calendar functionality with reminders and many tablet games. Moreover, the platform includes a way to record health data, such as weight, blood pressure, and blood sugar, a plan for health improvement including exercise and weight, and finally a wearable device that will monitor health parameters, such as the number of steps and the amount of sleep. This set of features is manageable, and the extracted data are visible to family members and others that have permission via the aforementioned web interface. Gamification techniques are also used in the application to increase the user's engagement. The research question of this chapter is, "What is the current view of people over 55 towards mobile technology, and how can it become more appealing?"

The rest of the chapter is organized as follows. In Section 2.3 we discuss our motivation for carrying out this user requirements study and its experimental setup is described in Section 2.4. In Section 2.5 we present and discuss the results that we have obtained, and we make the relevant recommendations. Finally, we conclude our work in Section 2.6.

2.3 Motivation

The goal of the EDLAH2 project was to develop a platform and a set of tablet applications that are easy to use by older adults. In order to do so, one should first understand the needs and requirements of the users of the platform, the elderly. This is the reason why this questionnaire took place, in order to understand the needs, the expectations and the attitude of the users towards physical activity tracking, playing games, gamification and social interaction using a tablet. By understanding how older people feel about technology, their health, and social inclusion, then it is possible to apply all the results in the design of the tablet application and gamify the whole experience. Therefore, the goal of the presented work is to understand our target group, in order to design the platform and further orient the development of our project that could motivate people to improve their wellbeing by using a tablet app.

2.4 Methodology

The consortium partners of the project were based in Switzerland, Luxembourg, and the United Kingdom (UK). All partners participated in the creation and the grouping of the questionnaire. The questions were developed in English so that all consortium partners could contribute with suggestions and thoughts. Each partner proposed questions trying to understand how the contribution for every aspect and feature of the project should be made. In total, 41 questions were formalized and categorized into seven thematic areas. The detailed questionnaire, along with the user responses, can be found in the Appendix. All questions were closed-ended ones with predefined responses.

The survey was carried out in Switzerland and the UK. The originally developed questionnaire was used in the UK, while a translated version of it in German was used in the German-speaking part of Switzerland. The survey was created and completed using the Drupal Webform module. The participants were recruited from the partners' existing user base and contacts, from end users of other projects and activities, from newsletters, e-mails, and phone calls in order to diversify the sample with previously unknown participants.

In total, 133 people completed the questionnaire (N = 133). Among them, 59% were Swiss, and 39% were British. There were also 2 Germans and 1 Italian. There was an almost equal amount of women and men that participated, with 68 women and 65 men. While the majority of the participants that is 66% were from 65 up to 79 years old, 16% were less than 65, and 17% were more than 79 years old.

Half of the interviewees (52%) have obtained higher education qualifications and a quarter of them (24%) have a work-based background. The majority of the participants (84%) live in their own home. As a matter of fact, all of the participants from Switzerland live in their own home, while all users living in a care village come from the United Kingdom. Most interviewees have no mobility impairments. At least three-quarters of the participants walk unaided, drive a vehicle, be it a car, motorbike or bicycle, and more than half of them (55%) use public transportation.

Most of the participants (71%) can still live on their own, without the need for any domestic support. Although memory failure problems increase with age, the majority of our sample claims to either not have any memory problems (20%), or to have some but not frequently (66%). Most participants (59%) do not suffer from either anxiety or depression, while 32% of them rarely do so. A quarter takes no medication, while the rest ought to take at least one type of drug.

2.5 Results and Recommendations

In this section, the results of the questionnaire are presented, and the most significant ones are discussed. The section consists of eight subsections, one for each of the seven thematic categories the questions belonged to and one last for compound results.

2.5.1 Games in General

It is very crucial for our project to understand the attitude of the users towards games in general. Considering that the goal of the project is to motivate people to use the tablet application and to improve their wellbeing by using gamification techniques, understanding their perception of games is essential before promoting healthier lifestyles with games.

Most of the participants (80%) play games regularly. Only 11% of them do not play at all, while 16% of them are playing daily. Classic board games like dominoes, card games, crosswords, scrabble and bingo, do not seem to be very popular among our target group since the distribution is skewed (towards the "several times a year" and "never" responses). 17% of the participants never play such board games.

The main reasons behind not playing games more frequently are lack of inspiration (30%), lack of time (27%), and lack of partners to play with (21%). On the other hand, for the people playing games, the main reasons for doing so are for fun (60%), for practicing the brain (56%) and for social inclusion (40%). Since our platform includes multiplayer features like a leaderboard, we have also asked about the users' experience with multiplayer games and about the type of such games they like. Most of the users (38%) are playing on their own, and 12% do not like playing multiplayer games. From the rest, 32% are playing games in pairs, 28% in a team and 27% with lots of other individuals.

2.5.2 Computer Games

Since one of the targets of our application was to create tablet games exciting and suitable for our target group, it is important to understand their opinion on them. Therefore this group of questions focuses on tablet and computer games.

It seems that just over half (54%) of the users would play video games either on a computer or a tablet. Comparing this result to the previous one of 80% of people playing games in general, we can already highlight the gap that exists between older adults and technology. 27% would prefer to play a single player game, while those that would happily play multiplayer games would prefer to do so with people they already know, either with friends (56%), or relatives (39%), or acquaintances (23%).

More than two thirds (69%) are not interested in competing against others and only a quarter of them (26%) would be interested in getting a prize from the game they play on the tablet. Regarding the prize, most (43%) seem to be indifferent to it, while the ones that would happily earn one would prefer something physical, either cash or discount vouchers (26%), or a free cup of tea (13%), instead of some digital reward, since every such reward was ranging between 2% and 8%.

2.5.3 Competition and Consumer Behavior

Regarding the questions used to understand the attitude of the users towards competition, on top of a previous question about competing other people on a tablet game, we are trying to understand their point of view regarding more aspects that involve competition. We also try to understand their consumer behavior and how it can be influenced when winning some prize is an incentive.

It seems that consumer needs (53%) are considered to be the most important driving force behind peoples' shopping decisions, with 39% stating that collecting points on loyalty cards also motivates them. Most users (79%) often receive some kind of points or use loyalty cards during shopping, but do not necessarily choose where to shop or what to buy based on loyalty rewards.

Three-quarters of our sample rarely or never play the lottery or participate in competitions. On the contrary, there is an almost uniform distribution among people that frequently watch competitive sports live or on TV and among those that never do so. Regarding the attitude of the interviewees towards winning, participation seems to be the main goal of the game for half of the interviewees, while there is a fair amount of them that strictly do not want to compete (17%).

2.5.4 Social Background

Another goal of the application is to enhance the social life of its users. Elderly will be able to receive e-mails, photos, and phone/video calls within our platform. Hence the importance of understanding their social background first.

The majority of the participants (82%) think that social contact is important, but not all of them can imagine engaging with more people using technology since only 40% think that it is possible to do so. Three-quarters of them (77%) have regular contact with their family through phone calls and visits at most on a weekly basis, with 24% of them on a daily basis. Only a minority of them do not go out that often (7%) and do not see people often (11%). All the rest do so at most every week.

Providing that we are creating a platform that monitors and collects data about different aspects of behavior and wellbeing, it is of crucial importance for the end users to know how their data are being treated and who is able to view them. Given that the users know exactly who and which data, permission would have been granted by more than half of the interviewees to certain members of the family (56%), as well as to certain healthcare professionals (59%). Around 24% would not share their data with others. There was, however, a respectable number of participants that were not entirely sure what data sharing means, so app developers should give attention and careful clarifications.

2.5.5 Health

A big part of the project is about proposing ways to improve one's health through gamification. Several health-related metrics are monitored and shared with predefined members of the family or healthcare professionals. Health tracking devices are also planned to be utilized by the platform.

It is clear that most elderly would like to improve their fitness level (83%) and would enjoy mental exercise (84%). It is not clear though, whether they would like some encouragement from others in order to do so, with 53% of them liking this idea and 42% not.

Regarding wearing a measuring device, the results are again split, with 54% interested in trying to use one and 43% not. The blood pressure monitor (45%), the clip-on pedometer (41%), the panic alarm (35%), and the steps measuring bracelet (32%) seem to be the most popular options.

Our target group would not freely share health-related information with certain family members, with 52% of them voting for this. 38% would like to be able to see a health report on the tablet, and 35% would not. It is noteworthy that 20% do not really know if they would be interested in this feature, a fact that we can most probably attribute to the technology-generation gap.

Regarding the most appealing aspects of health improvement training, getting a clear goal on what should be done (36%), and training with other people (29%), i.e., enhancing their social life, seem to be the most popular ones. Notably, only one participant would be interested in being able to compete with others, and only 9% of them would be interested in receiving some achievement for every successful exercise. These last results, as well as the significant number of people not replying to this question (25%), possibly demonstrate once more that the elderly are not up-to-date with the current trends in technology.

2.5.6 Local Opportunities and Volunteering

The platform also includes a way to volunteer locally with skills. 45% of the participants would like to see local volunteering opportunities while less of those (35%) would like to see those on a tablet. The majority of the participants (54%), however, would hesitate to share their skill set with local organizations through the tablet app.

During this phase of the questionnaire, many participants explicitly told us that they had been already doing voluntary work, but they were not particularly interested in using the tablet to share or see such opportunities.

2.5.7 Technology Acceptance

The target application runs on a tablet and addresses many aspects, some even new-found for our target group. As previously discussed, there is a technology-age gap, and thus, it is essential to understand the older adults' view on several aspects of technology, understand their previous experience, and evaluate their technology acceptance.

More than three quarters (76%) have an internet connection, either at home or a data plan on their phone and often use a computer, tablet, or smartphone. 18% of them do not have an internet connection, and 15% do not own any device mentioned above. Most of the people already not using a tablet would be interested in learning to do so, either by exploring and learning it themselves (35%), or with some practical help (18%), or by written instructions (15%). People that do not use technology, either do not understand it (6%), or do not see the need (3%), or do not trust technological devices (2%). This target group is an undoubtedly challenging one to be convinced to use any new means of technology.

The most prevalent reasons for using technology is staying in touch with people (75%), staying up-to-date with the world (72%) by reading the news or checking the weather forecast, learning new things (50%), and maintaining a hobby (37%). Regarding data security, some people are not worried at all (30%) since they show trust in the platform they use, while more are worried, but would happily use the platform (40%) or would first ask family advice beforehand (16%).

2.5.8 Compound Results

Among the elderly, the target group of our application, there seem to be different age subgroups in terms of familiarity and interest towards technology. In Figure 2.1, we can notice that as the age goes down, the frequency of using a computer or a mobile device rises. This is why it is important to motivate users of all ages to overcome the

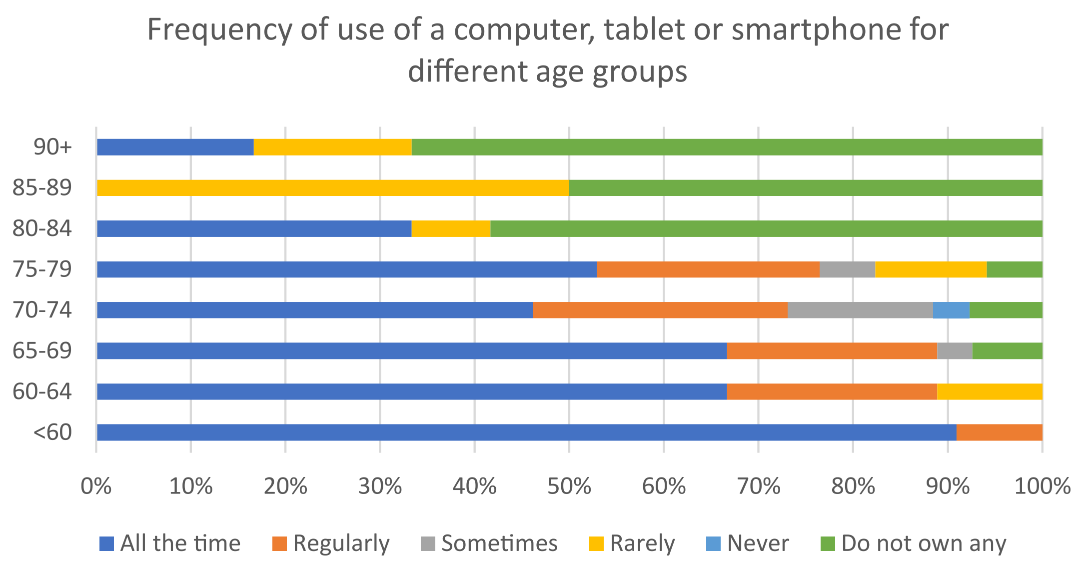

Figure 2.1: Frequency of use of computer, tablet or smartphone for different age groups.

initial fear of using technology and then to communicate all possibilities that open correctly.

Similarly, as the age rises, there seems to be less interest in learning to use modern portable devices, as seen in Figure 2.2. Figure 2.3 implies that people already familiar with computer, tablets, or smartphones are naturally more willing to try using new types of devices, such as wearable ones.

There is a correlation between the frequency of playing games and the interest of a person to play video games on a tablet. The more often people play some game, the more interest they are likely to have in playing tablet games as seen in Figure 2.4. From Figure 2.5 it also seems that social people, at least those that have regular contact with family members, are willing to use technology as a means of socializing, in case they are not already doing so.

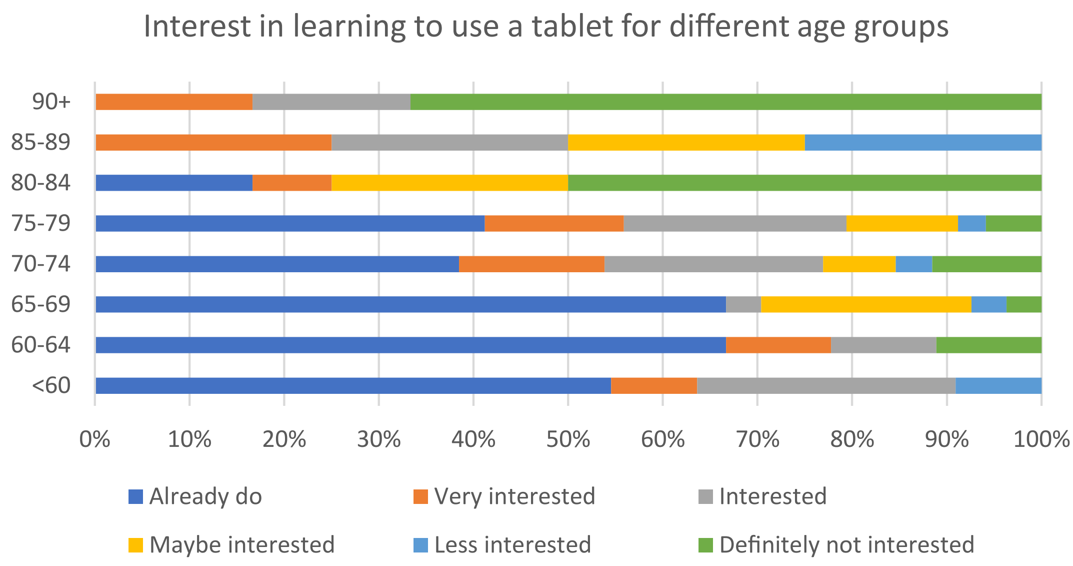

Figure 2.2: Interest in learning to use a tablet for different age groups.

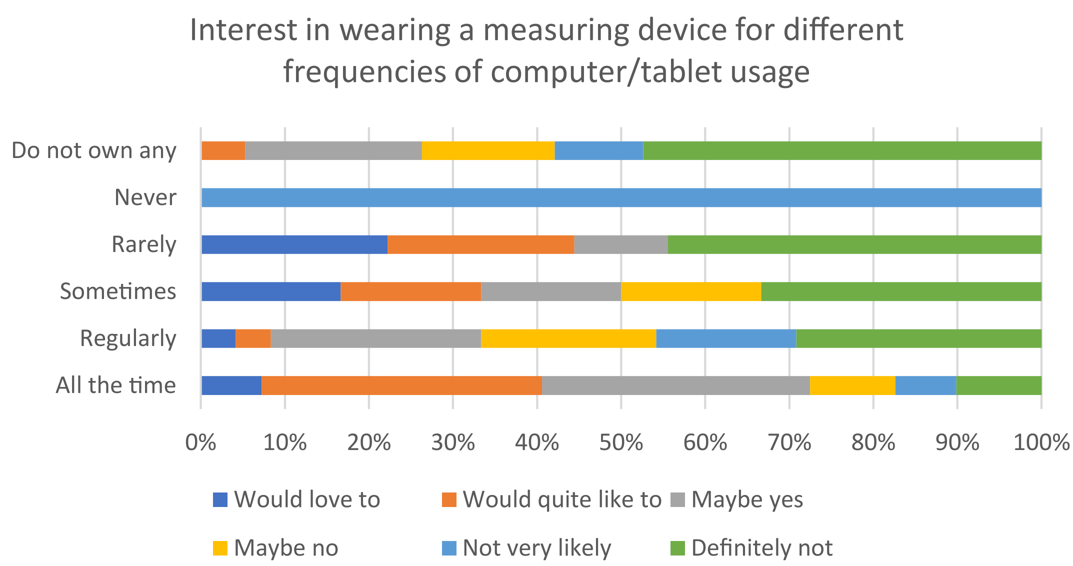

Figure 2.3: Interest in wearing a measuring device for different frequencies of computer/tablet usage.

2.6 Conclusion

Our work provides an insight into how older adults perceive technologies, health monitoring and tracking, games, and social interaction, in support of researchers

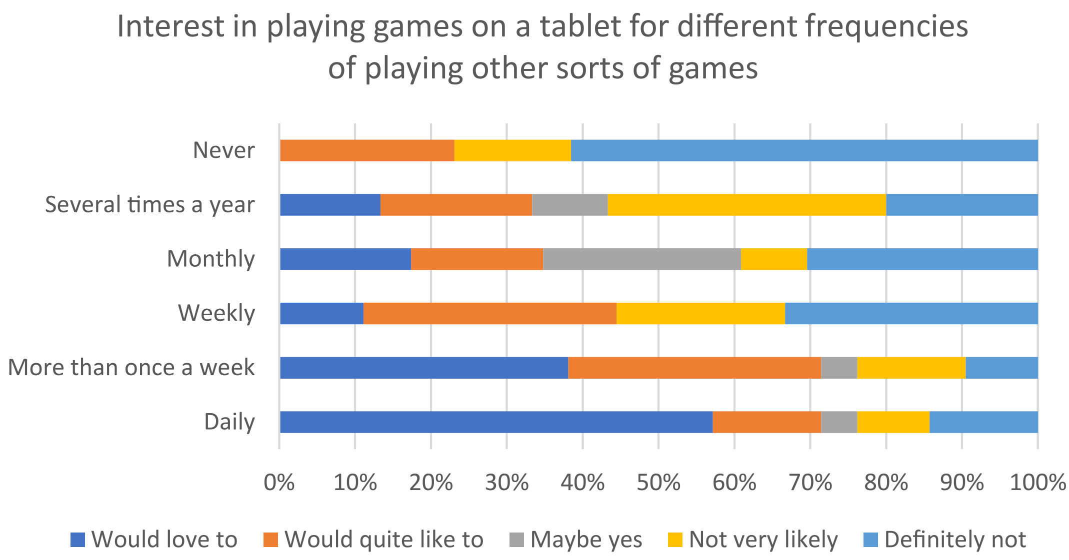

Figure 2.4: Interest in playing games on a tablet for different frequencies of playing other sorts of games.

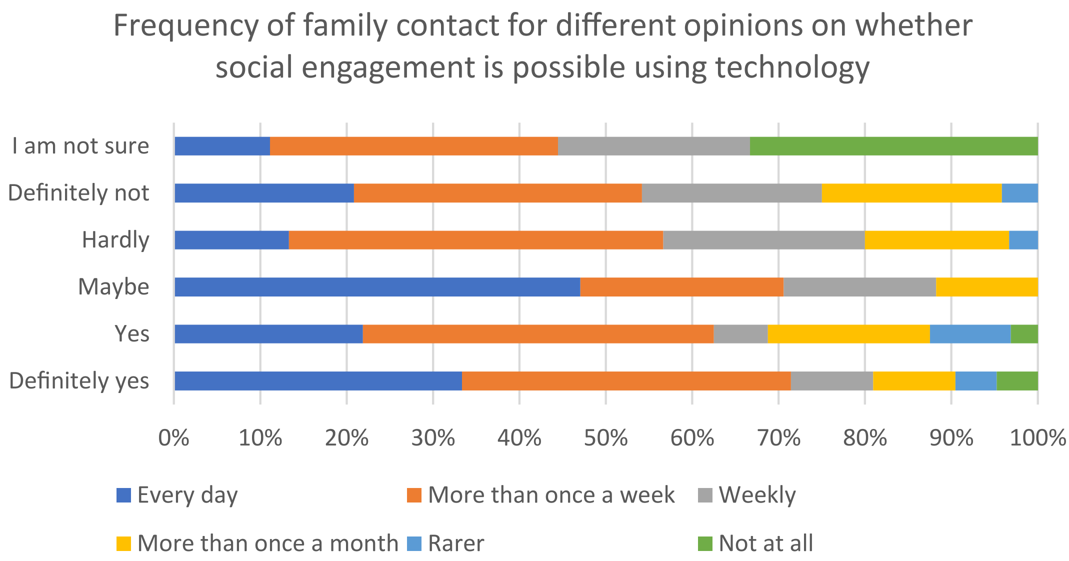

Figure 2.5: Frequency of family contact for different opinions on whether social engagement is possible using technology.

and developers who design applications for this target group. Designers should create human-computer interactions that allow seniors to remain active members of their family and society while pursuing their interests. In general, older adults seem to be interested in exploring technology, although sometimes they are unaware of its capabilities. This is why it is crucial to approach older adults with proper encouragement and clear explanations when presenting new applications to them.

There are, however, limitations of the presented work. Some of the questions in the questionnaire may be considered biased. For example, the question "Why don't you play games more frequently?" is biased, and the corresponding responses may not apply to everyone. There is also a considerable amount of no responses in some questions, with the highest percentage of 52.2% in the "What is the most prevalent reason for not using a smartphone, tablet, or computer?" question. This can be attributed to either the lack of a proper reply among the available responses or the misuse of the survey software that would not warn when there were no responses to some questions. The questionnaire was also not based on validated questionnaires but was developed to guide the development of a tablet application properly.

3 Indoor Room-Level Localization Using Bluetooth Low Energy Devices

3.1 Chapter Abstract

During the last decades, location-based services have become very popular. Global Navigation Satellite Systems (GNSS), such as the Global Positioning System (GPS) and GLONASS, are used for outdoor positioning and navigation, but are not suitable for indoor positioning, mostly because walls and roofs attenuate the microwave signals. For this reason, different technologies such as Wi-Fi access points and Bluetooth beacons are used for indoor localization. Currently, there are several available commercial systems, and there is no standard for an Indoor Positioning System (IPS). Despite that, those systems get better and better over time, and they have achieved an impressive accuracy.

The problem though, is that even if the only requirement of an IPS is room-level localization, those systems are most of the times not cost-efficient and not easy to set-up since they often require time-consuming calibration procedures. This chapter presents a low-cost, threshold-based approach and introduces an algorithm that takes into account both the Received Signal Strength Indication (RSSI) of the Bluetooth Low Energy (BLE) beacons and the geometry of the rooms the beacons are

This work has been published as: *'A BLE-Based Probabilistic Room-Level Localization Method'*, Kyritsis, A.I., Kostopoulos, P., Deriaz, M. and Konstantas, D., In International Conference On Localization and GNSS (ICL-GNSS 2016), Barcelona, Spain, June 2016.

placed in. Performance evaluation was done via measurements in an office environment composed of three rooms and in a house environment composed of six rooms. The experimental results show an improved accuracy in room detection when using the proposed algorithm, compared to when only considering the RSSI readings. This method was developed to provide context awareness to the European AAL research project SmartHeat and the Swiss national Innosuisse project F2D. However, the proposed IPS can be used in any modern application that requires indoor localization contextual information.

3.2 Introduction and Related Work

In the past two decades, there has been a continuous rise in interest in location-aware applications. After the invention of the Global Positioning System (GPS) [34], more and more devices have included a GPS receiver and have been using this technology. Especially with the rise of the smartphones, Global Navigation Satellite System (GNSS) receivers have become available in the market at low cost, and are nowadays ubiquitous. At the time of writing this thesis, USA's GPS, Russia's GLONASS, and China's BeiDou Navigation Satellite System (BDS) are operational GNSS providing global coverage. European Union's Galileo is scheduled to be operational by 2020. While a GNSS is an exemplary solution for most outdoor applications, it is not suitable for indoor environments. Therefore, new technologies and systems have been invented that can be used for indoor localization.

One common category of such systems is that of the inertial ones, namely those that use an inertial measurement unit tracking technique, such as the pedestrian dead reckoning [35]. Sound-based systems also exist using, for example, ultrasound anchor beacons with known positions [36]. There are also systems that use other spatially dependent environmental properties such as magnetic fields, visual object recognition, and light. Last but not least, there are hybrid systems that are implementing multiple technologies [37], or that are using multimodal sensing [38].

However, the most widespread indoor localization technique is by using radio transmissions. Methods that use radio signals include Wi-Fi devices that are popular and widely deployed and Bluetooth beacons that are low-cost [39]. Those systems either estimate the distance between the transmitter and the receiver by employing path-loss models or employ location fingerprinting to infer a position. The measured radio signal quantities typically include the link quality, the Time Of Arrival (TOA), the angle of arrival, the time difference of arrival, the Signal-to-Noise Ratio (SNR), and the RSSI. The RSSI is the relative received signal strength in a wireless environment, in arbitrary units. RSSI is an indication of the power level being received by the antenna and therefore, the higher the RSSI number, the stronger the signal.

Our approach to indoor localization is based on the use of BLE beacons, using the RSSI value, since it is available in all standard wireless communication devices. The important feature of our approach that distinguishes it from other systems based on Bluetooth is that it does not only rely on radio signal quantities. It also takes into account the geometry of the rooms the beacons are placed in, i.e., the height and the surface area. This information, therefore, becomes a requirement of the indoor localization system we are proposing. Since the requirements of the localization system to be developed were minimal cost and easy setup process for the end user, we used the minimum amount of BLE beacons, that is one BLE beacon per room attached to the ceiling in the center of it, and we opted to develop a more sophisticated algorithm for room detection.

In ubiquitous computing, the need for location information is critical, and several context-aware applications are in need of an indoor positioning system for room localization. Our motivation in developing this indoor localization system with room-level accuracy is to provide contextual information to the international research project named SmartHeat [3]. The project aims to provide a system that efficiently heats a house, room by room, based on the users' habits and preferences. It will also be used to provide location information to the F2D fall detection system [40], as a way to provide the system with context awareness in order to improve its accuracy as well as the reaction time of the user's carers. The same solution, though, can be easily deployed and used for any application that requires room-level localization.

Ideally, in line-of-sight conditions, the performance of such a system can be accurate. On the other hand, the radio frequency signals indoors are prone to disturbances due to shadowing, fading, the multipath propagation phenomenon, and device imperfections. These can lead to significant errors when estimating distances based on the radio signal quantities since the signal can significantly fluctuate. These errors can be confronted by not using the newest reading of the signal quantity exclusively, but by averaging a set of the latest ones [41].

Location information is essential for a wide range of ubiquitous and pervasive computing applications. This is the reason why the topic of determining the position of a device has been the subject of many studies. In this section, we give an overview of some existing systems and implementations that use Bluetooth as well as other technologies as a means to achieve room-level localization. All different implementations have had to balance the technologies used in terms of cost, precision, accuracy, portability, ease of installation, deployment, and use.

One of the first indoor badge positioning systems is the Active Badge system [42]. Active badges were used to emit a globally unique infrared signal and were carried by people. Receiver sensors were placed in each located place such as a room, in order to detect the signals sent by the active badges and to infer a position for each badge. Although the sensors and the badges were cheap, the sensors had to be connected to a central server, and the cables raised the cost of the system, despite the room-level accuracy that it provided. The use of a central server was also not suitable for our application.

Another way of indoor localization is by using ultrasound signals. Inspired by bats that use those signals to navigate at night, several such systems have been developed. The Active Bat positioning system [43] is using tags that periodically broadcast a short pulse of ultrasound. Ceiling mounted receivers at known positions receive the pulses mentioned above. Using the TOA radio signal quantity and the trilateration method, a 3D position for every tag can be calculated. Generally, the performance of the ultrasound technology is hindered by reflections and by obstacles between receivers and transmitters. Although the system has achieved impressive accuracy in positioning, the use of a large number of receivers by the Active Bat and the interconnection between them, limit the scalability of the system.

Conversely, the Cricket indoor localization system [36] uses ultrasound emitters attached on the walls or ceilings at known positions and receivers attached to objects to be located. The system uses TOA again and the trilateration localization technique to infer a location. On top of this, radio frequency signals are used for the synchronization of TOA and for proximity positioning to address fault tolerance issues. Although the system was not targeting room-level accuracy, fewer ultrasound emitters can be used to achieve this, leading to a proportional decrease in both cost and accuracy. The problem with this approach, though, is that both the transmitters and the receivers need more power since they have to handle both ultrasound and radio frequency signals at the same time.

A large body of indoor localization approaches uses the Wi-Fi technology, as to take advantage of the vast availability of wireless access points in urban areas. The RADAR system [44] uses the existing WLAN technology and employs RSSI and SNR with the triangulation localization technique. Another system named WILL [45] also uses the existing Wi-Fi infrastructure and mobile phones to localize the user indoors. On this occasion site survey is not needed, and thus the deployment is easy and rapid. Although Wi-Fi positioning is one of the most popular indoor positioning techniques, most of the times the Wi-Fi access points are not deployed with the ideal geometry and density for positioning, and thus are not optimized for indoor localization.

Another wireless sensor network based indoor location estimation system uses the ZigBee communication standard for room detection [46]. It considers the behavior of the RSSI through walls, floors, and ceilings, and using a decision algorithm estimates a position. A blind node is located by using the reference nodes, that are placed one per room. The system exhibits good performance for its simplicity, although a wrong room indication often occurs when the blind node is located in the vicinity of a wall, due to the unpredictable indoor multipath effects and the potentially small path loss through the intersecting material. The boundary locations were also the biggest challenge that we faced in our approach.

The use of Bluetooth technology for positioning has been evaluated more than a decade ago [47]. Since then, the introduction of the BLE radio protocol provided even more opportunities for indoor localization. BLE beacons are flexible in the sense that they are small in size, plus they do not need to be plugged in and are power efficient. Either deriving a location from fingerprinting techniques [48], or ranging techniques that use path-loss models [37], researchers have focused on increasing the accuracy of the positioning. Although a room estimation can often be derived from such systems, they are usually not optimized for it. Our research has focused on developing an easy to set up BLE-based system for room localization while keeping the cost as minimal as possible. The research question of this chapter is, "How can room-level indoor localization be effectively achieved by using one Bluetooth beacon per room?"

The rest of this chapter is organized as follows. In Section 3.3 we present the system we designed, and we provide the necessary background. We experimentally evaluated the performance of the proposed system in an office and a house environment, and the results are presented in Section 3.4. Finally, our conclusions are drawn in Section 3.5.

3.3 System Overview

3.3.1 RSSI and Propagation Model

In RSSI-based localization, the signal sent from the anchor beacon to the mobile device is used to map the RSSI to a distance by means of a propagation model. The correct calibration of the propagation model is crucial since the way RSSI is transformed into a distance significantly affects the accuracy of the positioning. The widely known method we use to model wireless signal propagation loss [49] is expressed as:

$$r = r_0 - 10nlog_{10}(\frac{d}{d_0}) + X_\sigma \tag{3.1}$$

where r and r_0 denote the received signal power at the real distance d and at a reference distance d_0 respectively. X_σ is a random variable representing the noise in the measured r and n is the path loss exponent, that depends on the transmission channel, the transmitter, and the receiver. Using $d_0 = 1$ meter as the reference distance, and assuming X_σ to be a zero-mean Gaussian distribution, the simplified model is used as follows:

$$r = p - 10nlog_{10}(d) \tag{3.2}$$

where r is the received signal power at a distance d, p is the received signal power of the receiver from a transmitter one meter away, and n is again the path loss exponent.

3.3.2 Room Dimensions and RSSI Thresholds

Let S be the surface area of a room and h be the height of it. Assuming a square room as on Figure 3.1, the radii of the inner and the outer tangent circles are calculated with Equations 3.3 and 3.4 respectively.

$$r_{in} = \frac{\sqrt{S}}{2} \tag{3.3}$$

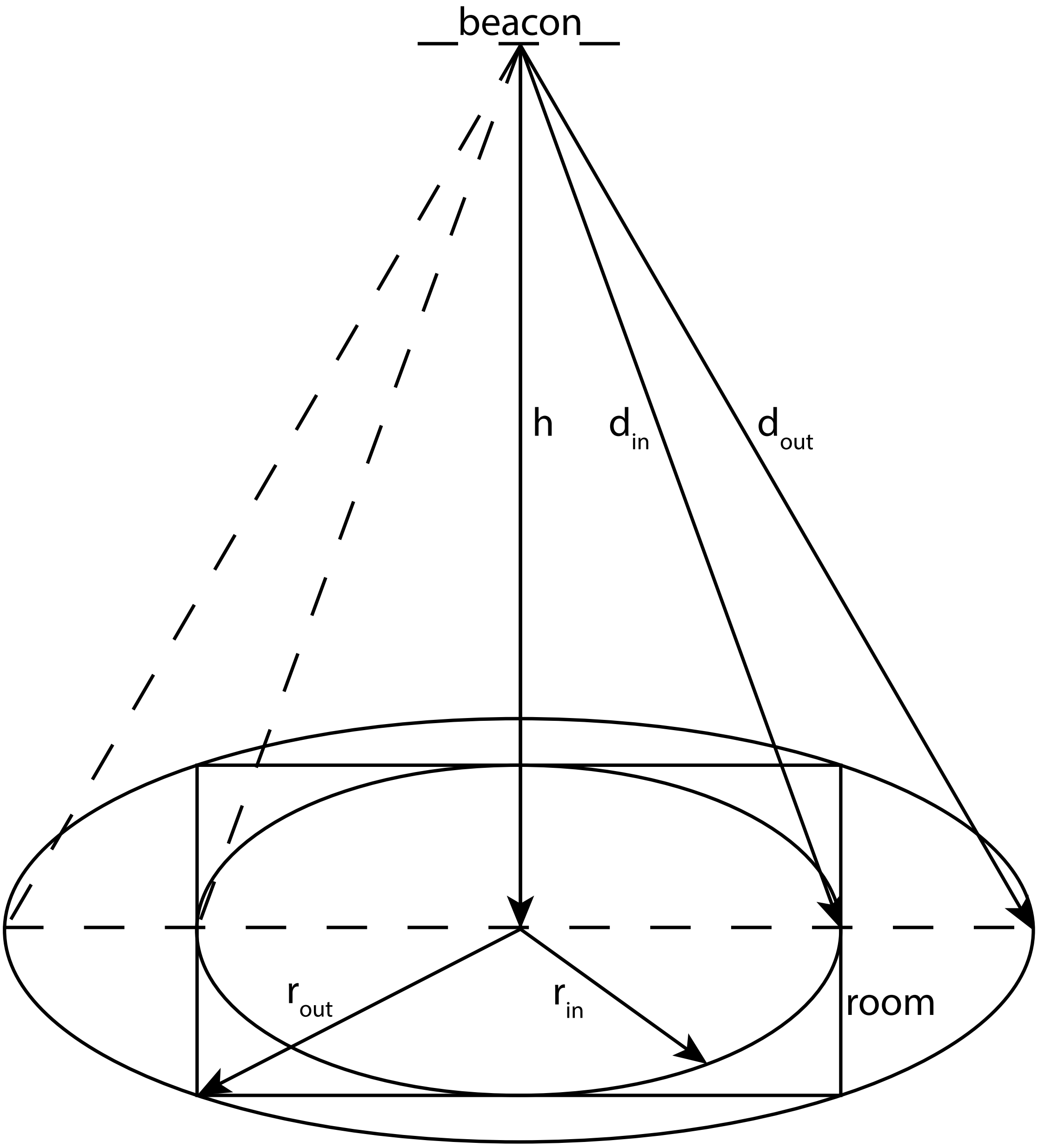

Figure 3.1: Dimensions in a square room.

$$r_{out} = \sqrt{\frac{S}{2}} \tag{3.4}$$

Now using the Pythagorean theorem, the hypotenuses are calculated with Equations 3.5 and 3.6 respectively.

$$d_{in} = \sqrt{h^2 + \frac{S}{4}} \tag{3.5}$$

$$d_{out} = \sqrt{h^2 + \frac{S}{2}} \tag{3.6}$$

Eventually, by substituting the calculated distances of the hypotenuses into the propagation model of Equation 3.2, the expected RSSI values at those distances are obtained. Defining the inner and the outer RSSI thresholds as the expected RSSI values at the inner and the outer tangent circles of the aforementioned square room, respectively, where there is a line of sight to the beacon and no interference. The thresholds are calculated with Equations 3.7 and 3.8.

$$threshold_{in} = p - 10nlog_{10}(\sqrt{h^2 + \frac{S}{4}}) \tag{3.7}$$

$$threshold_{out} = p - 10nlog_{10}(\sqrt{h^2 + \frac{S}{2}}) \tag{3.8}$$

3.3.3 RSSI Classification and Localization Algorithm

In our approach, one BLE beacon is attached to the ceiling in the center of every room. For every beacon, the $threshold_{in}$ and $threshold_{out}$ are calculated as described previously, taking into account the dimensions of the rooms. Based on the RSSI readings of the beacons, they fall into one of the following categories:

- The ”Strong” category (S), when $RSSI > threshold_{in}$
- The ”Medium” category (M), when $threshold_{in} > RSSI > threshold_{out}$
- The ”Weak” category (W), when $threshold_{out} > RSSI$
- The ”Not found” category (NF), when there is no reading for a specific beacon

The ordering of those categories based on their significance is given by Equation 3.9.

$$\mathrm{S} > \mathrm{M} > \mathrm{W} > \mathrm{NF} \tag{3.9}$$

At any given moment, for every beacon, a set of its N latest RSSI readings is averaged, before every beacon can be classified into one of the categories mentioned above. While a relatively large N may increase the robustness of the system, in the sense that it copes with noise and signal variations, this particular choice may introduce lag to the localization method, because old readings are taken into account for a longer period of time. This can be confronted, though, by giving more weight to the latest readings than to the old ones.

After the classification of the beacons, the most significant non-empty category is picked. If only one beacon falls into this category, then the procedure ends, and the presence is assumed in the room in which this specific beacon was placed. When multiple beacons fall into this category, then a score is calculated for every beacon that is equal to the difference of its measured RSSI and its lower threshold. The lower threshold is equal to $threshold_{in}$ when S is the most significant non-empty category, $threshold_{out}$ when M is the most significant non-empty category, and when W is the most significant non empty category, it is the global minimum of the RSSI readings

Table 3.1: Respective lower threshold for each category.

Category	Requirement	Lower threshold
Strong	$RSSI > threshold_{in}$	$threshold_{in}$
Medium	$threshold_{in} > RSSI > threshold_{out}$	$threshold_{out}$
Weak	$threshold_{out} > RSSI$	-127
Not found	no RSSI reading	N/A

set by the BLE specifications, which is -127 [50]. The corresponding lower threshold for every category along with the requirements for the beacons to fall into each one of them are given in Table 3.1.

Then, a probability is calculated for every beacon as the fraction of its score by the sum of all the scores of the beacons that fall into the aforementioned most significant non-empty category. In case a single room estimation is needed, the beacon with the highest probability dominates and in the final case of a draw, the beacon that is placed in the biggest room does. The whole procedure is presented in pseudo-code in Algorithm 1 and is depicted in Figure 3.2.

The way the algorithm is designed, it intrinsically favors presence in bigger rooms, in the sense that when the same RSSI is received by two beacons, the one in the bigger room will have lower thresholds and will either be in a higher category, or if not it will have a higher probability than the beacon in the smaller room. For this reason, the algorithm works best for the boundary locations in a room that are farther away from the center of the room, than the center of a smaller adjacent one.

3.4 Experiments and Evaluation

3.4.1 Experiment Methodology

For our experiments, we used a Samsung Galaxy S6 (the SM-G920F international variant) smartphone as the receiver and the Kontakt.io Smart Beacons, as seen in Figure 3.3, as the transmitters, set in their default configuration settings (transmission power = $-12dBm$ and interval between transmissions = $350ms$).

Algorithm 1 Room localization algorithm.

```
Initialize a pair of thresholds, threshold_in and threshold_out, for every beacon
according to the room it is in
while algorithm in use do
    for each beacon i do
        get recording_i after the BLE scanning
        if beacon i not found then
            add beacon i to the NF category
        else if recording_i > threshold_in then
            add beacon i to the S category
        else if recording_i > threshold_out then
            add beacon i to the M category
        else
            add beacon i to the W category
        end if
    end for
    if category S is non empty then
        highest_category ← S
        lower_threshold ← threshold_in
    else if category M is non empty then
        highest_category ← M
        lower_threshold ← threshold_out
    else if category W is non empty then
        highest_category ← W
        lower_threshold ← minimum_possible_BLE_value
    else
        highest_category ← NF
    end if
    if highest_category = NF then
        unknown location, no known beacon around
        continue with the next iteration
    else
        sum_scores ← 0
        for each beacon i falling into highest_category do
            score_i ← recording_i − lower_threshold
            sum_scores ← sum_scores + score_i
        end for
        for each beacon i falling into highest_category do
            probability_i ← score_i / sum_scores
        end for
    end if
end while
```

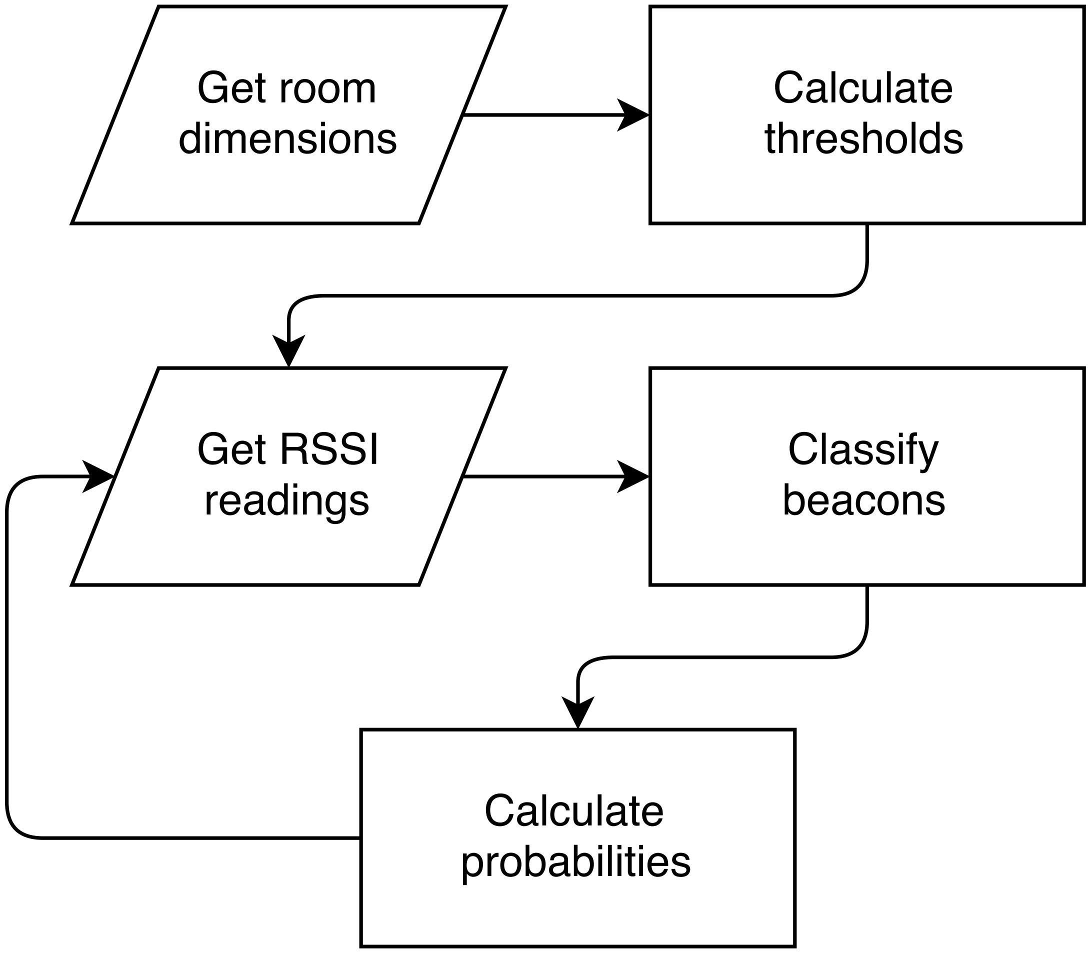

Figure 3.2: Steps of the localization method.

We gathered RSSI readings at grid locations in every room throughout the area. At every point, we collected a total of 200 averaged RSSI readings for every beacon. For our tests, we have empirically set $N = 10$, where N is the size of the set of the latest RSSI readings of every beacon that is averaged. The receiver was placed on a non-conducting surface at roughly one meter from the floor.

3.4.2 Propagation Model Calibration

In order to construct the specific propagation model for our application, we placed a BLE beacon in the center of a corridor. Then, we took multiple measurements at several points with a known distance from the beacon, ranging from 0.5 to 7 meters.

Figure 3.3: Kontakt.io BLE beacons used for indoor positioning.

By constructing the line of best fit described by Equation 3.2, the estimated values of the propagation model parameters were $p = -70.09$ and $n = 1.95$. The calibrated propagation model is presented in Figure 3.4.

3.4.3 Deployment in Two Locations

We have deployed our indoor positioning system in two different environments. The first is a typical office environment composed of three rooms, as seen in Figure 3.5, housing eight people. The area is divided by thick concrete walls and wooden doors. The second is a house environment composed of five rooms of different sizes and one corridor, as seen in Figure 3.7. Due to the corridor being oblong in this occasion, we have divided its total area into two equal ones, so that we can abstractly consider that

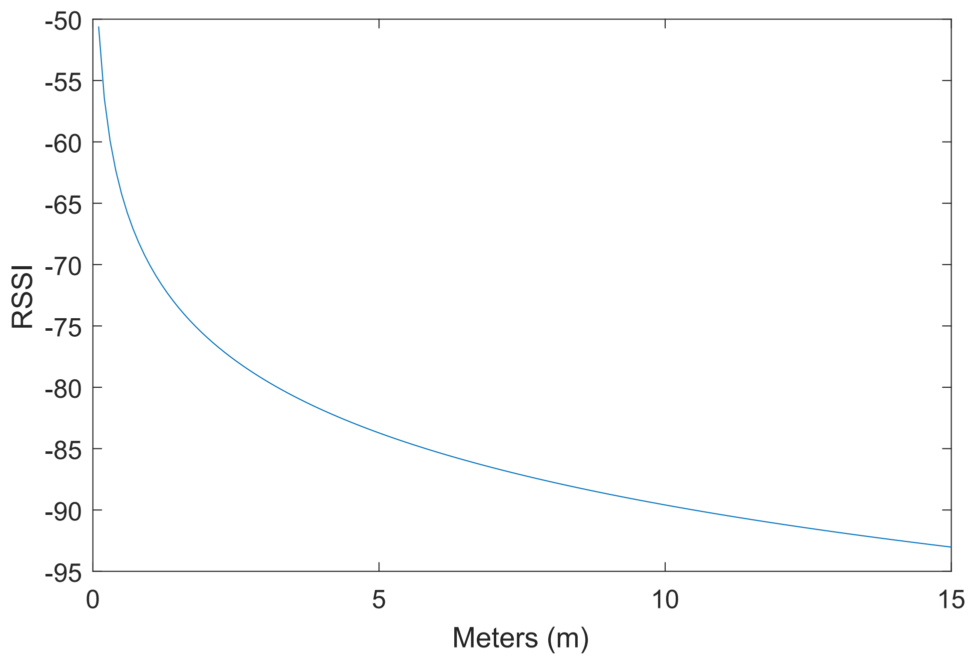

Figure 3.4: Calibrated propagation model.

the house is composed of a total of seven rooms. The area is divided by thin concrete walls and wooden doors.

3.4.4 Comparison

We compare the performance of our indoor positioning system with room-level accuracy with the one without the thresholds our algorithm introduces. That is a naive system that only considers the magnitude of the RSSI readings and assumes presence in the room with the highest one. Due to the result of the naive system being discrete, only the room with the highest probability given by our system is taken into account. The same recordings dataset was used for the comparison. The error of a point in a room corresponds to the percentage of incorrect classifications from 200 recordings taken at that point. The error of a room corresponds to the percentage of incorrect classifications from all recordings taken from all points in that room (9 points x 200 recordings per point = 1800 recordings per room).

Table 3.2: Per room error comparison in the office.

Room	Error without the algorithm (%)	Error with the algorithm (%)	Relative error improvement (%)	Fisher exact test statistic value
A	0	0	0	1
B	10.06	8.17	+18.79	0.0558
C	12.06	8.33	+30.88	0.0003

We are assessing the statistical significance of the results using Fisher's exact test [51]. The null hypothesis for our experiment is that the introduction of our algorithm will not improve room localization. Therefore, the alternative hypothesis is that our algorithm will improve room localization. We are using a significance level of $a = 0.05$.

3.4.5 Office Environment

In this experiment, RSSI readings were collected at 27 different points (9 for every room) as depicted by the circles in Figure 3.5. The green points are the ones for which the error was improved with the introduction of the localization algorithm, while for the red one the error deteriorated. Table 3.2 presents the average error per room and Table 3.3 presents the specific locations in the office for which the error has changed. For the rest of the points that the error remained unchanged (white points), the average error was 0.29%. As seen in Figure 3.6, the average error of the points of room B has improved by 18.79% and the average error of the points of room C has improved by 30.88%.

Using an alpha of 0.05 as the cutoff for significance, on a room level analysis, the improvement that our algorithm brings is statistically significant in room C of the office environment. On a point level analysis, points 13 and 19 had a statistically significant result.

3.4.6 House Environment

In this experiment, RSSI readings were collected at 63 different points (9 for every room) as depicted by the circles in Figure 3.7. Once more, the green points are the ones

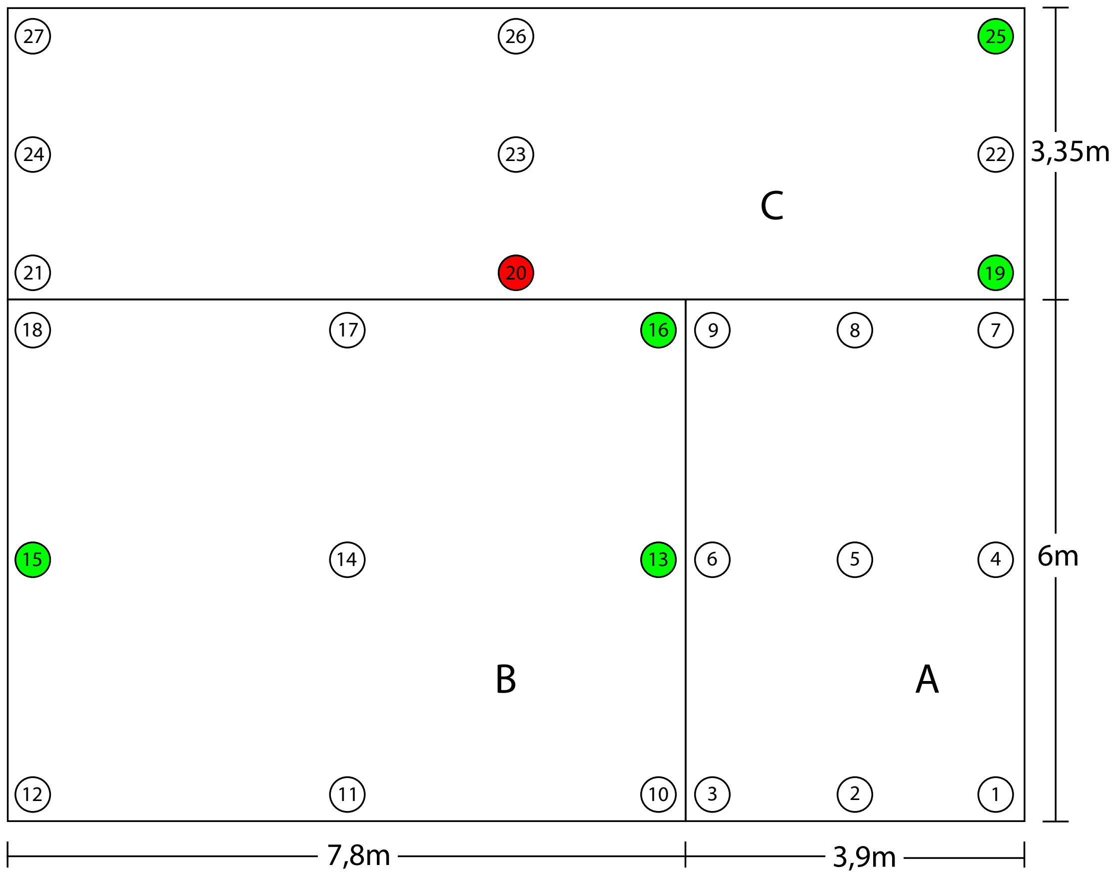

Figure 3.5: Office evaluation area and targeted locations (green for the points with an error improvement, red for an error deterioration).

for which the error was improved with the introduction of the localization algorithm, while for the red ones the error deteriorated. Table 3.4 presents the average error per room and Table 3.5 presents the specific locations in the house for which the error has changed. For the rest of the points that the error remained unchanged (white points), the average error was 6.07%. As seen in Figure 3.8, the average error of the points of room A has improved by 8%, of room B by 18.91%, of room E by 13.52%, of room G by 8.9%, while the average accuracy of the points of room D has deteriorated by 9.09%.

Using an alpha of 0.05 as the cutoff for significance, on a room level analysis, the improvement that our algorithm brings is statistically significant in room E of the house environment. On a point level analysis, points 16, 43, 44, and 62 had a statistically significant result.

Table 3.3: Locations in the office with an error change.

Point	Error without the algorithm (%)	Error with the algorithm (%)	Relative error improvement (%)	Fisher exact test statistic value
13	53.5	40	+25.23	0.0091
15	11.5	10.5	+8.7	0.8732
16	19.5	17	+12.82	0.6048
19	86.5	50	+42.2	<0.00001
20	21.5	25	-16.28	0.4777
25	0.5	0	+100	1

Table 3.4: Per room error comparison in the house.

Room	Error without the algorithm (%)	Error with the algorithm (%)	Relative error improvement (%)	Fisher exact test statistic value
A	9.72	8.94	+8	0.4564
B	8.22	6.67	+18.91	0.0863
C	0	0	0	1
D	1.83	2	-9.09	0.8081
E	21.78	18.83	+13.52	0.0312
F	15.33	15.33	0	1
G	18.72	17.06	+8.9	0.2072

Table 3.5: Locations in the house with an error change.

Point	Error without the algorithm (%)	Error with the algorithm (%)	Relative error improvement (%)	Fisher exact test statistic value
2	2.5	2	+20	1
3	25	24.5	+2	1
9	29.5	23.5	+20.34	0.2126
11	9.5	10	-5.26	1
16	54	39.5	+26.85	0.005
32	7.5	8.5	-13.33	0.8541
34	0.5	1	-100	1
40	46.5	44.5	+4.3	0.7633
43	45	24	+46.67	0
44	100	96.5	+3.5	0.0148
62	45	30	+33.3	0.0027

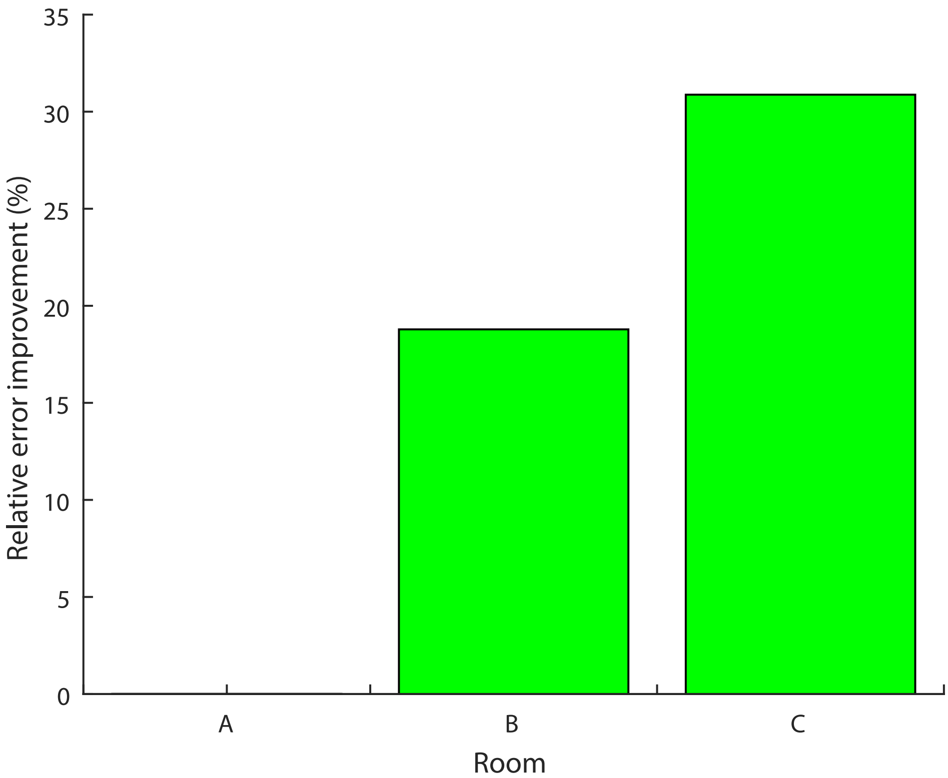

Figure 3.6: Relative error improvement in the office evaluation area.

3.4.7 Discussion

The presented algorithm was designed in order to improve the accuracy in the boundary locations. These are the locations in a room that are farther away from the center of the room, than the center of a smaller adjacent one. As seen from the measurements, accuracy was improved in the following such locations: points 13, 16 and 19 in the office evaluation area and points 3, 9 16, 40, 43 and 44 in the house evaluation area.

The deterioration of the accuracy in some points was mainly due to the combination of the RSSI fluctuating and the fact that the presented algorithm intrinsically favors presence in bigger rooms. This especially holds true for point 20 in the office evaluation area, where although the RSSI from the beacon in room C was in average

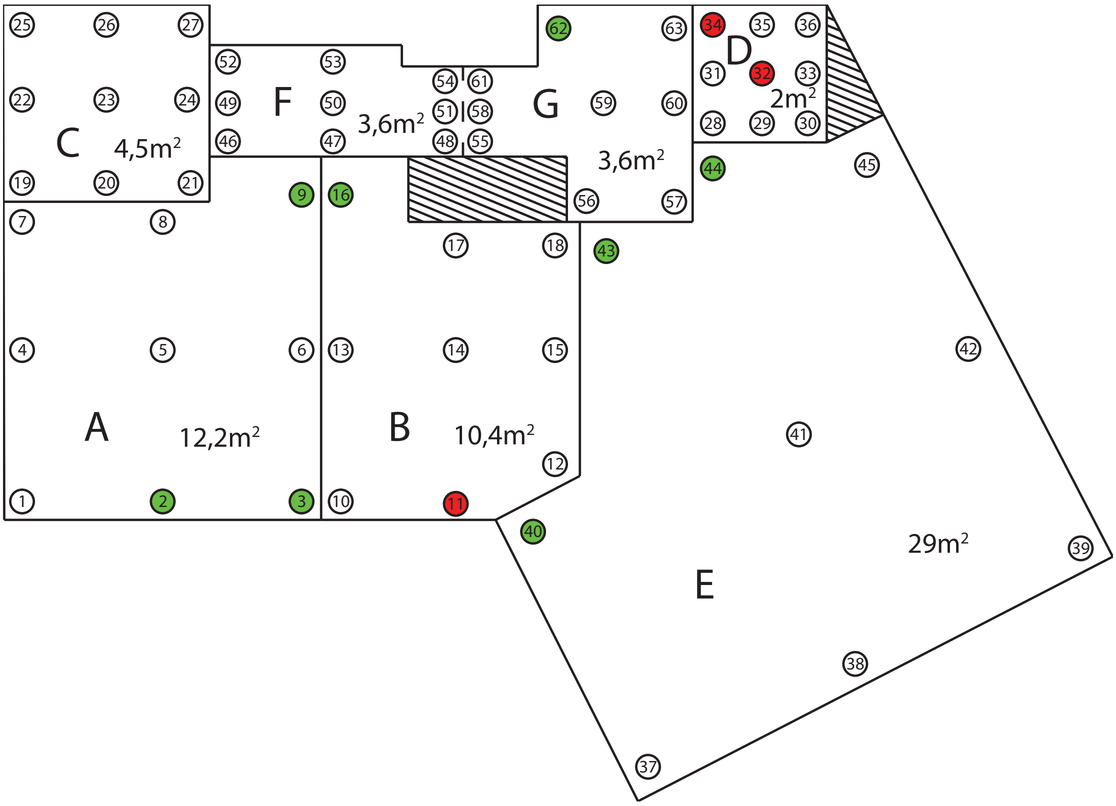

Figure 3.7: House evaluation area and targeted locations (green for the points with an error improvement, red for an error deterioration).

higher than the RSSI from the beacon in room B, the fluctuation of the signal along with the thresholds introduced by the algorithm eventually decreased the localization accuracy.

This limitation of the proposed algorithm can also be noticed in room D of the house evaluation area. By being the smallest room in the localization area, it is the least favored room regarding the calculated thresholds. Any fluctuation of the RSSI of the beacon located in this room has, therefore, a stronger influence to the category classification of the beacon as also inferred from Figure 3.4.

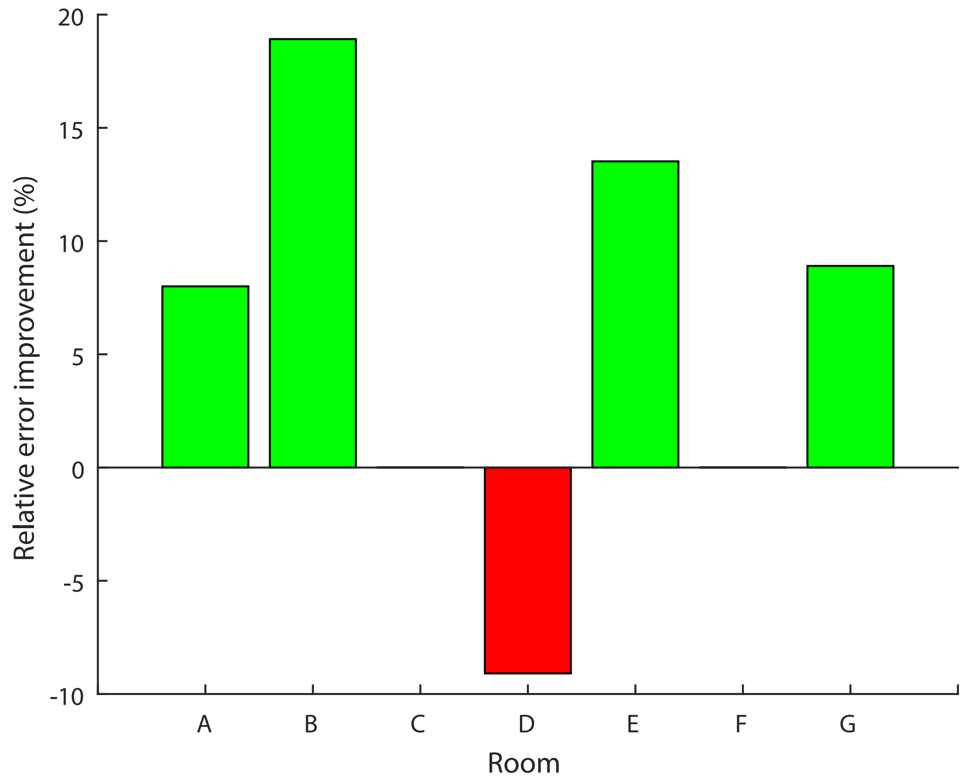

Figure 3.8: Relative error improvement in the house evaluation area.

3.5 Conclusion

In this chapter, we have presented an easy to deploy BLE-based indoor positioning system with room-level accuracy. The system only requires the geometry of the rooms and BLE beacons attached to the ceiling in the center of every room. The presented algorithm computes two RSSI thresholds for every room, and based on them, categorizes the RSSI readings from every relevant beacon and finally estimates room localization probabilities for the user. The proposed system can be easily deployed and used by any application requiring room-level indoor localization information.

In order to evaluate the performance of our system, we have deployed it in two different locations, an office environment composed of three rooms and a house environment composed of six rooms. After comparing it with the no-threshold approach, we saw an improvement of room estimation accuracy, especially in the boundary locations of the rooms. These are the locations in a room that were farther away from the center of the room, than the center of a smaller adjacent one. Overall, out of the 90 points that measurements were taken, the algorithm that we propose on average managed to improve the localization accuracy of 13 of them, decreased the accuracy of 4 and did not affect the rest.

There are, however, limitations of the presented work. We have evaluated its usability in two different living scenarios, but that is far from proving that our approach works in any case. There are countless different house and office arrangements and items that can may deteriorate signal propagation and reception. We have also not evaluated our approach at different times of the day in the same scenario to evaluate the impact of temporal factors like the usage of nearby Wi-Fi networks or the power network.

4 Stress and Mood Detection from Smartphone Usage Patterns

4.1 Chapter Abstract

Stress is undoubtedly an integral part of modern society, and people feel often overwhelmed by the pressures imposed by different aspects of modern life. At the same time, smart devices and especially smartphones have entered our lives for good, and their usage is nowadays widespread. These devices can be used for a variety of things and are meant to facilitate several aspects of our daily life. The question that we are examining in this chapter is whether smartphones can be used to detect stress and prevent chronic stress situations. We have organized an experiment in which we were collecting various data from the smartphones of four users during their daily lives. The experiment lasted for four weeks, and the participants were providing feedback four times a day regarding their subjective perception regarding their happiness, activeness, and stress levels. We explore how data coming from different sensors can be synchronized and fused. Then we build several machine learning models of different modeling traditions and manage to achieve a lift of predictive performance for all psychological states when compared against the corresponding naive models: these are the models that would always predict the psychological state with the highest occurrence in the collected dataset. Smartphones can, therefore, serve as non-invasive devices when creating a stress and mood detection system, and other smart devices like smartwatches able to capture physiological signals may further increase the performance of such a system.

4.2 Introduction and Related Work

Decades ago, researchers identified that physical and mental health are closely intertwined [52]. With the widespread use of computers, studies have been held on devices and systems that can recognize and interpret human emotions. A new interdisciplinary field among computer science, psychology, and cognitive science was born, and this modern branch of computer science was called affective computing [53]. Stress is one among the different emotional states that affective computing targets to detect. It is an essential problem in modern societies. Stress can damage physical health, and detecting it early prevents it from becoming chronic and reduces costs to society.

Defining stress, however, is not straightforward as it is a subjective phenomenon [54]. Non-formally, stress can be defined as the body's method of reacting to a dangerous or challenging situation. When the body senses a threat, the brain instantiates a response called "fight-or-flight" to defend itself, and a signal is sent to the hypothalamus [55]. The hypothalamus is controlling the involuntary body functions, such as breathing and heartbeat, via the autonomic nervous system. This system has two components, the sympathetic nervous system, and the parasympathetic nervous system. In a stressful situation, the sympathetic nervous system is activated, and stress hormones such as epinephrine and cortisol are released to arouse the body, increase the heart rate and tighten the muscles. When this mechanism works optimally, the senses become sharper, and the individual becomes more alert, focused, and energetic. Then, when the dangerous situation is gone, the parasympathetic nervous system calms the body down and dampens the stress response.

This stress reaction takes place when an individual faces a stressful situation. Stressful situations, however, can be subjective, and their appraisal and interpretation can differ from person to person. Measuring the perceived stress is challenging, and a common way to approach this is via self-reports [56]. The adverse effect that stress has on memory and the fact that people tend to forget stressful events pose another challenge in the reported perceived stress levels [57]. Stress is beneficial up to a certain level for the individual to survive dangerous situations. After this level is exceeded, stress starts to become harmful and damages health and quality of life.

Apart from impacting people on a personal level, excessive stress has a negative impact on society and the economy. Office environments, where constant high mental workloads exist and where technological progress brings changes and requires continuous adaptation, contribute to stress for office workers [58]. Two-thirds of all employees in the EU are experiencing stress, and one fifth believes that stress is their most pressing health issue at work [59]. Stress is the second most reported health-related issue at work in Europe, after musculoskeletal disorders [60]. The most common reasons for sick leaves in Europe are also related to stress [61]. The cost that stress brings in the economy of Europe is in the billions of euros. The cost is split between employers, that pay due to reduced productivity and sick-leave absenteeism, and to the society, that pays for worker compensations and medical costs.

Stress is typically categorized into two kinds: acute stress and chronic stress [62]. Acute is the short-term stress that results from unpredictability and a poor sense of control. Chronic stress, on the other hand, results from repeated exposure to stressful events. Humans are good at managing acute stress by releasing stress hormones to help the body and the mind deal with the present situation. Our stress response system, however, is not designed for constant activation and extensive use of it due to chronic stress may lead to health problems.

Several methods have been proposed to invoke stress in a laboratory environment on demand. Subjects are asked to perform specific stressful tasks, and at the same time, researchers measure stress signals. A widely used stress induction test is the Stroop Color and Word Test [63]. In this test, the subjects are initially asked to read the names of the color words and at a later stage to name the color of the font and ignore the color that is written. Another stress-inducing test is a cardiovascular one named the cold pressor test [64], where the participant's hand is put in ice water. The researchers can measure each participant's pain threshold and tolerance by monitoring the blood pressure and the heart rate, by listening to the participants express the pain they feel, and by allowing them to remove the hand from the ice water once the pain becomes unbearable. A last commonly used test is the International Affective Picture System (IAPS) one, that consists of a standardized set of pictures for studying emotion and attention [65]. The pictures consist of various casual and extreme ones that are meant to induce stress and to arouse emotions to the viewers.

There are many ways that stress symptoms can be observed and thus measured. The activation of the sympathetic nervous system ignites many physiological, behavioral, and psychological symptoms. Psychological symptoms include intense emotional responses, and an increased number of negative emotions, such as anxiety, anger, fear, and depression. The way to quantify those symptoms is via the use of self-reported data, or via periodically interviewing participants of an experiment.

A more objective way to quantify stress, however, is by monitoring physiological information. One of the most prominent ways to detect stress is via heart activity. The Heart Rate (HR) and more importantly the Heart Rate Variability (HRV) are two essential metrics [66]. An electrocardiogram (ECG) can be employed to measure the electrical activity of the heart, and can nowadays even be performed with a smartwatch [67]. Brain activity is also affected by stressful events and emotional changes. It can be measured with an electroencephalogram (EEG), and by applying signal processing on certain frequency bands, stress can be detected [68]. Other physiological signals to detect stress include measuring the muscle and the neural activity via an electromyogram (EMG) [69], measuring the Electrodermal Activity (EDA), also known as Galvanic Skin Response (GSR) [70], measuring the skin temperature, the respiration rate, blink rate variations, and the pupil diameter [58].

Another way to detect stress is by noticing behavioral changes. Stress has an effect on the behavior of individuals. Behavioral reactions attributed to stress include stress-related facial expressions [71], and changes to the human vocal production, in terms of pitch (fundamental frequency) and speaking rate [72]. Posture is also a good indicator of human feelings, and may therefore also provide an insight on stress levels [73].

Stress may also be noticed through mouse and keystroke dynamics by identifying emotional states by identifying each person's typing rhythm [74]. An extension to this method, the method that our work is focused on, is the detection of stress using smartphones [75]. Many information regarding users' behavior can be extracted in a non-intrusive manner and exploited to detect stress.

Studies towards stress detection can be categorized into the controlled laboratory ones, the restricted and the unrestricted ones. In laboratory controlled environments, usually physiological signals of individuals are monitored during the time they are exposed in various stressors. In controlled experiments, users are monitored with cameras or other sensors in controlled environments like in offices or classrooms. Lastly, in the unrestricted studies, the daily lives of individuals are monitored with various intrusive or non-intrusive sensors, while a ground truth estimation is acquired with the use of subjective questionnaires.

On one end of the experiments spectrum, controlled laboratory experiments can count as preliminary works towards daily life stress detection, as they provide researchers with an initial insight into which sensors and data analysis techniques can be used to detect stress. In those studies, ground truth collection is not required as the researcher has control over the stress levels the participants are exposed to in the experiments. The most characteristic signals used for stress detection in such experiments come from heart activity and are the HR and the HRV. Different time and frequency domain features are constructed before classification to differentiate between stressed and relaxed states [76]. GSR is the next most used signal for stress detection, as the body sweats and skin conductance increases [77] under stressful or emotional situations. Controlled stress detections experiments have also been carried out by using speech data [78] or brain activity using EEG signals [79]. Multimodal monitoring has also been explored to improve stress detection accuracy even further.

Stress detection experiments in restricted environments act as a bridge between laboratory research and real-life stress detection environments. Offices are among the places that contribute to stress the most, and the limited movement of individuals in the offices during working hours makes those individuals excellent participants for stress detection experiments. Relevant experiments have been carried out among call center agents by monitoring the EDA during stressful and non-stressful calls [80]. In automobile environments, movements are restricted, and therefore, ambient sensors can be used in conjunction with other potentially intrusive ones. Stress among drivers can fluctuate, especially when among traffic jams in crowded cities, and using different biological sensors the drivers' stress level can be estimated [81]. The campus environment is a semi-restricted one that resembles the unpredictability of daily life experiments. Researchers have focused on improving the quality of life of students by evaluating their mental well-being [82]. Although this research is not directly aiming at detecting stress levels, the collected dataset can be used to do so.

On the other end of the spectrum, and aligning with the goal of this chapter, an ideal stress detection system should be able to detect stress during the daily lives of its users. Data that can be used for stress detection in daily life scenarios can come from behavior patterns measured with ubiquitous devices such as smartphones, and from physiological signals measured with wearable sensors [83]. The most significant challenges in those scenarios are the reliability of the collected ground truth and the causality relationship between smartphone usage and induced stress. Stress detection accuracies in daily life schemes tend to be lower than the ones in restricted, and laboratory environments [75], and such stress detection systems are usually accompanied by stress alleviating methods that are proposed to the user when stress is detected.

In our system, we are collecting data from a smartphone. These data include sensor data, along with user-smartphone interactions. We are also collecting the subjective perception of the users regarding their mood and stress levels via a questionnaire. The research question of this chapter is, "How can smartphone data be translated to mood and stress states?"

The rest of this chapter is organized as follows. In Section 4.3 we present the system we designed. We experimentally evaluated the performance of the proposed system in a daily life experiment, and the results are presented in Section 4.4. Finally, our conclusions are drawn in Section 4.5.

4.3 System Overview

Technology has revolutionized our daily lives over the past decade. In a UK based study conducted by Ofcom, the regulator and competition authority for the UK communications industries, shifts in human behavior due to technological advances were researched [84]. 78% of people own a smartphone nowadays, up from 17% a decade ago. They check their smartphones on average every 12 minutes. 65% of people under 35 check their phones within 5 minutes of waking up, and 60% of them check their phones 5 minutes before going to bed. The average time people spend on their smartphones has risen to 2 and a half hours each day while 78% claim that they could not live without one anymore.

It is, therefore, apparent that smartphones have entered our lives for good. Given the number of hours that each person interacts with the phone every day, it is possible that specific behavior patterns can be identified. By monitoring different interactions and events that take place on the phone, we are conducting research on whether the stress level of users can be evaluated throughout the day.

4.3.1 Mobile Application

For this study, we developed an Android application that would collect data about the daily routines of the users. The application is collecting data while running in the background of the smartphone as a service. Attention was given to minimizing battery consumption in order not to interfere with the everyday use of a smartphone. We do not require the use of any other connected device and want to make the experience as non-intrusive as possible. The application is also periodically and securely transmitting and saving all the collected information in a database on a server we deployed. All data are anonymized and only contain a numeric user identifier to protect user privacy.

4.3.2 Subjective Feedback

During the time the application is running as a background service, the users are periodically asked to leave feedback regarding their subjective perception of their mood. Since detecting stress was the primary target of our application, reporting stress was the main feedback we asked for. However, instead of displaying the "Stress" word in the provided questionnaire, we preferred using the word "Relaxation" as to prevent any potential negative anchoring cognitive bias [85]. In the questionnaire we developed, we also include questions regarding the subjective "Happiness" and "Activeness" of the users in order to explore different states of emotions we could identify with the collected data. Those questions were inspired by the Circumplex Model of Affect [86], a commonly used model to study human emotions.

A screenshot of the experience sampling method we use is presented in Figure 4.1. We opted to use analog bars for the user interface and not classes, like stressed and relaxed for the "Relaxation" question, because we wanted to avoid any initial bias and wanted to enable the users to evaluate their subjective mood on a virtually continuous spectrum. Other similar experiments have used a similar sampling method with a scrolling bar resulting in a continuous value [87]. Having a visual analog scale like the one we opted for can often maximize the precision of the answers [88], in comparison to the widely used five- or seven-point Likert scales [89]. The values that we are getting for each question range from 0 to 99.

4.3.3 Monitored Data

Our application monitors different data and information from the smartphone. Data are saved in the form of events. Each event includes at least the "user identifier", the "timestamp" the event took place, and the "type" of the event. Those events include:

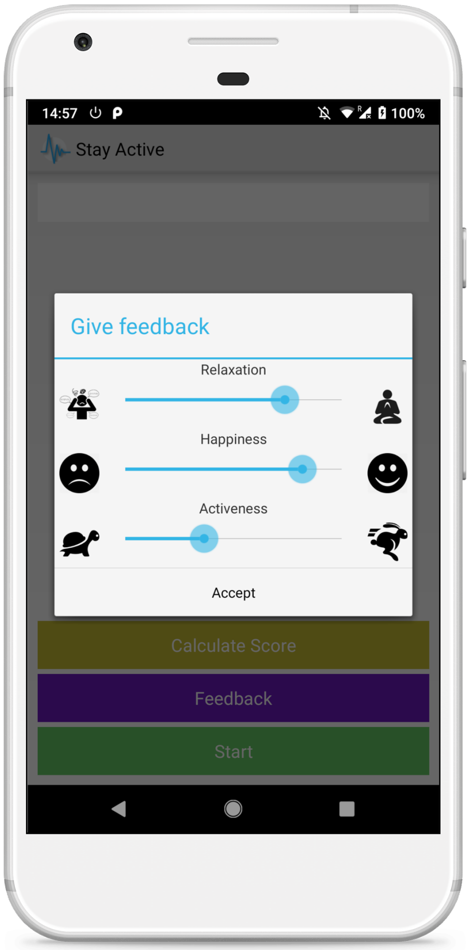

Figure 4.1: Screenshot of the implemented experience sampling method.

Activity The user activity was derived via Google's Activity Recognition API [90]. A new activity estimation is obtained at most every 30 seconds. The "type" of activity ranges between:

- ACTIVITY_BICYCLE
- ACTIVITY_STATIC
- ACTIVITY_UNKNOWN
- ACTIVITY_VEHICLE

- ACTIVITY_WALKING

A numeric variable called "confidence" ranging between 0 and 100 is also provided, declaring the certainty of the reported activity type.

Apps An app event is saved each time the user launches an application. The "type" of this event always takes the value APP_LAUNCHED, and a "name" variable saves the name of the application the user launched.

Calls A calls event is saved each time the user places, receives or is in a phone call. The "type" of a call event ranges between:

- CALL_CALLING
- CALL_IDLE
- CALL_OFFHOOK
- CALL_RINGING

Feedback A feedback event is saved each time the user provides feedback. The "type" of this event always takes the value FEEDBACK_GIVEN, and three numeric values, the "Activeness", the "Happiness" and the "Relaxation", ranging from 0 to 99 include the provided feedback.

Screen A screen event is created when the screen is either turned on or off, and the "type" ranges between:

- SCREEN_OFF
- SCREEN_ON

SMS An SMS event is saved each time the user sends or receives an SMS, and the "type" ranges between:

- SMS_SENT
- SMS_RECEIVED

Steps A steps event is created each time Android's step detector detects steps. The "type" of this event always takes the value STEPS_DETECTED, and a "count" variable contains the number of steps that were detected.

Touch A touch event is created periodically according to smartphone usage. The "type" of this event always takes the value TOUCH_SESSION, and an "amount" variable contains the number of screen touches that were detected.

Unlike in controlled laboratory environments, data collection is more prone to error in daily life scenarios. In an ideal scenario for our case, individuals should always carry their smartphones with them in order to capture as much correct data as possible. Improper placement of the phones may introduce problems, especially with activity recognition. Users should also pay attention and never let the smartphone run out of power to avoid potential data gaps.

4.4 Experiment and Evaluation

We conducted an experiment in which we will evaluate whether user perceived stress can be estimated by using data acquired from a smartphone. We did not require the participants of our experiment to use any other connected device, like a smartwatch. We only required that they installed the previously described Android application to their smartphone, gave permissions to let the app run in the background, and we periodically asked for user feedback about their stress and mood levels.

In this section, we start by describing the experiment that was conducted. Then we present the data understanding, the data preparation, and the feature engineering steps that we performed. We conclude by presenting the machine learning models we built along with their predictive performances.

4.4.1 Experiment Methodology

We managed to record data of 4 male participants of ages 26-30 for a period of 4 weeks. More participants were initially recruited but did not manage to keep up with the experiment for the required period of 4 weeks. We requested the users to complete the given questionnaire four times a day via Android notifications. The times that the notifications triggered every day were 08.00, 11.30, 15.00, and 20.30 in order to capture mood estimations throughout the day. One would argue that having those fixed times of the notifications would introduce an anticipation bias to the responses after a couple of days in the experiment. However, the participants were not obliged to reply at that specific moment and could freely respond during their subsequent interaction with their smartphone. All recorded data were periodically and anonymously sent to our server.

4.4.2 Data Understanding and Preparation

After the end of the experiment, all available data were aggregated on the server's database and were later downloaded on a desktop machine for offline processing. All processing and modeling were done using the R statistical software. The raw data were initially filtered based on the user ids and based on the experiment dates, in order to keep only the data entries of interest.

4.4.2.1 Construction, Integration and Filtering of the Data

The available dataset was ordered based on the user id and then based on the timestamps, as to provide a meaningful consistency. The psychic perception of a user is the result of all things that occurred before the moment of capturing that perception. This is the assumption that was necessary for our data construction technique.

During our experiment, the user was providing feedback for three different aspects, the "Activeness", the "Happiness" and the "Relaxation". Therefore three distinct machine learning models should be trained, and to do so, three distinct datasets should be created, one for each outcome. Moreover, we needed to create a single row of input values (features) for each feedback provided. Since many different events exist before each feedback event, we needed to merge the information from all previous events into a single representative one, along with the response value that follows. This single line in the final dataset should be a full summary of all that happened previously and led to a specific response.

4.4.2.2 Exploratory Data Analysis

An exploratory data analysis was performed on the dataset before the feature engineering phase. All existing events in the dataset have a specific type that refers to the type of action the smartphone recorded. This is a categorical variable, and Figure 4.2 presents the frequencies of occurrences for all distinct type values.

Regarding the activities that were detected, the activity recognition system we used reported a confidence level along with each activity entry. Figure 4.3 presents the frequency of the confidence values. The majority of the values suggest that we can trust in general the activities contained in the dataset. This is also suggested by the statistical mean of 87.9% and the standard deviation of 18.7% of the confidence values.

The "count" and "amount" variables contain the number of steps and the number of screen touches that were detected and saved per "steps" or "touch" event. Figures 4.4 and 4.5 present the values of both those variables. They both follow a typical Pareto distribution, where most of the instances have low values. Regarding the "apps" events, there were 205 different apps the participants used, some common among them and some unique per user.

Concerning the target values, the frequency histograms of the "Activeness", the "Happiness" and the "Relaxation" responses are presented in Figures 4.6, 4.7 and 4.8

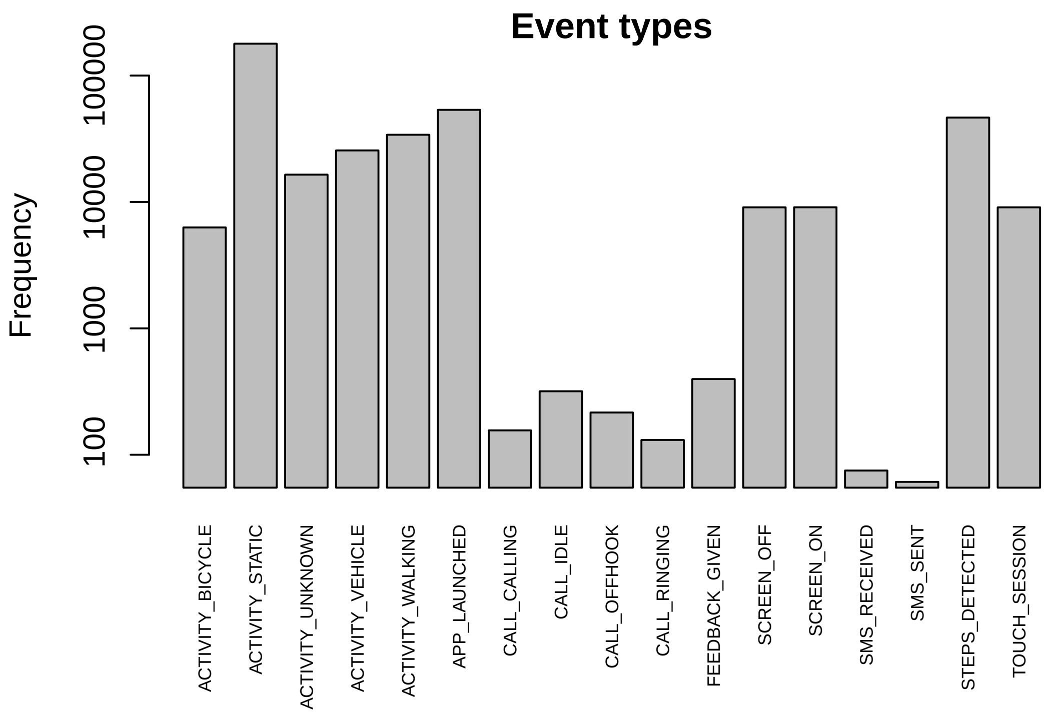

Figure 4.2: Frequency of appearance of different events in the dataset.

respectively. The average response rate across all participants was 82.71%. In all three cases, we notice that many values are concentrated in the middle of the spectrum. Their volume is inconsistent with the rest of the frequencies. This is attributed to the design of the mobile application. When the user was being prompted to leave mood feedback, the default values of the of all 3 bars were set to 50. This led to the anchoring effect, and many users did not change this value at all. We consider these measurements unreliable, and we will later discard them from the dataset. By discarding those responses, the response rates drop to 77.29% for the "Activeness", 75.63% for the "Happiness", and 77.71% for the "Relaxation" responses. A better application UI design, in this case, would not have any initialized values when asking for feedback.

Moreover, for all three histograms, one can visualize two or three Gaussian distributions. One is for values above 50, and the rest can be discerned for values

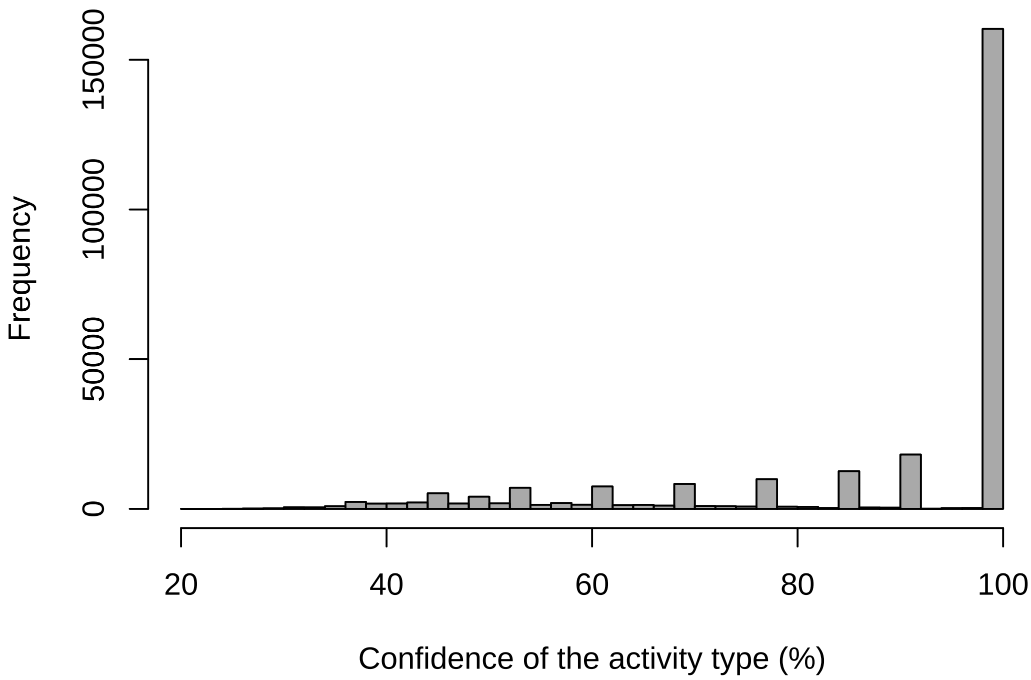

Figure 4.3: Frequency of appearance of confidence levels for the activity events in the dataset.

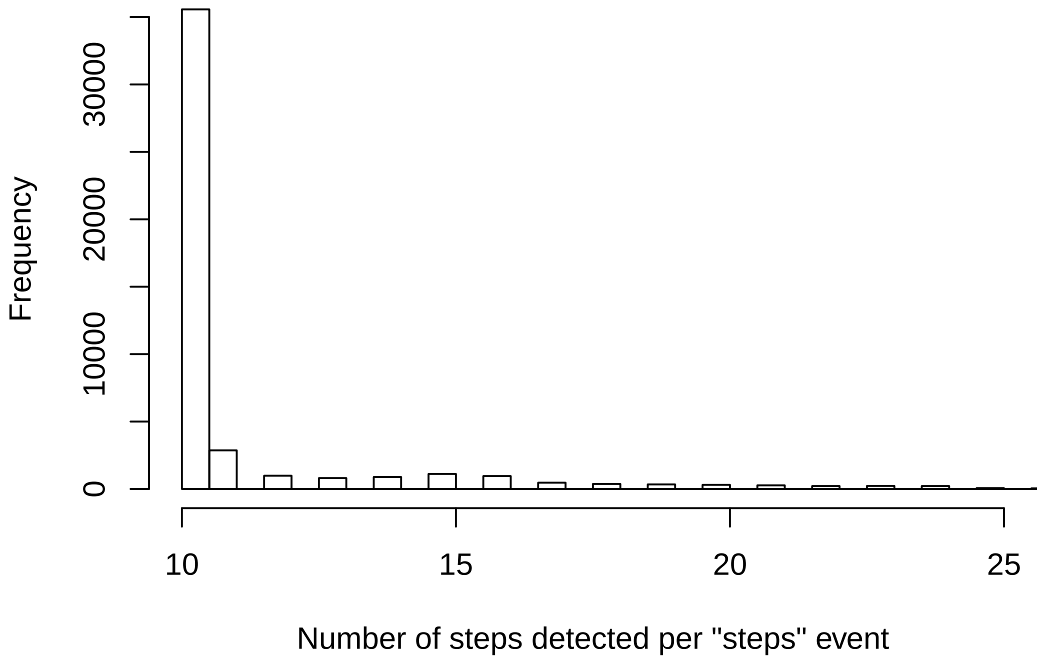

Figure 4.4: Number of steps per "steps" event in the dataset.

below 50. Finally, "Activeness", "Happiness" and "Relaxation", have all a mean greater than 50 in our dataset.

All three response values are numeric and extend from 0 to 99. However, it might be counter-intuitive to estimate these mental perception attributes in this context. Who can discern the difference between 67 and 72 in terms of "Activeness" or "Relaxation"? Thus, we have decided to aggregate these variables in three meaningful classes, by selecting cut-off points (see [87], [91]): "low" for values 0 to 33, "medium" for values 34 to 66 and "high" for values 67 to 99. Having grouped all feedbacks to the three classes, Figures 4.9, 4.10 and 4.11 present the frequencies of each class for each variable. Table 4.1 presents the probabilities of each feedback variable belonging to one of the three constructed classes in our dataset.

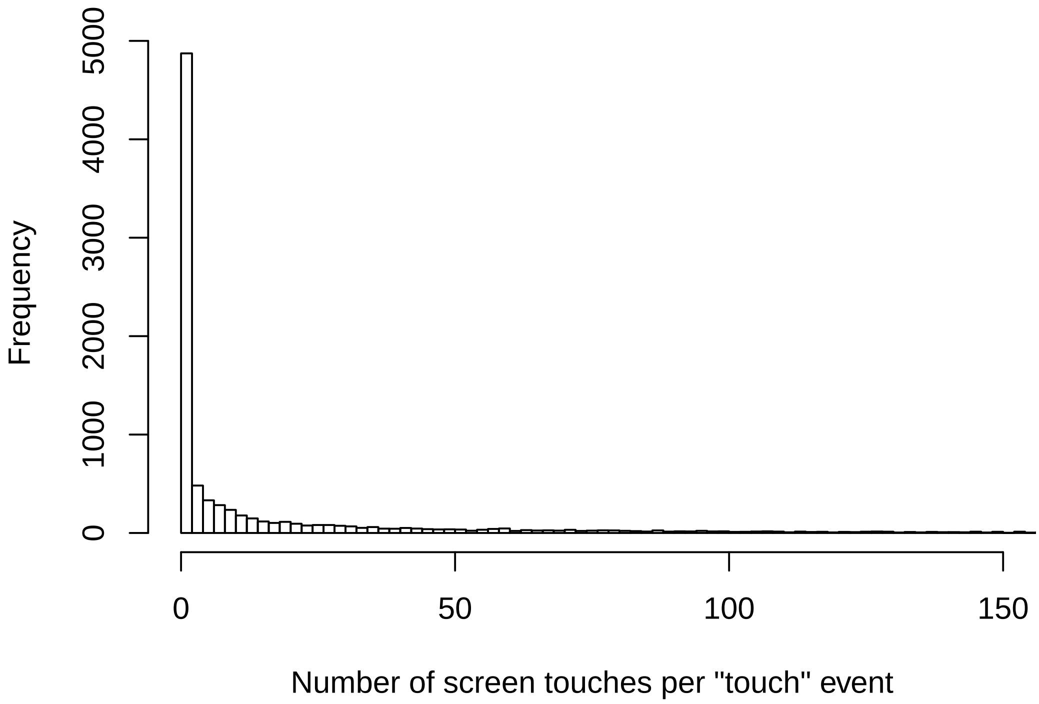

Figure 4.5: Number of screen touches per "touch" event in the dataset.

Table 4.1: Frequency of each class for all target feedback variables.

Feedback variable	**Classes (%)**		
	low	medium	high
Activeness	34.5	50.7	21.3
Happiness	11.8	43.8	47.1
Relaxation	11.8	37.8	50.4

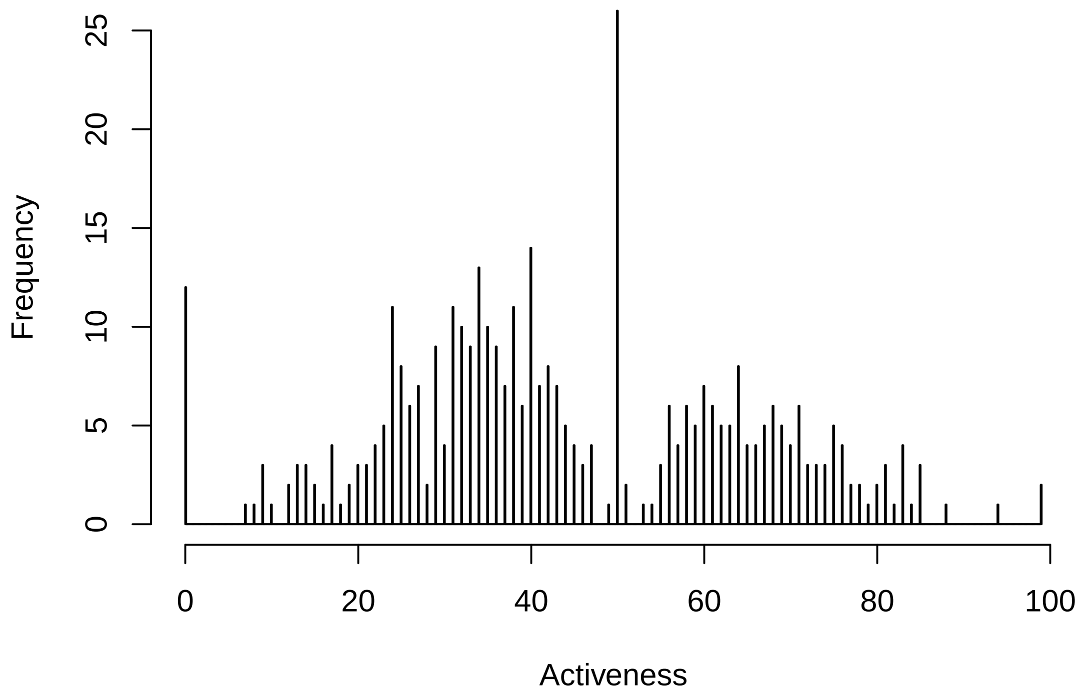

Figure 4.6: Histogram of the "Activeness" responses in the dataset.

4.4.2.3 Feature Engineering

From the recorded events that were described above, within each given time-frame, a total number of 226 features were constructed. The following features were extracted per monitored event:

Activity The number of activity instances among the different activity types:

- ACTIVITY_BICYCLE
- ACTIVITY_STATIC
- ACTIVITY_UNKNOWN
- ACTIVITY_VEHICLE
- ACTIVITY_WALKING

The total amount of features from this set of events is 5.

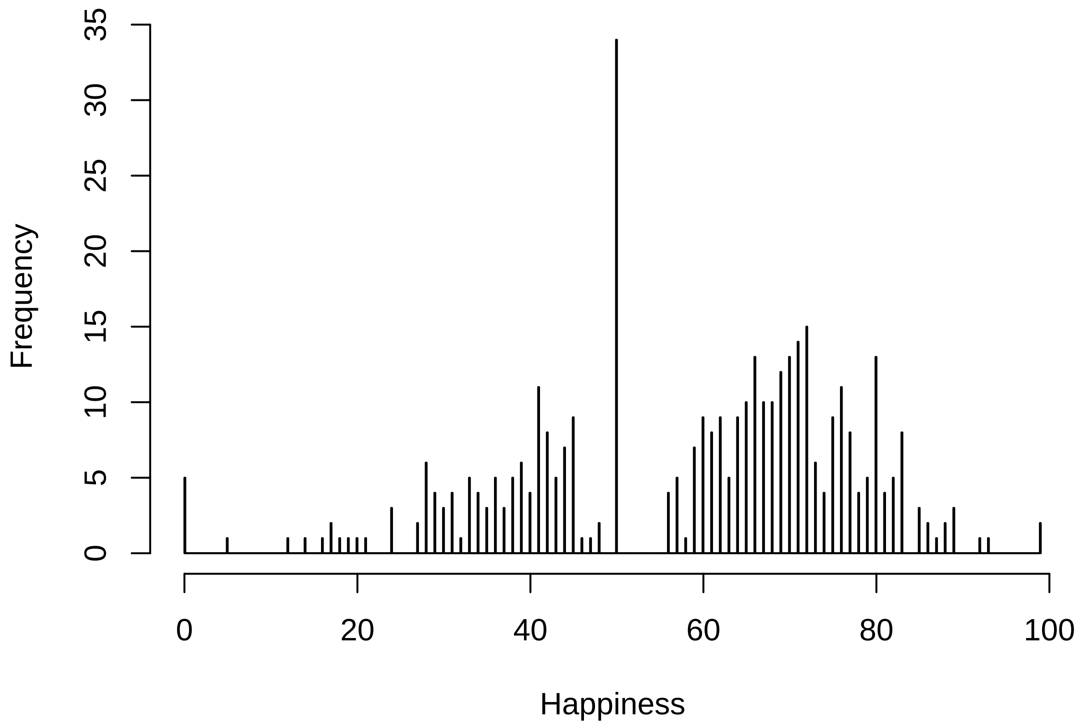

Figure 4.7: Histogram of the "Happiness" responses in the dataset.

Apps From the app related events, the total number of app executions was monitored. Moreover, for each of the 206 executed apps across all participants, the total number that each individual app was executed was also monitored.

So a total number of 207 features were constructed from the app events.

Calls The number of call events among the different call types:

- CALL_CALLING
- CALL_IDLE
- CALL_OFFHOOK
- CALL_RINGING

The total amount of features from this set of events is 4.

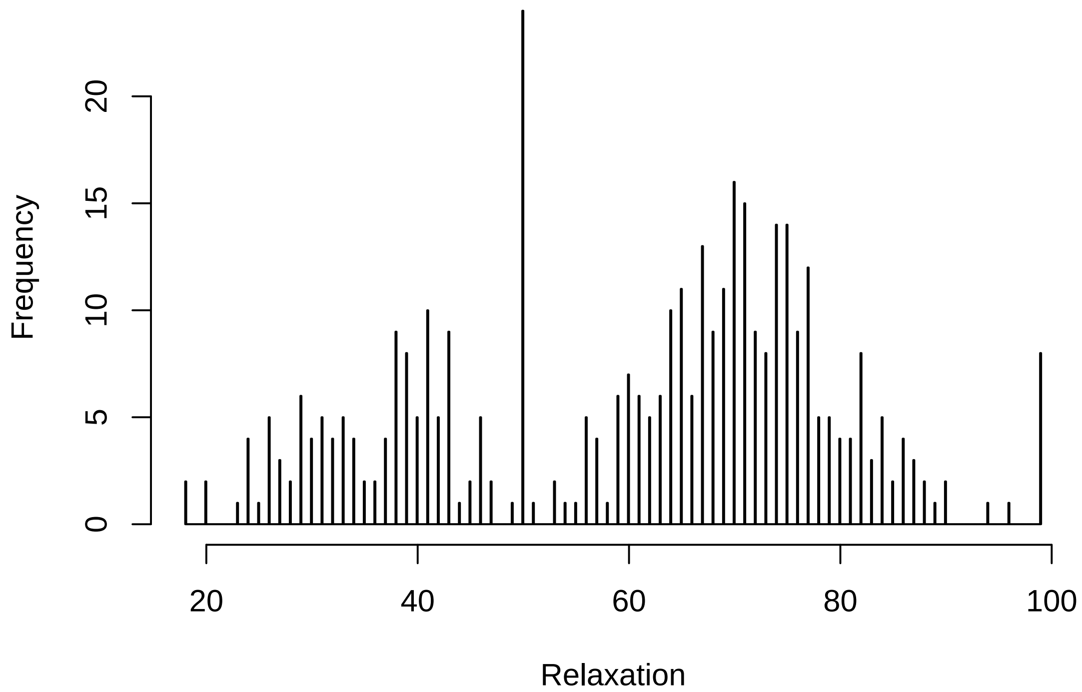

Figure 4.8: Histogram of the "Relaxation" responses in the dataset.

Screen From the screen events, the number of times that the participant turned on the screen of the smartphone was monitored. There was 1 feature that was constructed from the screen events.

SMS The number of SMS events among the different SMS types:

- SMS_SENT
- SMS_RECEIVED

The total amount of features from this set of events is 2.

Steps From the steps events, the total number of steps, the number of detected walking sessions along with the mean number of steps per walking session of the participant were calculated. So there were 3 features that were constructed from this set of events.

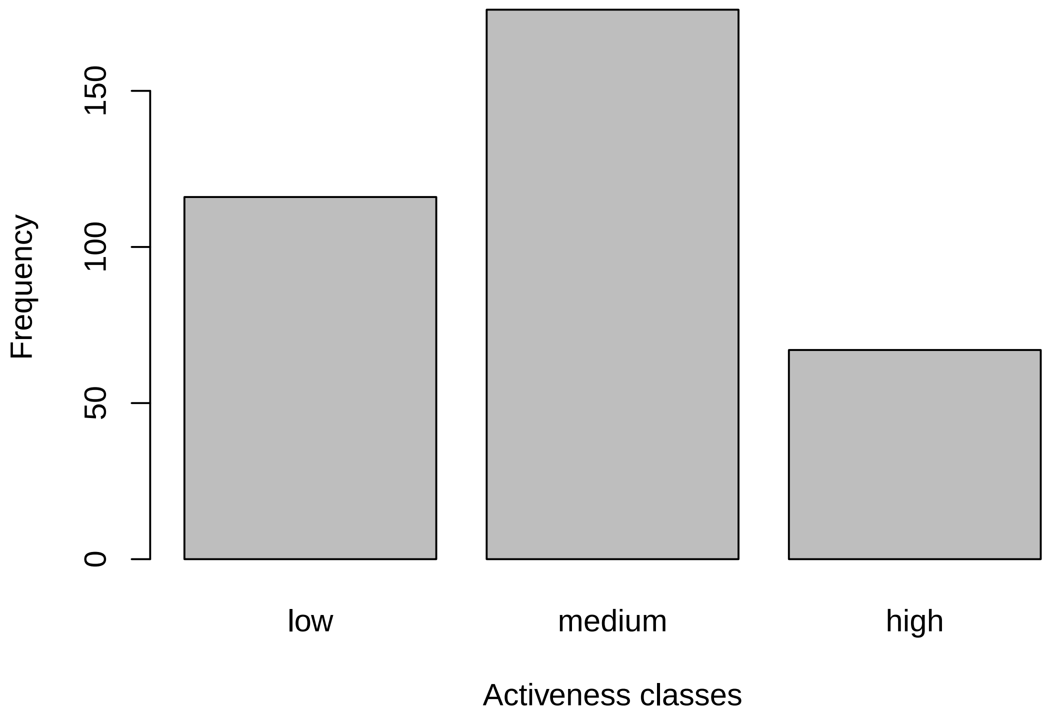

Figure 4.9: Frequency of appearance of the "Activeness" classes.

Touch As with the steps events, for the touch events, the total number of screen touches, the number of smartphone usage sessions along with the mean number of screen touches per smartphone usage session of the participant were calculated. So there were 3 features that were constructed from this set of events.

Miscellaneous Additionally, for each instance, the weekday was used as a categorical variable, so 1 more miscellaneous feature.

4.4.2.4 Data Summary

Table 4.2 summarizes all features in the dataset except for the 206 app-specific constructed ones and the categorical weekday feature. The 25%, 50%, and 75% columns show the corresponding percentiles. A percentile indicates the value below

Figure 4.10: Frequency of appearance of the "Happiness" classes.

which a certain percentage of observations among the whole set of observations fall. The value of the 50%, therefore, corresponds to the median.

Concerning the weekday feature, the dataset was relatively balanced with 49 instances of Monday, 56 of Tuesday, 56 of Wednesday, 53 of Thursday, 58 of Friday, 55 of Saturday, and 44 of Sunday.

4.4.3 Modeling

We are facing a multiclass classification problem. In order to compare the predictive performances of the models we are training, we are presenting the notion of a naive model. The naive model is a theoretical model that always predicts an output of the most frequent class within the available data. In terms of "Activeness", the naive model is the assumption that all predictions are set to "medium" and the naive model's predictive accuracy is equal to 50.7% (the "medium" class frequency in our

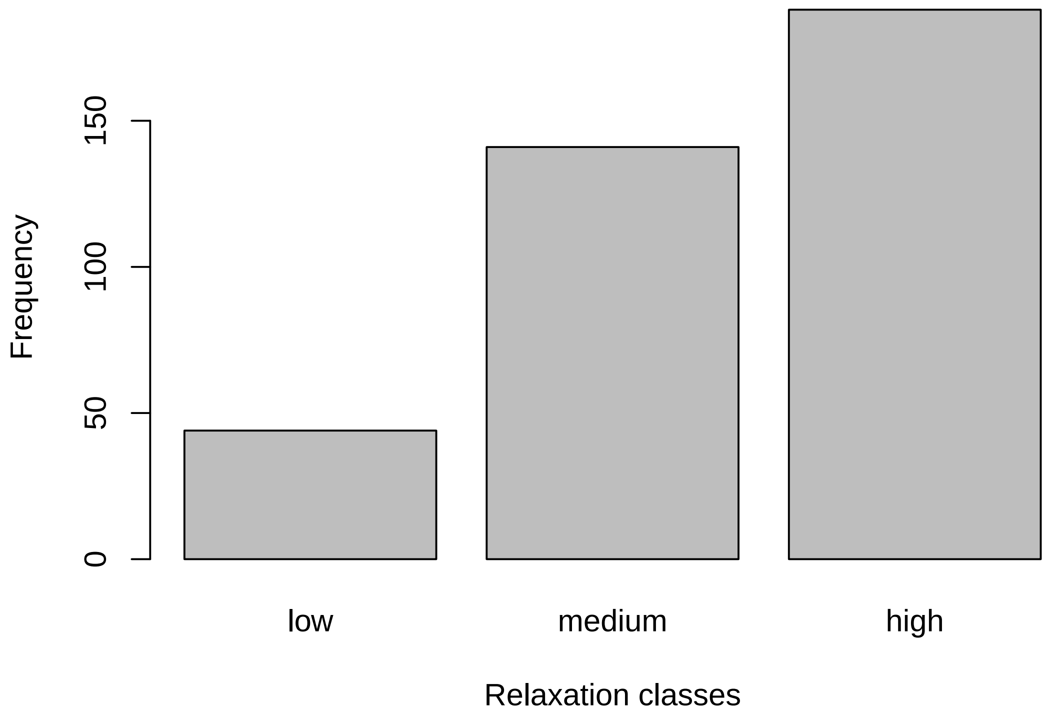

Figure 4.11: Frequency of appearance of the "Relaxation" classes.

Table 4.2: Dataset summary for the computed features.

Feature name	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Activity bicycle	14.59	0	0	0	1	474
Activity static	445.6	0	3	421	663	3342
Activity unknown	41.56	0	0	20	59.5	511
Activity vehicle	66.89	0	0	4	70	768
Activity walking	89.09	0	0	21	94.5	2328
Apps launched	134.8	0	60	103	178.5	814
Phone calling	0.41	0	0	0	0	9
Phone idle	0.8	0	0	0	1	12
Phone offhook	0.54	0	0	0	0	10
Phone ringing	0.32	0	0	0	0	5
Screen on	22.67	0	11	18	30	162
SMS received	0.2	0	0	0	0	5
SMS sent	0.16	0	0	0	0	10
Steps sessions	120.5	0	7	55	133.5	2769
Usage sessions	22.65	0	11	18	30	161
Steps	1780.3	10	444.5	1058	1967.5	29124
Mean steps P.S.	43.99	10	11.01	12.4	18	4031
Screen touches	692.9	2	209	488	955	7222
Mean screen touches P.S.	31.05	0.67	12.35	23.44	37.78	333

dataset). Regarding "Happiness", the respective naive model is the assumption that all predictions are set to "high" where the predictive accuracy is equal to 47.1% (again the "high" class frequency in our dataset). Finally, the "Relaxation" naive model is setting all predictions to "high" and under the same reasoning, the predictive accuracy is equal to 50.4%.

4.4.3.1 Modeling Techniques

Multiple machine learning algorithms were tested to evaluate the performance of different modeling traditions on our dataset. More specifically the classifiers that we tested are the Random Forest (RF) [92], Gradient Boosting Machines (GBM) [93], Partial Least Squares (PLS) [94] and the Support Vector Machines (SVM) [95]. All models were trained with the goal of maximizing the predictive accuracy metric, and the 10-fold cross-validation scheme was used. All tests were run on R using the caret package [96].

4.4.3.2 Predicting Activeness

Figure 4.12 and Table 4.3 present the predictive accuracies of all the models we trained to predict "Activeness". Besides those results, t-tests were performed on the models' performances in pairs. Practically, for each pair the difference of the mean accuracies of two models considering their respective variances, a t-test was performed at 95% confidence interval. This is to account for the fact that the accuracy results are not stable but present a considerable variance. Based on the t-test results, it can be seen that RF and GBM outperform PLS and SVM. However, no clear decision in a statistically significant way can be made on whether RF or GBM performs better.

Since we needed to select one model for our study, we selected the trained RF model as it is the one that delivers the best predictive performance. Comparing the predictive performance of RF against the naive model (the model that always predicts "medium") where the predictive accuracy is equal to 50.7%, we can observe that our machine

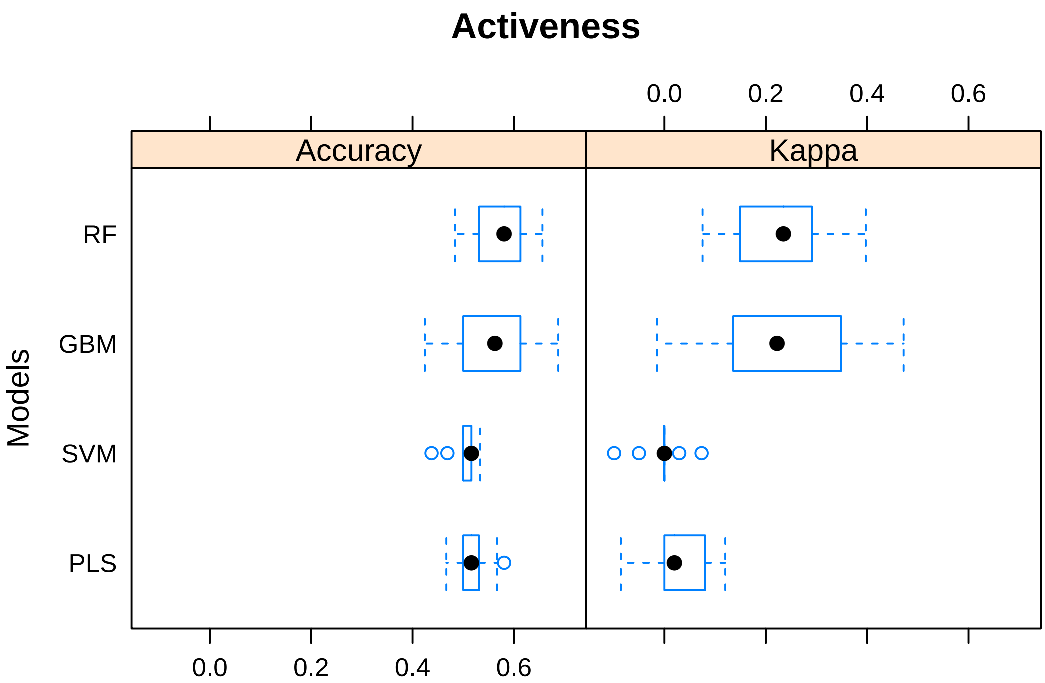

Figure 4.12: Predictive accuracy of multiple classifiers for predicting "Activeness".

Table 4.3: Predictive accuracy of multiple classifiers for predicting "Activeness".

Model	Min	Mean	Max
PLS	0.467	0.517	0.581
RF	0.484	0.571	0.656
GBM	0.424	0.558	0.688
SVM	0.438	0.511	0.533

learning approach provided a lift of predictive performance equal to 6.4% (57.1% - 50.7%). Table 4.4 presents the confusion matrix of the trained RF model. Table 4.5 reports the 5 most significant features for classifying "Activeness" using the selected, trained RF model.

4.4.3.3 Predicting Happiness

In a similar manner, Figure 4.13 and Table 4.6 present the predictive accuracies of all the models we trained to predict "Happiness". Besides those results, t-tests were

Table 4.4: Confusion matrix of the trained RF model for predicting "Activeness".

	true low	true medium	true high
predicted low	9.5	5.7	2.1
predicted medium	20	42.4	11.1
predicted high	0.8	3.2	5.2

Table 4.5: Most significant features for predicting "Activeness".

Rank	Feature
1	Mean screen touches per session
2	Steps
3	Mean steps per session
4	Steps sessions
5	Activity static

performed on the models' performances in pairs. Based on the t-tests, PLS, RF, and GBM outperform SVM. However, no clear decision in a statistically significant way can be made on which of the PLS, RF, and GBM model performs the best.

Since we needed to select one model for our study, we selected the trained RF model as it is the one that delivers the best predictive performance on average. Comparing the predictive performance of RF against the naive model (the model that always predicts "high") where the predictive accuracy is equal to 47.1%, we can observe that our machine learning approach provided a lift of predictive performance equal to 14.2% (61.3% - 47.1%). Table 4.7 presents the confusion matrix of the trained RF model. Table 4.8 reports the 5 most significant features for classifying "Happiness" using the selected, trained RF model.

Table 4.6: Predictive accuracy of multiple classifiers for predicting "Happiness".

Model	Min	Mean	Max
PLS	0.406	0.607	0.813
RF	0.438	0.613	0.774
GBM	0.452	0.612	0.75
SVM	0.375	0.556	0.719

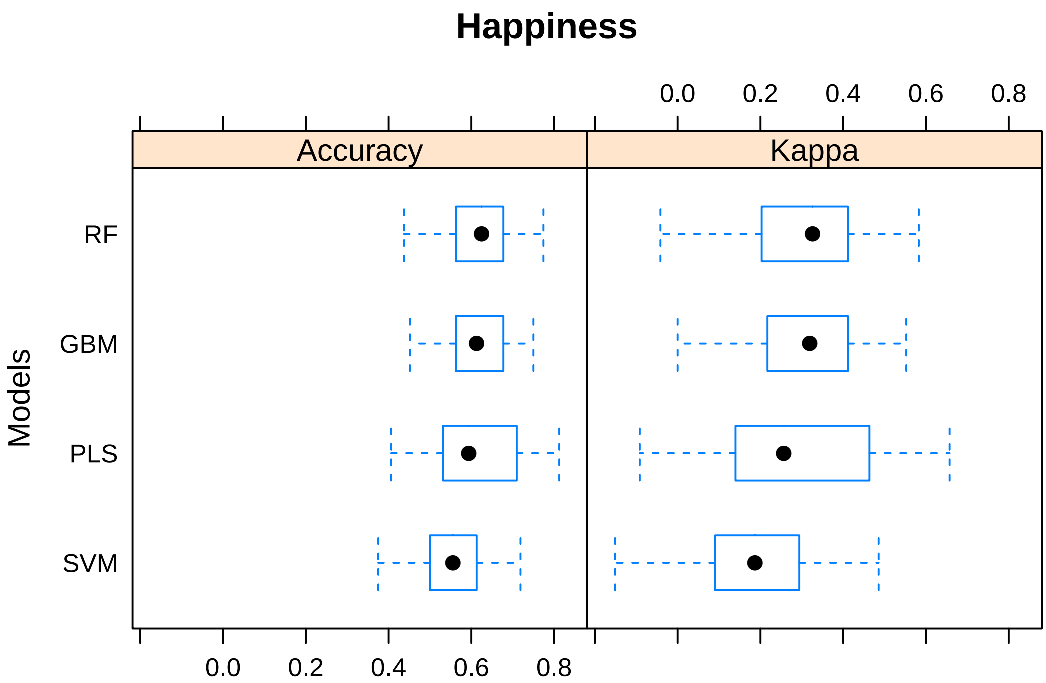

Figure 4.13: Predictive accuracy of multiple classifiers for predicting "Happiness".

Table 4.7: Confusion matrix of the trained RF model for predicting "Happiness".

	true low	true medium	true high
predicted low	0.4	0.1	0.0
predicted medium	6.3	26.8	12.4
predicted high	2.7	17.1	34.1

Table 4.8: Most significant features for predicting "Happiness".

Rank	**Feature**
1	Activity unknown
2	Mean steps per session
3	Steps
4	Activity static
5	Mean screen touches per session

Table 4.9: Predictive accuracy of multiple classifiers for predicting "Relaxation".

Model	**Min**	**Mean**	**Max**
PLS	0.394	0.58	0.697
RF	0.515	0.628	0.794
GBM	0.471	0.603	0.788
SVM	0.406	0.54	0.719

Table 4.10: Confusion matrix of the trained RF model for predicting "Relaxation".

	true low	true medium	true high
predicted low	0.3	0.8	0.0
predicted medium	9.0	23.8	8.9
predicted high	3.1	15.4	38.7

4.4.3.4 Predicting Relaxation

Finally, Figure 4.14 and Table 4.9 present the predictive accuracies of all the models we trained to predict "Relaxation". Besides those results, t-tests were performed on the models' performances in pairs. Based on the t-tests, PLS, RF, and GBM outperform SVM. However, no clear decision in a statistically significant way can be made on which of the PLS, RF, and GBM model performs the best.

Since we needed to select one model for our study, we selected the trained RF model as it is the one that delivers the best predictive performance on average. Comparing the predictive performance of RF against the naive model (the model that always predicts "high") where the predictive accuracy is equal to 50.4%, we can observe that our machine learning approach provided a lift of predictive performance equal to 12.4% (62.8% - 50.4%). Table 4.10 presents the confusion matrix of the trained RF model. Table 4.11 reports the 5 most significant features for classifying "Relaxation" using the selected, trained RF model.

4.5 Conclusion

In this chapter, we have presented an approach towards detecting mood states in a non-invasive way. We have presented a mobile application that gathers non-sensitive

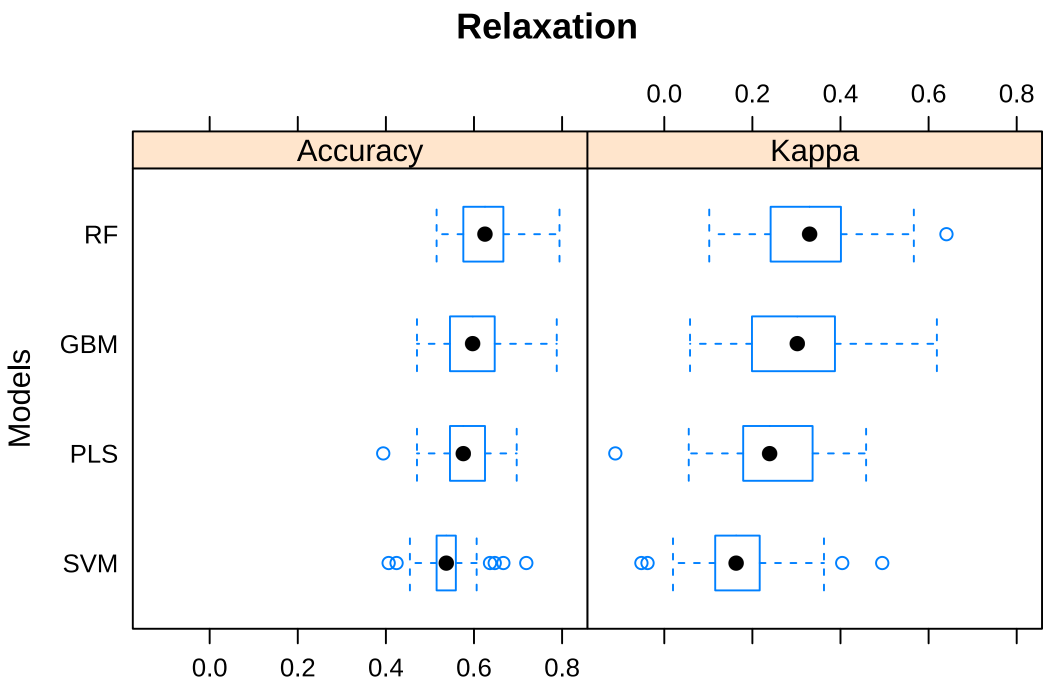

Figure 4.14: Predictive accuracy of multiple classifiers for predicting "Relaxation".

Table 4.11: Most significant features for predicting "Relaxation".

Rank	Feature
1	WhatsApp app
2	Steps
3	Steps sessions
4	Mean steps per session
5	Screen touches

information about user activities and interactions with the smartphone. We organized an experiment in which 4 participants used our application for a period of 4 weeks, sharing their data with us and periodically providing us feedback at different times during the day regarding their subjective perception of their well-being. The initial goal was to collect feedback and to predict stress levels, but inspired by the Circumplex Model of Affect, we have also included feedback responses regarding the perceived activeness and happiness levels of the users. We trained multiple machine learning models to predict all target variables. We achieved predictive performances up to 57.1%, 61.3%, and 62.8% when predicting activeness, happiness, and relaxation (or stress), respectively. Even though our dataset was not of sufficient size to provide definitive results, it is evident from our experiment that it is possible to infer mood states by using smartphone interaction data, without sacrificing privacy.

There are, however, limitations of the presented work. The study focused on the correlation of mood and stress with the smartphone data, and not on causality. It was not evaluated whether a particular behavior was inducing mood and stress variations or if the behavior was the result of those. Moreover, the dataset was created using only 4 male participants of a small age variance, and therefore the collected data are skewed in that sense, so it is a stretch to draw conclusions in general. By having fewer participants in a sample, it is less probable for that sample to represent the general population as a whole closely. Furthermore, each individual's outcome has a larger impact on the overall results. There is still value, however, in smaller experiments such as the one presented, in the sense that they can encourage larger and more expensive studies if the results are promising enough.

5 Anomaly Detection Techniques in Mobile App Usage Data among Older Adults

5.1 Chapter Abstract

We are living in an era of demographic aging, and new technologies that support independent living are constantly being created. In this context, more and more mobile applications are developed for this target group. In this chapter, we are presenting a multidimensional application that targets older adults. We are monitoring the usage of all different aspects of the app, the amount of daily activity in the form of daily steps and the resting time throughout the day from a connected bracelet the user is wearing. Data amounting to 402 user-days of 6 different users are collected. A set of different datasets are manufactured, and various anomaly detection techniques are employed to identify the abnormalities in the datasets. The results demonstrate that clustering can be of use to detect anomalies in the older adults' patterns that could be the trigger of appropriate actions, like informing family members or professional caregivers.

This work has been published as: *'Anomaly Detection Techniques in Mobile App Usage Data among Older Adults'*, Kyritsis, A.I., Deriaz, M. and Konstantas, D., In IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom 2018), Ostrava, Czech Republic, September 2018.

5.2 Introduction and Related Work

After the massive adoption of the Internet, new forms of technology have been developed that have an impact on almost every aspect of daily life, including health. The term eHealth was created to refer to technologies and applications in the service of health and wellbeing [97]. With the latest advances in mobile communications, and with the widespread use of smartphones and connected devices, there has been a high number of health-related mobile applications. These applications can focus on specific medical conditions [98], enable doctors to provide their professional services at a distance [99], and target to promote behavior change for health improvements and disease management [100].

We are living in the ubiquitous computing era where connected devices form the Internet of Things (IoT) and produce data faster than we can logically process. The plethora of sensors that every modern smartphone includes, along with the advances in telecommunication technologies, enable the creation of context-aware applications [101] that power the IoT. Tracking various aspects of wellbeing with mobile applications has become a habit. Numerous consumer electronic devices can continuously monitor users and can assist in healthcare services [102].

Outlier detection, also known as anomaly detection, is a broad domain that has many applications in different fields. Such applications include abnormal behaviour detection in network traffic [103], fraud detection in credit card usage [104], video surveillance systems [105], etc. According to the requirements of each problem, different frameworks [106] and approaches for anomaly detection have been proposed [107].

This study was conducted in the frame of the European Active and Assisted Living (AAL) project named EDLAH2 (Enhanced Daily Living and Health 2) [4]. The goal of the project is to make the usage of smart technology easy and to promote wellbeing and health among older adults. This is achieved with the development of a tablet application targeting this age group. Figure 5.1 presents the home screen of the app.

Figure 5.1: Screenshot of the home screen of the EDLAH2 app.

The functionality of the app includes an easy to manage photo library with photos sent by family members, integrated video/audio communication, a web browser, calendar functionality, and some tablet games.

Additionally, the platform includes a connected wearable device that will monitor health parameters, such as the number of steps and the amount of resting time (minutes of sleep reported by the wearable). The tablet application is reporting app usage and health data to a web server, where family members and professional carers that have permission can view statistics and information. The goal of this chapter is to explore ways of detecting abnormal behavior among older people. To do so, we apply various anomaly detection techniques to data about the usage of the home screen tablet app and the recorded health data.

There have been many studies on detecting abnormal behavior in humans. Body-worn sensors can be used for activity recognition in order to build a model of normal activities [108]. Then, activities that largely deviate can be characterized as abnormal.

Using this approach, however, in uncontrolled environments is extremely difficult, if not impossible, because of the infinite number of activities that should be included in the training dataset and be labeled as normal [109].

Detecting anomalies has also been a topic of interest in computer vision [110]. In relevant studies, human behaviors, motion patterns, and activities are modeled from video footage. Statistical-based methods are used to characterize behaviors, even in crowded scenes [111]. Video-based approaches can be employed to extract useful information in surveillance and public areas. However, due to privacy concerns, and due to the fixed locations of the cameras, these techniques can not be used for applications where constant personalized monitoring is required.

An approach to building a personalized model able to identify abnormal instances would be to track users indoors. In a similar study [112], motion and door sensors were used to track the activity of an older adult indoors. Binary dissimilarity measures, such as the classical hamming distance and the fuzzy hamming distance, are then used to measure the degree of resemblance between activity patterns. When designing a smart home environment [113], such sensors along with electricity power usage and bed/sofa pressure sensors can be installed to monitor the day-to-day activities of the inhabitants.

A similar problem to inferring behavior through the usage patterns of a mobile application is the problem of predicting the next app that will be executed on smartphones [114]. Home screen applications can monitor various spatiotemporal contexts [115], including the time of day, the location of the user and the previously opened app. The app prediction system can then build a personalized prediction model that will exploit the relationships between those attributes and the next application that will be executed by the user. By being able to predict smartphone app usage, one can identify usage patterns and even identify anomalies when significant deviations from the expected patterns are noticed.

In this study, we aim at detecting anomalies in the data that were recorded from the developed tablet application. Those data include app usage logs and activity data from the connected bracelet. Before proceeding with extracting useful features, we examine some approaches that have been developed for anomaly detection.

The rest of the chapter is organized as follows. In Section 5.3 we present various anomaly detection techniques that can be applied for human abnormal behavior detection. In Section 5.4 we discuss our case study along with the evaluation of the abnormal behavior detection techniques that we used. Finally, we conclude our work with Section 5.5.

5.3 Anomaly Detection Techniques

It is generally assumed that the anomalies that need to be detected are scarce in the given dataset. The approach for detecting anomalies can be either supervised or unsupervised, depending on whether the training set is labeled or not. Approaches can be further categorized into discriminative and generative, parametric and nonparametric, and into univariate and multivariate [107].

5.3.1 Dataset Characteristics

5.3.1.1 Univariate Techniques

Univariable methods are the ones that examine one variable a time. A way to identify abnormal observations is to use a variance or standard deviation-based measure. Abnormal cases would be the ones that are several standard deviations away from the mean. This measure is called the z-score. However, this approach is problematic because the anomalies will influence the mean and the standard deviation in the first place, so it is less likely that they will be later identified as anomalies. In the case of data with a normal distribution, for example, roughly 1 every 22 observations will

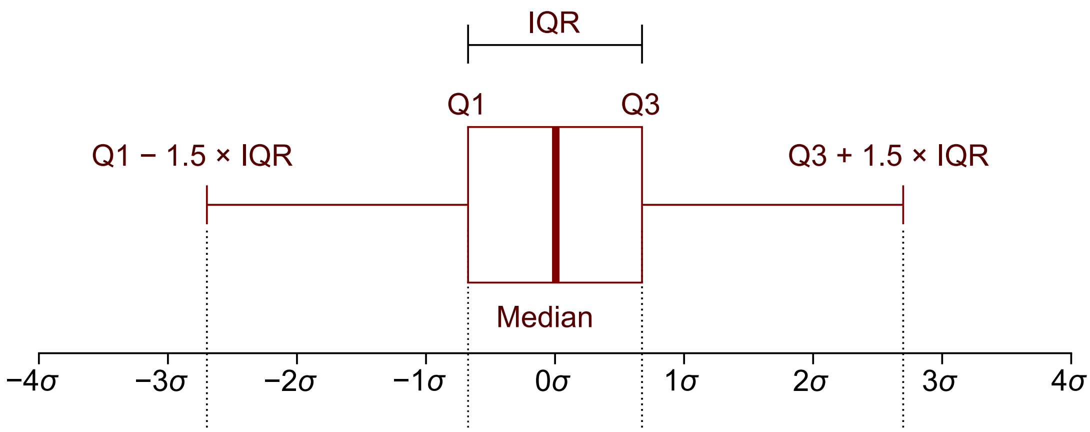

Figure 5.2: Example of a box plot.

differ by at least 2 standard deviations from the mean, and 1 every 370 will deviate by at least 3 standard deviations. Any more than that may imply anomalies in the dataset.

More common techniques use quartiles or percentile-based measures. In those approaches, the distance between each observation from the rest of them is based on the Interquartile Range (IQR) or the middle 50% of the scores. Groups of numerical data are typically depicted with box plots, with an example of such a plot presented in Figure 5.2. The box of the plot represents the area between the first (lower) and the third (upper) quartiles. The bar in the box represents the median, and the whiskers (fences) typically represent a distance of 1.5 IQRs away from the lower and the upper quartiles. IQR is equal to the distance between the upper and lower quartiles. The observations outside the area defined by the whiskers are typically characterized as outliers.

5.3.1.2 Multivariate Techniques

In bivariate approaches, a pair of parameters is examined at a time. Such methods include distance measures, where the distance of an observation is calculated from the center. A more comprehensible approach is the bivariate normal distribution with various confidence regions that are depicted with ellipses over a scatter plot. Observations that fall out of a selected confidence ellipse are the anomalous ones.

Kernel density estimates are topographical maps that follow the density of the data and can have irregular shapes.

Clustering is usually used for unsupervised learning problems. There are many available clustering algorithms, and the best one depends on the application and the characteristics of the available data [116]. A popular algorithm is the k-means clustering one. This algorithm partitions the available dataset into a predefined number of clusters by attempting to minimize intra-partition distances. Some other methods do not need to specify the number of clusters in advance. The mean shift and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) are two such methods. The mean shift algorithm assumes that there is a probability density function and tries to place the centroids of the clusters at the maxima of that function. DBSCAN is an algorithm that assumes clusters in dense regions of data.

5.3.2 Model Characteristics

5.3.2.1 Discriminative Models

A way to approach an outlier detection problem is first to build a model based on a given dataset and then to compute a score for each new observation. The given dataset contains both normal and anomalous observations. The discriminative techniques consist of a similarity function that measures the similarity between two observations. Using this function, the observations are clustered so that within a single cluster the similarity is maximized, while at the same time, the similarity between different clusters is minimized. An observation's anomaly score is defined as the distance of the observation from the centroid of its closest cluster. The parameters of discriminative techniques roughly consist of the definition of the similarity function and the clustering method.

5.3.2.2 Generative Models

In unsupervised generative techniques, a model is trained using a dataset that is known to be clean of anomalous observations. Then for every new observation, the probability of generation of such observation from the trained model is calculated. An observation is categorized as anomalous in case this probability is low, or normal if the probability is high.

5.3.2.3 Parametric and Nonparametric Models

A learning model that can summarize data using a predefined number of parameters is called a parametric model. These models are used in occasions where there is prior knowledge of the problem. This technique may simplify learning but may limit learning capabilities.

On the contrary, nonparametric techniques do not make strong assumptions about the underlying distribution or the form of the mapping function. These methods seek to best fit the training data while being able to generalize to unseen data.

5.4 Case Study

5.4.1 System Overview

A tablet that had the EDLAH2 application installed was given to 6 older adults above 65 years old. The users were also asked to wear an accompanying connected bracelet, the Xiaomi Mi Band 2. All of the participants were living in their own house. The data that have been collected amount to 402 user-days of tablet usage and activity tracking.

The tablet application was periodically sending the app usage information along with data collected from the bracelet to a web server. The collected data are summarized in Table 5.1. Data from the bracelet, i.e., the number of steps and the amount of resting time, were grouped and sent whenever the bracelet synced with the tablet, and a

Table 5.1: Data monitored by the EDLAH2 application.

Category	Content
Browser	Interaction with the web browser
Games	Interaction with a game
Launch_activity	Execution of a feature of the app
Photos	Interaction with the photos feature of the app
Resting_time	Time windows of resting time
Steps	Hourly number of steps

Wi-Fi connection was available. Data about the usage of the different features of the tablet app were sent to the server immediately on every interaction of the user with the tablet. The different app features that were tracked include:

- the number of times the EDLAH2 application was opened,
- the number of times the web browser was opened,
- the number of websites that were browsed,
- the number of times the game center feature was accessed,
- the number of games that were played,
- the number of times the health-oriented features were accessed,
- the number of times the photos feature was accessed,
- the number of photos that were seen,
- the number of videos that were watched, and
- the number of times that the calendar was browsed.

5.4.2 Feature Extraction

The R programming environment was used in order to filter out incomplete days and to process the available data. A number of different ways to build features that will be

Table 5.2: Characteristics of the 4 datasets in test.

Dataset No	No of features	Missing values imputation	Missing values = 0	Steps & resting time	Usage per feature	Aggregated usage
Dataset1	12	✓		✓	✓	
Dataset2	12		✓	✓	✓	
Dataset3	2		✓	✓		
Dataset4	3		✓	✓		✓

later used for clustering were explored. Table 5.2 summarizes the characteristics of the 4 datasets that were built and tested. The data were aggregated on a per-day manner, so each dataset contains 402 entries. Dataset1 includes features about steps, resting time, and the independent usage of every aspect of the app (10 features), amounting in total to 12 features to be used for clustering. For the missing values, imputation was employed using the k nearest neighbors methods with $k = 3$. The missing values solely belonged to app usage classes. The resting time feature accounted for the number of minutes the user was estimated to rest during the day.

Dataset2 had the same number of features as dataset1, but instead of using data imputation, all missing values were set to 0. Dataset3 was formed by keeping only 2 features of dataset2, the step count, and the resting time. Finally, for dataset4 the step count and the resting time features of dataset2 were also used, but instead of independent features for every aspect of the app usage, all those features were combined into a single aggregated app usage feature.

Figure 5.3 presents the histogram of the steps feature, Figure 5.4 presents the histogram of the resting time feature, and Figure 5.5 presents the histograms of all different app features that we are monotoring in dataset1.

5.4.3 Univariate Outliers

Initially, we have proceeded with detecting outliers from independent variables. Figure 5.6 presents the boxplots for each variable of dataset4. An outlier, in this case, is defined as an observation that is located outside the whiskers of the boxplot.

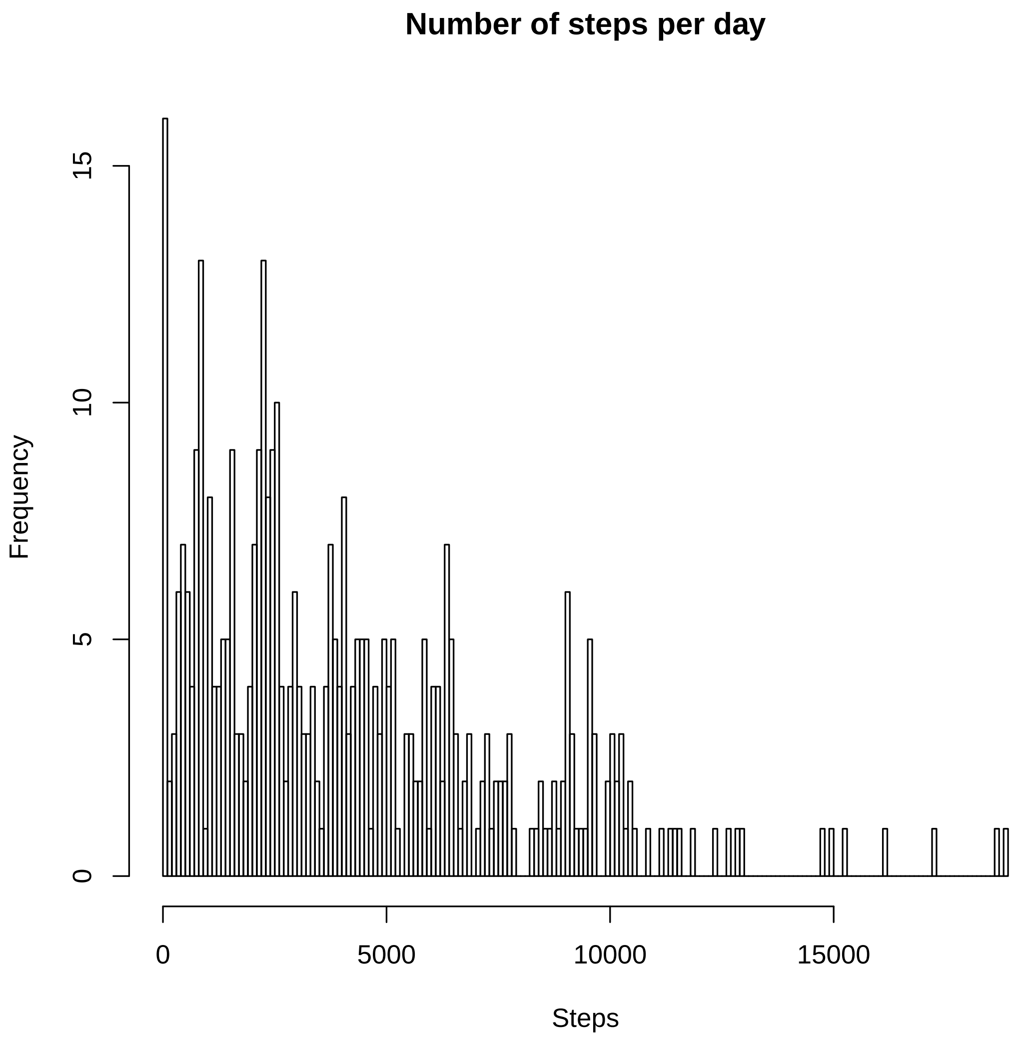

Figure 5.3: Histogram of the steps feature.

At this stage, we have identified that some days had a value of 0 for the steps feature or for the resting time one. This can be attributed to several possibilities. It might have been that the user was not wearing the bracelet, or that the battery was depleted, or that the algorithm measuring the steps and the resting time malfunctioned. However, it might have been a problem that the user had, a case that an alarm should trigger from the platform.

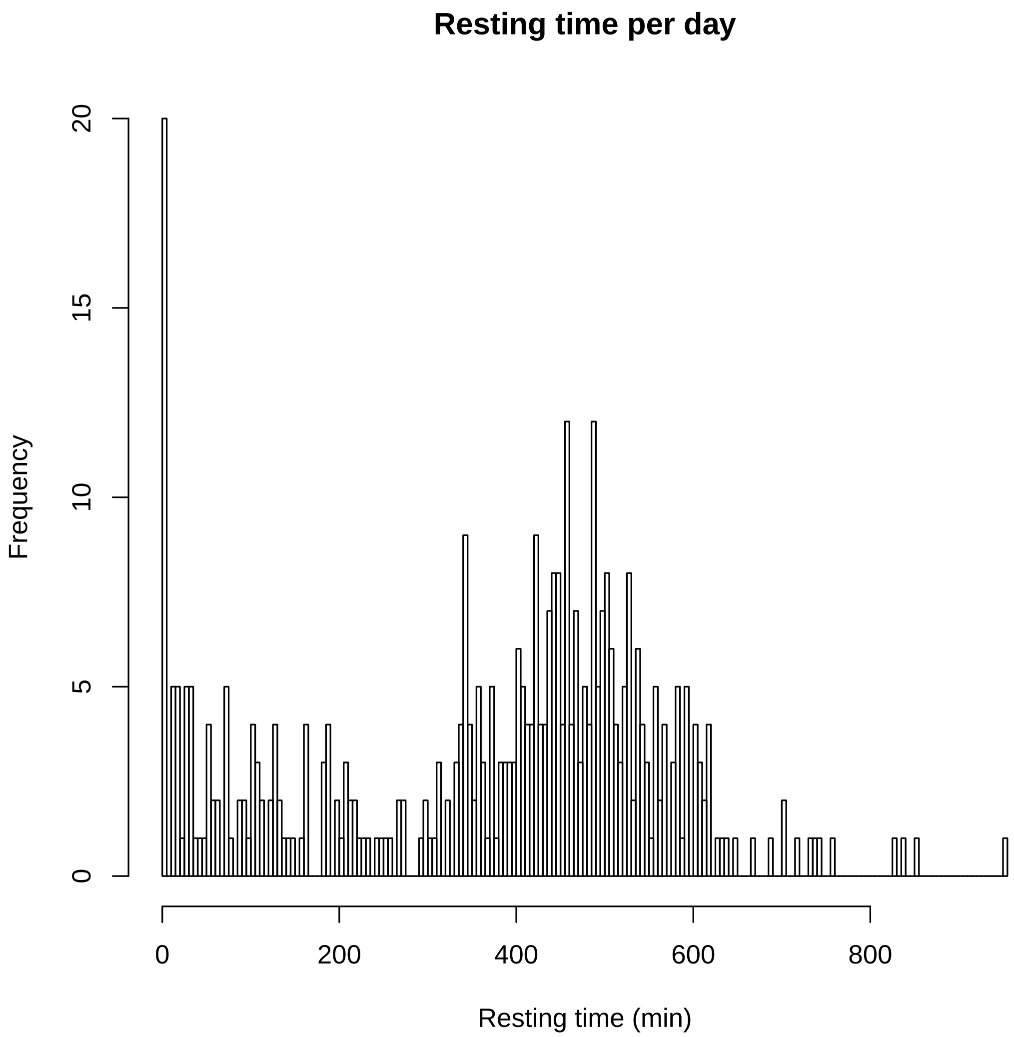

Figure 5.4: Histogram of the resting time feature.

5.4.4 Bivariate Outliers

Our next approach is to evaluate abnormality in our data in pairs. We have selected the bivariate normal distribution confidence ellipse method. Figure 5.7 graphically presents the results from this technique from the features of dataset4 for a selected confidence level of 90%. The observations that are not encircled by the ellipses in the scatterplots are considered to be anomalous. This approach gives a first impression of

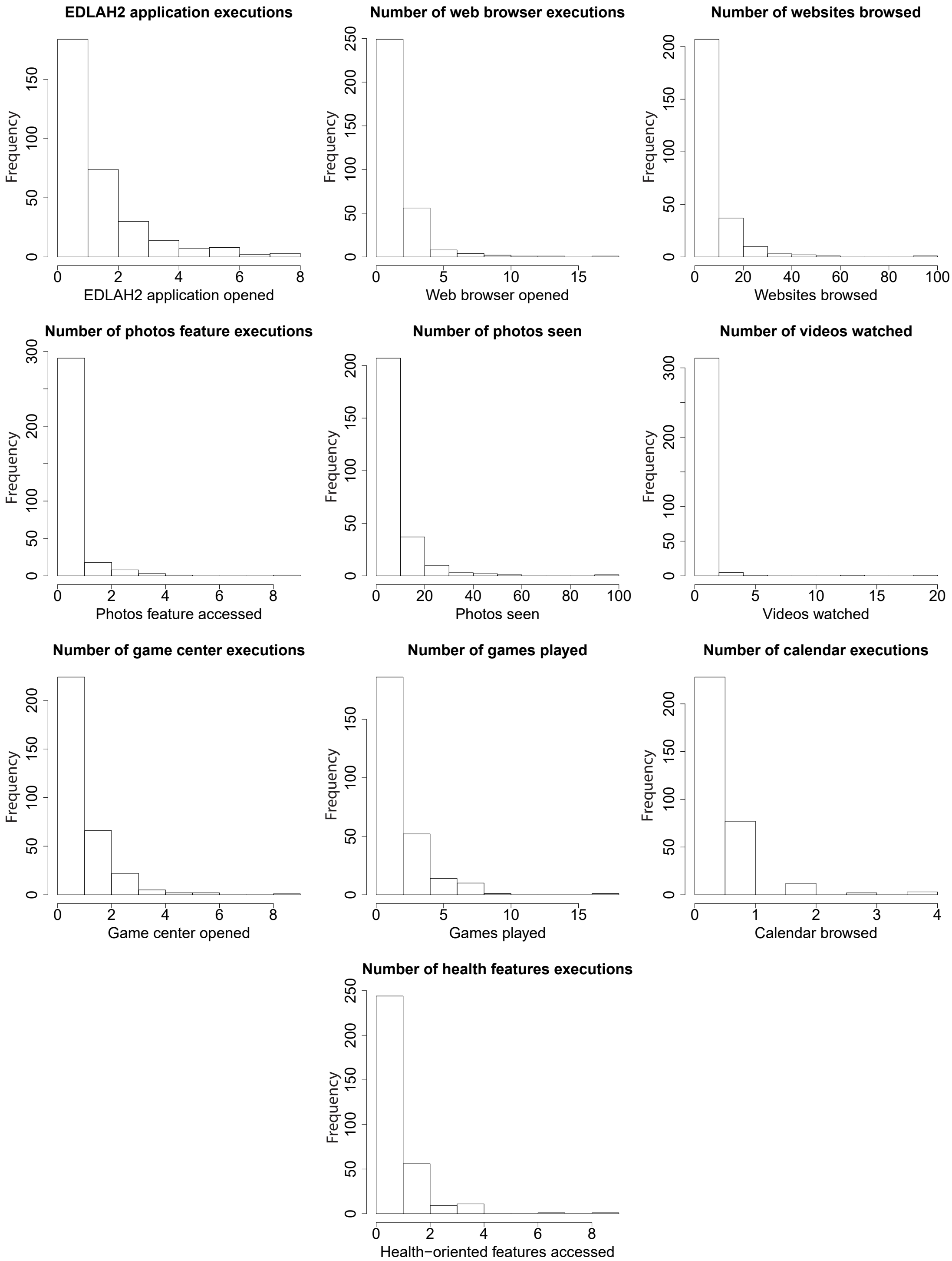

Figure 5.5: Histograms of all the different monitored app features.

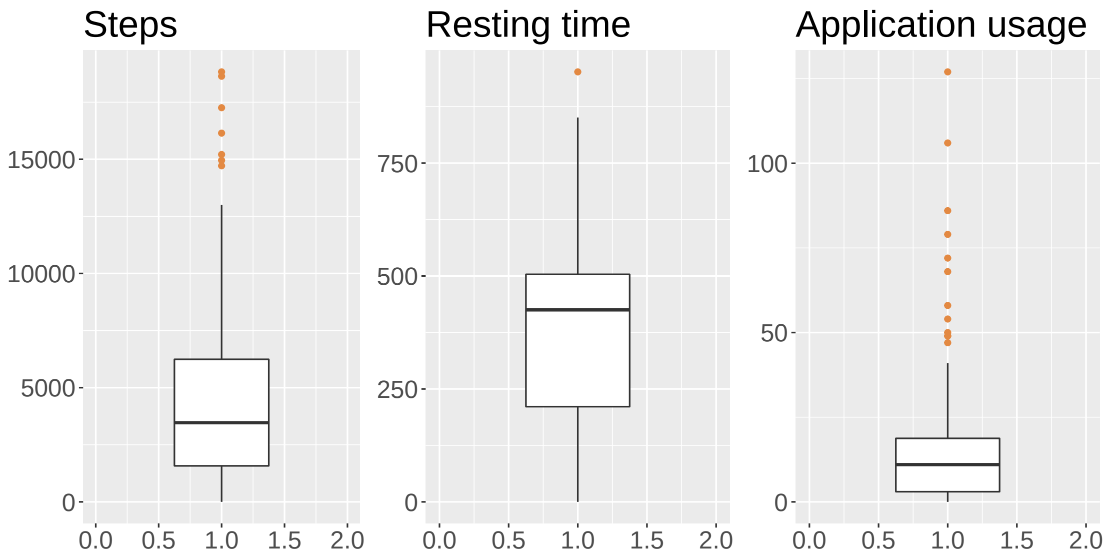

Figure 5.6: Boxplots for each feature of dataset4.

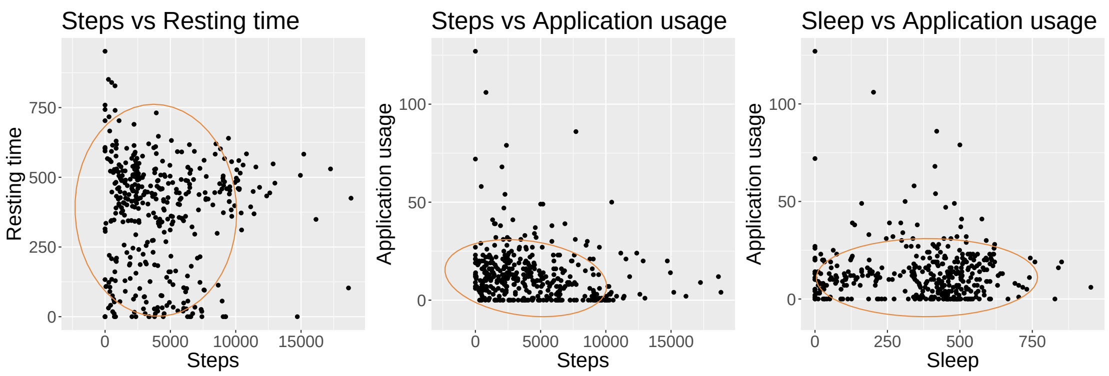

Figure 5.7: Scatter plots with 90% confidence ellipses for each pair of features of dataset4.

how the variables relate to each other and enables the identification of patterns that exist across all variables.

5.4.5 Multivariate Outliers

5.4.5.1 Clustering

We have used different clustering techniques in order to detect anomalies. Theoretically, if daily routine data are clustered together, the odd data will be far from the normal cluster. Also, observations that are near the normal cluster centroid may be considered as normal, while the rest observations that are located far from the cluster centroid as abnormal. The clustering techniques that we have used are the mean shift, DBSCAN, and k-means clustering with $k = 2, 3, 4$.

5.4.5.2 Evaluation

How well a particular unsupervised learning method performs depends on why unsupervised learning is used in the first place. In our case, we are using clustering methods expecting that the majority of the data should be considered as normal behavior and will form a cluster with the observations relatively near the centroid of the cluster. At the same time, the observations of abnormal behavior will be excluded from the normal cluster and may form clusters on their own.

We took the step of manually labeling the dataset by using empirical intuitions that would characterize an observation and thus a day as abnormal. In order to characterize an observation as normal, and therefore safe not to trigger any alarms, we have set thresholds for the steps and the resting time features. We have assumed that during any normal day, the user should have walked more than 100 steps and rested between 4 and a half and 12 hours. This assumption was made together with the caregivers. We expect that our method should be able to detect otherwise so that a carer is informed via an alarm. We did not apply any thresholds to the app usage features. Obviously, days that are categorized as abnormal by this empirical approach may be a result of erroneous samples or days that the user was not wearing the bracelet. However, these might also be occasions that signal an alarming situation. We, therefore, characterized abnormal patterns as those that we assume to be of interest and should attract the attention of the carer.

To evaluate the clustering, we have assumed in all cases that the biggest cluster is the cluster containing the observations of normal behavior, while the rest contain abnormal behaviors. In the mean shift and the DBSCAN methods, the number of

Table 5.3: Sensitivity of the clustering algorithms.

Dataset	Mean shift (%)	DBSCAN (%)	2-means (%)	3-means (%)	4-means (%)
1	47.8 (18)	- (1)	31.7	29.4	65.5
2	36.7 (21)	25.8 (4)	28.3	32.7	33.8
3	56.7 (3)	37.5 (2)	97.5	56.7	44.3
4	41.7 (5)	43.3 (2)	89.2	54.4	52.5

clusters is automatically inferred, while in the k-means clustering the number of clusters is preselected. Standardization was used on all datasets before clustering. For each method, three values were calculated to evaluate the performance of the abnormal behavior detection system, the sensitivity (also called the true positive rate) given by Equation 5.1, the specificity given by Equation 5.2 and the Positive Predictive Value (PPV) given by Equation 5.3.

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{5.1}$$

$$Specificity = \frac{TrueNegatives}{TrueNegatives + FalsePositives} \tag{5.2}$$

$$PPV = \frac{TruePositives}{TruePositives + FalsePositives} \tag{5.3}$$

Tables 5.3, 5.4 and 5.5 include the sensitivity, the specificity and the PPV of all clustering algorithms for all datasets. The number of clusters that were inferred from the mean shift and the DBSCAN methods is displayed in the corresponding parentheses in the tables.

Table 5.4: Specificity of the clustering algorithms.

Dataset	Mean shift (%)	DBSCAN (%)	2-means (%)	3-means (%)	4-means (%)
1	69.9 (18)	67.9 (1)	67.8	67.4	93.4
2	68.6 (21)	66.8 (4)	66.67	68.4	69.9
3	69.9 (3)	68 (2)	95.4	92.5	89.7
4	68.2 (5)	70.1 (2)	95.2	91.3	91.9

Table 5.5: Positive predictive value of the clustering algorithms.

Dataset	Mean shift (%)	DBSCAN (%)	2-means (%)	3-means (%)	4-means (%)
1	17.1 (18)	0 (1)	14.7	15.5	88.4
2	14 (21)	12.4 (4)	21.7	49.6	56.6
3	13.2 (3)	2.3 (2)	89.9	88.4	88.4
4	3.9 (5)	22.5 (2)	89.9	86.8	88.4

We notice that the sensitivity of observing an abnormal day varies across clustering methods and datasets and peaks for the dataset3 and the dataset4 using the 2-means clustering method. The fact that this method performs equally well for both datasets indicates that the single app usage feature had little contribution to the clustering. How 2-means clustering worked for dataset4 of our case is presented in Figure 5.8 for each pair of the available features.

5.5 Conclusion

In this chapter, we have investigated the possibility of detecting abnormal behavior of older adults through monitoring the use of a tablet application along with the activity and resting habits of the users as these were monitored by a connected bracelet. A case study from 6 different users was investigated where 402 user-days were recorded. A set of datasets containing different features were built, and different univariate and multivariate techniques, including clustering algorithms, were evaluated. Some cases have both high sensitivities and specificities according to a rational first manual categorization that we have applied to filter some abnormal instances.

There are, however, limitations of the presented work. We have evaluated our abnormal detection system based on an assumption of the existence of a rational

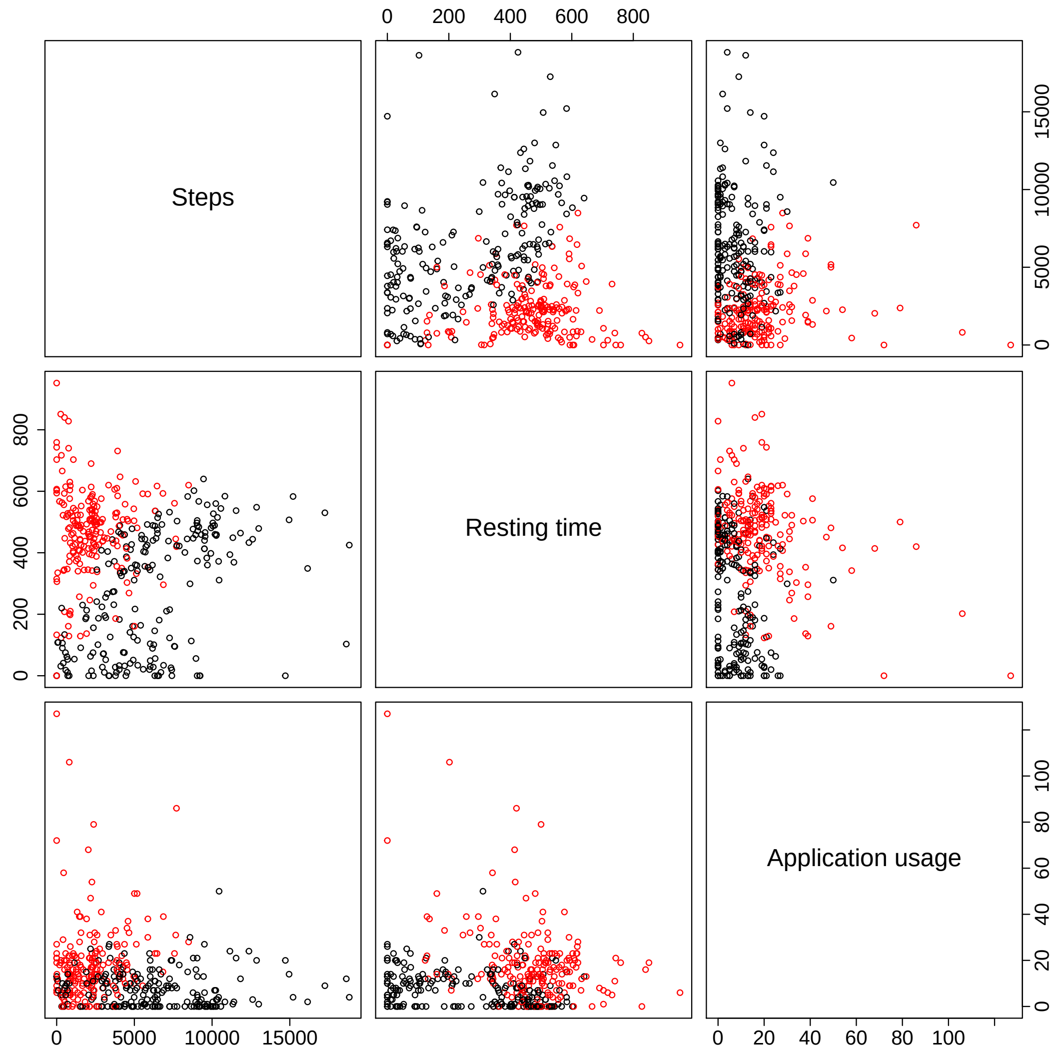

Figure 5.8: Scatter plots of 2-means clustering for each pair of features of dataset4.

number of steps and resting minutes within the day of a participant. This means that other abnormal instances that do not fall into this assumption will be undetected, fact that can not be avoided using an unlabeled dataset.

6 Considerations for the Design of an Activity Recognition System Using Inertial Sensors

6.1 Chapter Abstract

The last decade there has been increasing research interest in the field of human activity recognition in the frame of designing context-aware applications. There are a plethora of parameters that affect the performance of an activity recognition system. However, designers of such systems often either ignore some factors or even neglect their importance. In this chapter, we present and discuss in detail research challenges in human activity recognition using inertial sensors, and we analyze the significance of the existent parameters during the design and the evaluation of such systems. We exemplify the role of the aforementioned parameters with an experiment that was conducted, in which 11 people performed five different activities. Data were recorded from the inertial sensors of a wrist-worn smartwatch. We illustrate how various parameters of the system can be configured and demonstrate how they impact the whole system's performance. This work aims to be used as a concise reference for future endeavors in the field of human activity recognition using inertial sensors of mobile devices in general, and wrist-worn smartwatches in particular.

This work has been published as: *'Considerations for the Design of an Activity Recognition System Using Inertial Sensors'*, Kyritsis, A.I., Deriaz, M. and Konstantas, D., In IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom 2018), Ostrava, Czech Republic, September 2018.

6.2 Introduction and Related Work

Over the past decade, there have been great efforts towards Activity Recognition (AR) methods and techniques both by researchers and the industry. There are many applications that either require or would benefit from AR. Healthcare monitoring systems use sensors to track Activities of Daily Living (ADL) of older adults and assist the work of caregivers [117]. Besides the healthcare sector, other domains that benefit from AR include sports [118], entertainment [119] and the industrial sector [120].

Several commercial products rely on AR. All major video game console manufacturers have developed such systems. Nintendo recognizes gestures using the inertial sensors of the handheld controllers starting from the Nintendo Wii console [121]. Microsoft identifies activities by monitoring full-body movement using the Kinect camera [119]. Sony uses inertial sensor data from the controllers and tracks them in space using a camera with the PlayStation Move system. All these systems, while initially developed for entertainment, have also been used by researchers for rehabilitation purposes [122]. Modern Virtual Reality (VR) consumer products, like the HTC Vive, use both Inertial Measurement Units (IMU) in handheld controllers and cameras that detect user gestures and activities. Although these devices have initially targeted gaming, researchers have exploited their capabilities so that they can be used in other domains, like in health applications [123].

The proposed approaches for AR systems can be roughly divided into two categories, the inertial sensor-based ones [124], and the camera-based ones [125]. In the sensor-based methods, one or more inertial sensors, such as accelerometers and gyroscopes, are attached to the human body. Time-series techniques are applied to the collected signals to extract useful information. In the camera-based methods, different computer vision techniques are employed to obtain analytical results.

Camera-based AR methods use video sequences recorded by video cameras to detect human gestures and activities. AR is an important domain of research in computer vision, and its applications include patient monitoring systems and video surveillance systems [126]. Processing and feature extraction from raw videos target the finding of specific characteristics such as colors, shapes, and body motions that can describe human activity. These features can also be used for body model reconstruction [127]. Despite all the progress made for AR using vision-based methods, they pose particular limitations. They are intrusive and thus can not be used in applications where privacy is a requirement. Moreover, due to the fixed locations of the cameras, these techniques can not be used for real-time applications where constant monitoring is required.

Inertial sensor-based AR techniques overcome these last limitations. The increased availability of such sensors due to the omnipresence of smartphones and smartwatches has enabled the use of AR techniques in ubiquitous computing. Sensor-based AR systems are either knowledge-driven or data-driven. Knowledge-driven approaches use prior domain knowledge to build an abstract model and apply the model to the recorded data [128]. On the contrary, data-driven approaches work by extracting correlations between data and gestures and eventually build a model for classification [124].

While there are already a lot of works and applications of AR, very few of them [109] discuss the parameters and the choices that impact the performance of an AR system. Usually, those works present a single application specific best solution. The scope of this chapter is to discuss the various design parameters that are crucial when performing AR. An AR system is developed to show how different variables impact the performance of it. Developing the optimal AR system would require a massive dataset, a lot of experiments and computing power for parameter tuning, and is out of the scope of this work. There is a large variety in every step of creating an AR system, starting from the data acquisition and the feature engineering up to the training and classification phases, and every single step can significantly impact the system as a whole. The research question of this chapter is, "What are the main considerations in the design of an inertial sensor-based activity recognition system, and what are the tradeoffs?"

The rest of the chapter is organized as follows. In Section 6.3 we discuss the considerations and the parameters that affect every AR system in the design, implementation, testing, and evaluation phases. In Section 6.4 we present the experiment that we have conducted, and in Section 6.5 we evaluate a variety of tests and show how each parameter tuning may impact the performance of the overall system. Finally, we conclude the chapter in Section 6.6.

6.3 Design Considerations

There are many challenges when designing an AR system. In this section, we are presenting a list of considerations and parameters that impact every AR system during the design, implementation, testing, and evaluating phase. Figure 6.1 summarizes the considerations that are further discussed.

6.3.1 Data Characteristics

6.3.1.1 Intra-class Variations

The first challenge of any AR system is to be robust to intra-class variations. Intra-class variations mainly exist because the different activities are performed differently by various people. For example, if inertial sensors are attached to the wrists of two people running, it is very improbable that we observe a similar swing in both cases. Intra-class variations can also exist among activities performed by the same person. For example, when a person is walking to catch a bus and when the same person is taking an after-work walk. Theoretically, the prediction from both cases should belong to the "walking" class, but in practice, also depending on where the inertial sensors are placed, the recorded data might significantly differ. To tackle with this issue, the designer of the system should capture a significant amount of training data, both from a single person but also from several others, for the dataset to capture as much

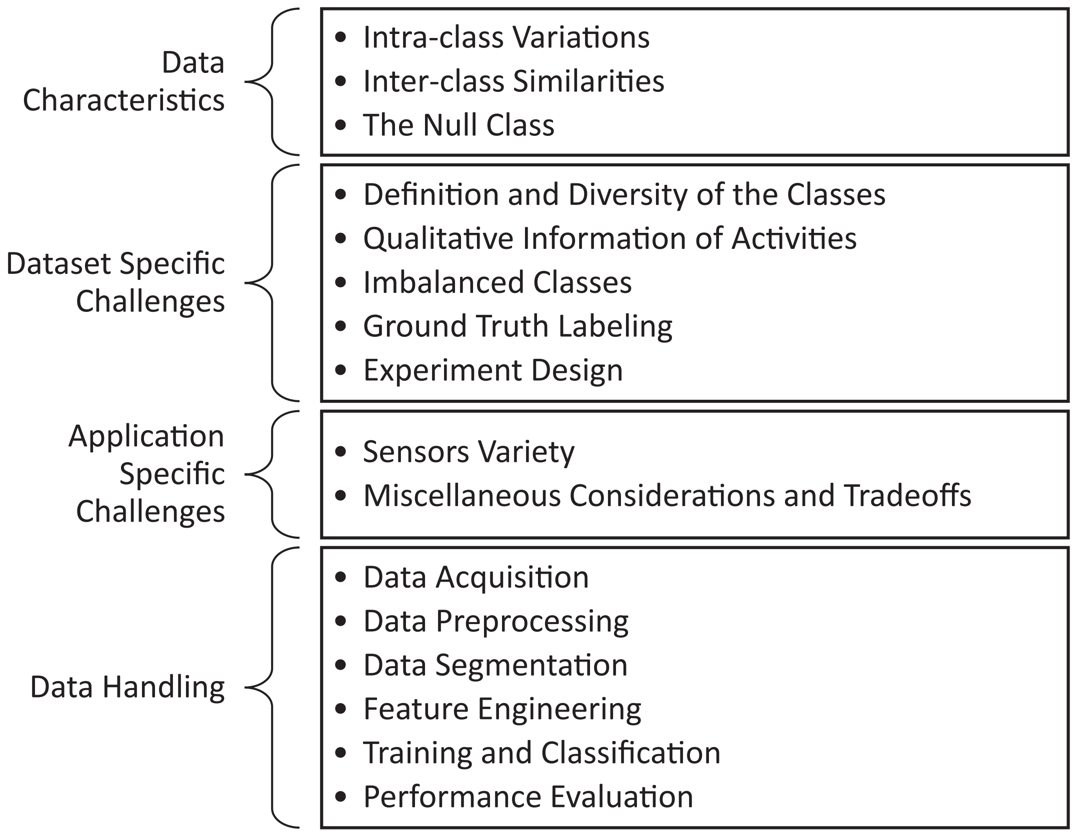

Figure 6.1: Summary of the design considerations of an activity recognition system.

variability as possible. It is crucial for a model to avoid overfitting and be robust across different people.

6.3.1.2 Inter-class Similarities

Another challenge arises when a system aims to identify classes that are different, but the recorded sensor data show very similar characteristics. An example would be to differentiate between the activity of walking with a stroller and the one of walking with a shopping cart. To tackle this issue, the designer can either take more and different sensors into account or observe other correlated activities that take place in parallel [129].

6.3.1.3 The Null Class

In a continuous data collection setting, like in an all-day running AR system, the indifferent to the designer classes for identification may be more and may form the majority of the collected dataset. In that case, the irrelevant activities that have similar characteristics to the relevant ones form the null class. The null class is extremely hard to model since it encapsulates an arbitrary and theoretically infinite number of activities. The simplest way would be to identify the null class with the data patterns that significantly differ from those of the desired activity. Other methods like self-training [130], may allow the designer to use the null class during the training of the AR model.

6.3.2 Dataset Specific Challenges

6.3.2.1 Definition and Diversity of the Classes

When designing an AR system, it is essential to define the classes of the activities that are of interest to the specific application. Although this task seems to be of little importance, it can be tricky because human activities are highly diverse, can be performed in many ways and sometimes even the definition of them can vary. There have been researches on the definition of a taxonomy of activities [131] that can prove a good reference for AR systems designers.

6.3.2.2 Qualitative Information of Activities

While the vast majority of research for AR systems aims at detecting the activity that is being performed at a specific moment, little progress has been made on the extraction of its qualitative information. It would be interesting, for example, for physiotherapists to be able to know if their patients are executing the prescribed activities correctly, and if not, to understand the source of the problem. Such research has so far only targeted the sports sector, and the settings were too constrained [132].

6.3.2.3 Imbalanced Classes

Another challenge of any AR system arises when the classes to be modeled do not exist in similar quantities in the training dataset. This problem is profound in long-term monitoring settings because only a few activities take place frequently, e.g., walking, while others scarcely, e.g., doing squats. Possible ways for a researcher to ameliorate this situation include the recording of additional training data of the underrepresented classes, the generation of simulated training data, and the oversampling the smaller classes [133].

6.3.2.4 Ground Truth Labeling

A necessary, laborious, and time-consuming task for all supervised AR systems is the annotation of the training data. Post hoc annotation is possible for data captured from cameras by labeling the footage but is difficult to achieve with inertial sensor data. In laboratory settings, the researcher can annotate the data in real time, but in daily life situations, the user has to label the data with the ground truth with techniques like the experience sampling method [134].

6.3.2.5 Experiment Design

An essential aspect of any AR system is the data collection and the overall design of the experiment. So far, there are only a few general purpose datasets [135] that can be used for activity recognition, and there is no commonly agreed way to collect data. The recorded data depend on the designer of the experiment, and usually, every study has different priorities, such as a large number of participants, a large dataset, or clean data. Datasets are also not always publicly available to be reused in other experiments. To be able to have comparable and reproducible scientific results and focus more on the methods for data analysis, it is crucial for the scientific community to commonly agree on some standard data acquisition guidelines and datasets.

6.3.3 Application Specific Challenges

6.3.3.1 Sensors Variety

There is considerable variability in the available sensing equipment. Every sensor has different specifications given by the manufacturer of it, such as sampling frequency, accuracy, precision, and operating temperature range. Moreover, mobile devices can be used in different ways. For example, for the same activity, one would expect different recordings from the inertial sensors of a smartwatch worn on the wrist than from a smartphone kept in the pocket.

Smart devices embedded three-axis accelerometers and three-axis gyroscopes are most commonly used in inertial sensor-based AR systems [136]. Fusing both accelerometer and gyroscope data usually leads to a better recognition performance than when only using a single source of data. A three-axis magnetometer can also be used in conjunction with the sensors mentioned above in order to optimize the detection of the orientation of the user in space [137]. A nine-axis inertial sensor refers to a three-axis accelerometer, a three-axis gyroscope, and a three-axis magnetometer enclosed in a single module.

6.3.3.2 Miscellaneous Considerations and Tradeoffs

There are many tradeoffs that each AR system designer should consider. Some computer applications that rely on gesture recognition should run in real time, while for others such as monitoring long-term behavior, for example, an offline analysis may suffice. There are also tradeoffs associated with the accuracy, the power consumption, and the latency of the system. In case the application runs on a mobile device, power efficiency should be taken into account sometimes even opposed to the accuracy and the latency of it [138].

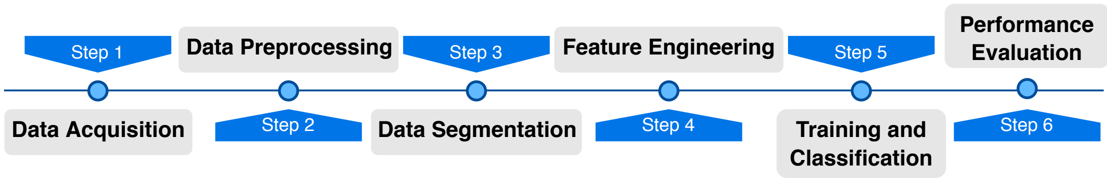

Figure 6.2: Summary of the data handling steps of an activity recognition system.

6.3.4 Data Handling

6.3.4.1 Data Acquisition

Figure 6.2 summarizes the data handling steps that are further discussed. The first step in any IMU-based AR system is to capture raw data using different inertial sensors, attached to different locations of the user's body. There are also advanced systems that use even more sensors in the environment to record additional data [139]. Some sensors provide multiple values, like for example the accelerometer that gives three recordings, one for each of the x, y, and z-axes. Each sensor is sampled at regular time intervals, typically every 1-10 milliseconds, and the recorded raw data correspond to a multivariate time-series dataset. The sampling frequency is different per sensor and sometimes can be set according to the requirements of the application, e.g., for power saving.

The collected sample size is an important feature of any empirical study. Sample size determination is the procedure of choosing the right number of observations that will allow inferences about a population from a sample. The required sample sizes may be calculated by knowing the population size, the margin of error, the confidence level, and a target variance.

6.3.4.2 Data Preprocessing

Before proceeding to extract features from the available raw dataset, the data need to be preprocessed and cleaned. Among the raw data, there might be artifacts caused by electromagnetic interference that need to be filtered out. Also, data streams from different sensors should be synchronized at that point. In case the streams have a different sampling rate, they should be resampled with the same frequency. Data may also be calibrated according to the sensor characteristics and normalized. Data preprocessing is a generic step, and the same actions should be applied to all input data with no exceptions [140].

6.3.4.3 Data Segmentation

In this part, we are identifying the segments of the preprocessed data that contain useful information about the activities to be detected. Each segment has a duration and is defined by the start time and the end time within the time-series dataset. It is hard to segment the dataset ideally because in daily life, there is not always a concrete pause between two consequent activities, and the boundaries are tough to define.

A widely used data segmentation method is the sliding time window one. In the sliding window approach, a window of a specific width is moved across the data and defines the start and the end of the segment. The higher the width, the higher the lag, since the AR system has to "wait" for a specific amount of time before a segment is full. The optimal size of the window is not known, is inferred during the testing phase and can influence the performance of the system [141]. Another variable of the segmentation phase is the step size. While a small step size will increase the number of segments and create some that potentially better contain information about an activity, it will also increase the computing load of the application, since some entries should be computed more than once.

Another way to segment the preprocessed data may take advantage of the fact that different activities have different intensities, and so the energy level of the IMU signals are distinctive. Other methods for data segmentation include the definition of either a rest position or a specific gesture [142]. Finally, segmentation can occur using external sources, like calendar entries with the start and duration of different activity sessions.

6.3.4.4 Feature Engineering

This step is about deriving features from the raw time series data. These features will form the feature vector that will be used for machine learning. We need to create features in order to reduce the dimension of the initial dataset, and therefore, to train faster and more cost-effective models. Depending on the features, they can be either extracted on the segmented windows or the entire activity. The most widely used features in AR research consist of signal based ones. These can be either time domain based statistical ones like the mean, the median, and the variance, or frequency-domain features, like the energy in specific frequency bands [143]. The more features, the more training data are needed to classify activities more accurately, and the more computational resources are required for classification. On real-time, mobile systems, and edge computing, for example, it is imperative to use a minimum amount of features that do not significantly degrade the performance of the system. There are methods to automatically reduce the dimensionality of the feature space by ranking and selecting the most important features [144].

6.3.4.5 Training and Classification

Before predicting activities based on newly recorded sensor data, we must first train the selected model. For supervised learning, a training set is needed that consists of N entries of feature vectors X with corresponding output labels y. The selected model is defined by a parameter set θ, which during the training phase is learned to minimize the classification error on the training set. Then for classification, the selected model with the trained parameter set θ is used to map a feature vector X to a set of confidence values y, that corresponds to the scores for every class that exists in the class set. With the scores vector and using either confidence thresholds or multiobjective optimization techniques, a single prediction class is selected by the trained model.

Over the years, machine learning researchers have proposed a large number of algorithms. Many of those have already been used by the AR community to classify activities and to solve application-specific problems. The used methods include among others Hidden Markov Models [145], dynamic Bayesian networks [146], Decision Trees [147], Support Vector Machines [148], etc. There have also been researchers that have implemented deep learning neural networks techniques [149] for AR. So far, no research derives one or a set of best machine learning algorithms for AR. The decision on that depends on the characteristics of the data. For example, if the dataset consists of many observations in a low dimensional space, then even the k-Nearest Neighbors (kNN) method may perform sufficiently well, while in other cases a more complex model may be necessary. There are also techniques to fuse the results of multiple classifiers to create a model that performs better than the submodels that it consists of [150].

6.3.4.6 Performance Evaluation

The last step when training a model for AR is to evaluate its performance [151]. Various performance metrics can be used for evaluation, metrics such as True Positive (TP) and False Positive (FP) rate, precision, accuracy, F-scores, and recall. The confusion matrix summarizes how many instances of the training set, either in absolute terms or a percentage, were correctly classified and how many were not. The Receiver Operating Characteristic (ROC) and the Precision-Recall (PR) curves also provide an insightful view of the predictive performance of the model.

The evaluation scheme that is typically used to evaluate AR models is the k-fold cross-validation scheme. According to this, the training set is partitioned into multiple folds. All of them but one are used to train the model, and then the remaining fold is used to evaluate the performance of the model. This procedure is repeated until all folds have been left out once. The performance results are then averaged to evaluate the performance of the predicted model. A hold-out strategy can also be used, where the model is trained once on a percentage of the available dataset and then evaluated on the rest of it.

6.4 Experimental Design

We conducted an experiment in which we will showcase how different design decisions of an AR system compare and impact the overall performance of the system. We focused on recognizing mobility activities from wrist-worn inertial sensors. Activity tracking applications running on smartphones use this kind of AR. At this point, however, we should note that in real life scenarios the problem of AR appears to be more demanding mostly because of the noise in the data, and the variability of the way the activities can be performed.

6.4.1 Setup

We recorded the wrist movements of 11 male participants of ages 25-35 that performed five different mobility activities. The mobility activities include walking, running, idling in the office, going up and going down the stairs. Each participant was asked to perform each of the five activities for around 35 seconds four times in total. With this, we guaranteed that the recorded dataset would be balanced. For the going up and going down the stairs activities, the 35 seconds target per recording could not always be satisfied, because the time of going up or down the available staircase varied according to the pace of each user. The resulting total dataset was roughly 128 minutes.

6.4.2 Sensors and Data

Wrist measurements were recorded using the inertial sensors of the Sony SmartWatch 3 watch running Wear OS. The watch was worn on the dominant hand of every participant, which was the right for ten participants, and the left for one of them.

The available IMUs were a three-axis accelerometer and a three-axis gyroscope. All recordings were timestamped, and the sampling rate for the accelerometer and the gyroscope was 150 Hz. During all recordings, the author of this thesis was observing and instructing the participants in order to correctly annotate with the ground truth and guarantee the cleanliness of the data. We used Matlab for feature extraction from the time series data and Python with the Scikit-learn module [152] for the machine learning experiments.

6.5 Evaluation

As previously discussed, there are many components of an AR system that can be implemented in a variety of ways, and each such decision impacts the overall performance of the system. In this section, we will methodically evaluate our system with a plethora of choices regarding its parameters. We have searched up to a certain extent, also considering that all steps are interdependent and need to be configured jointly to achieve optimal results. This challenge becomes even more prevalent in real time AR systems that need to be regularly optimized based on user feedback and need to adapt continuously.

Since the IMUs that we have used did not provide a constant sampling rate throughout the recordings, the raw sensor data were resampled with a sampling frequency of 60 Hz. This frequency was selected for this study as it is higher than the 20 Hz commonly required to assess daily living [153], and also lower than what typical off-the-shelf IMU components can achieve.

Both time and frequency domain features were computed for both sensors. The time domain features include the mean, the standard deviation, the median, the skewness, the kurtosis, the 25th and the 75th percentile, and the squared sum of the components under the 25th and the 75th percentile. Those were derived from the resultant vector. For the frequency domain features, a Fast Fourier Transform (FFT) was performed after normalization on the windows, and the features were computed per axis. Those

Table 6.1: List of extracted features per sensor.

Domain	Group	Features	No of features
Time	1	Mean Standard deviation	2
	2	Median Skewness Kurtosis 25th percentile 75th percentile Sq. sum of < 25th perc. Sq. sum of < 75th perc.	7
Frequency	3	Maximum frequency Sum of heights < 5 Hz Number of peaks < 5 Hz	9

features include the maximum frequency, the sum of heights of frequency components below 5 Hz and the number of peaks in the spectrum below 5 Hz, as it was noticed that most of the signal strength lied between 0-5 Hz. All the features extracted for this study are summarized and grouped in Table 6.1. Group 1 includes two basic time domain features, Group 2 contains the rest of them, and Group 3 includes all the computed frequency domain ones.

To evaluate the performance of our systems, we split the available dataset into a training set (80%) and a test set (20%). The 10-fold cross-validation scheme was used on the training set in order to train the model, the performance of which was evaluated on the aforementioned test set. The reasons behind using this setup were the following: We used cross-validation on the training set for hyper-parameter tuning. However, to correctly measure the performance of the final model, we needed another independent sample, that is the test set since all information of the training set has become part of the trained model.

6.5.1 Machine Learning Algorithms and Person Dependence

For the initial test, we fed the features of Group 1 of both the accelerometer and the gyroscope sensors, so four features in total, into multiple machine learning algorithms.

Table 6.2: Confusion matrix of the person-independent kNN classifier.

		Predicted class				
		Down	Idle	Up	Walk	Run
True class	Down	84.9%	0%	10.7%	4.4%	0%
	Idle	0%	100%	0%	0%	0%
	Up	7.5%	0%	86.7%	5.8%	0%
	Walk	4%	0%	6.3%	89.7%	0%
	Run	0%	0%	0%	0%	100%

The classifiers that we have evaluated are Logistic Regression (LR) [154], Support Vector Machines (SVM) [95], Random Forest (RF) [92], Decision Tree (DT) [155], Naive Bayes (NB) [156], Extra Trees (ET) [157] and k-Nearest Neighbors (kNN) [158]. The features were computed over a time window of 5s with a step size of 1s, so there was a 4s overlap between consecutive windows. This value was used for the time window as it is large enough to contain useful information regarding the performed activity. The step size was set to 1s as it is small enough to increase the number of the produced segments during segmentation.

We tested both a user-independent and a user-dependent approach. Figure 6.3 presents the box plot for all trained classifiers for the user-independent tests. Table 6.2 presents the confusion matrix for the kNN classifier, the best performing classifier for this case. We notice that the idling and the running activities achieve a perfect classification rate, even with a few features and a simple kNN classifier. We also notice that the classes that mostly suffer from misclassification are the going down and up ones, with misclassification rates of 15.1% and 13.3% respectively.

Table 6.3 summarizes the characteristics of all the features of the computed dataset for the user-independent test. Accelerometers measure linear acceleration in meters per second squared (m/s^2), and gyroscopes measure rotational motion in radian per second (rad/s). The mean, min, and max columns are self-explanatory. Note that there were no null values. The std column shows the standard deviation that measures how dispersed the values are. The 25%, 50%, and 75% columns show the corresponding percentiles. A percentile indicates the value below which a certain

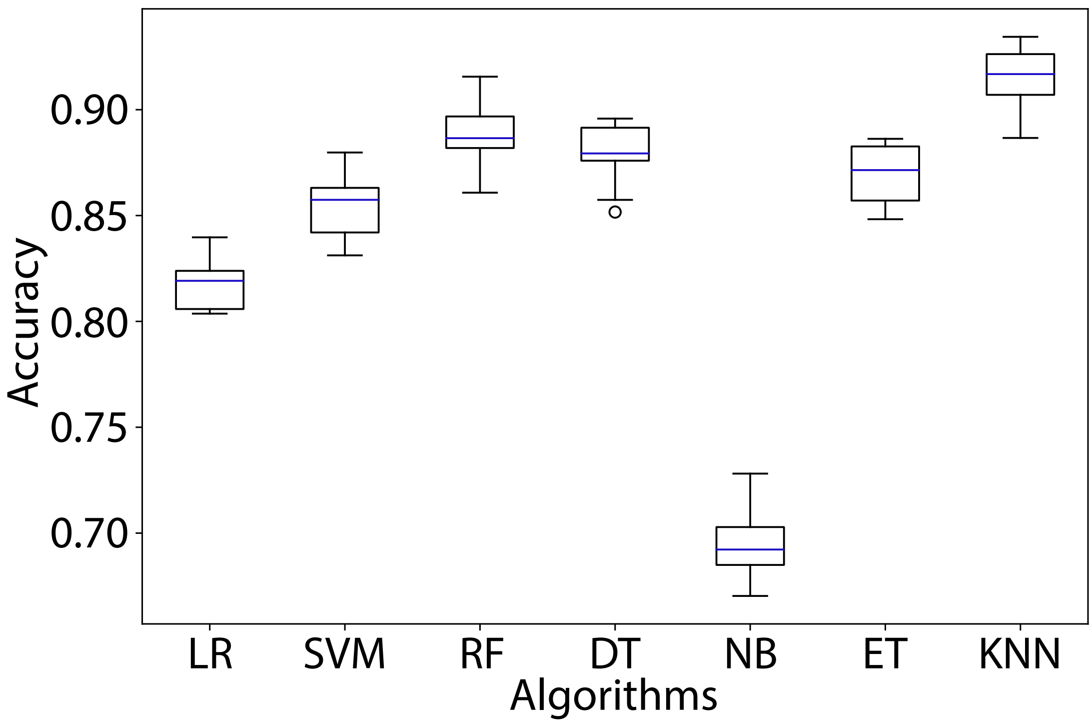

Figure 6.3: Box plot of multiple person-independent classifiers.

Table 6.3: Dataset summary for the user-independent test.

Feature no & name	Mean	Std	Min	25%	50%	75%	Max
1 mean Acc	5.39	4.60	0.06	2.69	4.06	6.21	19.43
2 std Acc	1.85	1.78	0.02	1.1	1.65	2.41	6.62
3 mean Gyr	1.7	1.09	0	1.01	1.54	2.49	6.24
4 std Gyr	0.84	0.49	0	0.54	0.79	1.16	3.91

percentage of observations among the whole set of observations fall. The value of the 50%, therefore, corresponds to the median.

For the user-dependent approach, the results of all the 11 users for all selected classifiers are presented in Table 6.4. As expected, personalized models have on average a better accuracy than person-independent ones, because they adapt to a specific person. There is even a case that a perfect classification rate was achieved (user 6, ET classifier). However, as in the case of user 3, when intra-class variations exist, that is when the user does not perform every activity in a consistent way, the model does not perform equally well even though it is a personalized one.

Table 6.4: Accuracy of multiple person-dependent classifiers.

User	**Algorithms**						
	LR	SVM	RF	DT	NB	ET	KNN
1	89.7%	96.4%	97%	97%	90.3%	98.2%	95.2%
2	95.5%	97.7%	97%	95.5%	93.9%	97%	97%
3	78.1%	83.2%	90.5%	85.4%	79.6%	86.1%	86.1%
4	87.3%	90.3%	89.6%	88.1%	79.9%	92.5%	91.8%
5	96.3%	97.8%	97.8%	94.9%	94.9%	95.6%	98.5%
6	97.7%	99.2%	99.2%	99.2%	96.9%	100%	99.2%
7	98.5%	99.3%	98.5%	98.5%	95.6%	98.5%	97.8%
8	96.3%	97%	94%	92.5%	91%	97%	97.8%
9	82.5%	81.7%	84.1%	86.5%	69%	83.3%	85.7%
10	92.6%	95.6%	93.3%	89.6%	81.5%	94.8%	91.9%
11	91.3%	96.4%	96.4%	94.2%	94.9%	95.7%	94.9%
Avg.	91.4%	94.1%	94.3%	92.9%	88%	94.4%	94.2%

Table 6.5 summarizes the characteristics of all the features of the computed datasets for the user-dependent tests.

6.5.2 Sliding Time Window Size

One of the parameters selected during the feature extraction phase is the size of the time window. To investigate how the time window size affects the performance of our AR system, we swept the time window for the values $T_w = 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20\ s$. We run all the classifiers for all these cases using the same four features of Group 1 of Table 6.1 for both sensors. On all tests, the step and thus the overlap between two consecutive windows was equal to half the time window size. For example, for the time window of 0.5s the selected step was 0.25s, for the time window of 1s the overlap was 0.5s, and so on. The results are presented in Figure 6.4.

Naturally, as the time window increases, the available observations in the observations set that can be extracted from the raw data decrease. We notice that the time window size affects the recognition performance of the system, but up to a point. When the

Table 6.5: Datasets summary for the user-dependent tests.

User	Feature no & name	Mean	Std	Min	25%	50%	75%	Max
1	1 mean Acc	4.29	4.08	0.08	1.99	2.62	4.35	13.11
	2 std Acc	1.55	1.36	0.04	0.86	1.05	1.23	5.18
	3 mean Gyr	1.12	0.58	0.02	0.74	1.11	1.63	2.02
	4 std Gyr	0.6	0.24	0.01	0.45	0.61	0.81	1.22
2	1 mean Acc	5.37	4.56	0.08	3.54	4.16	5	15.17
	2 std Acc	1.74	0.96	0.04	1.29	1.7	2.58	4.42
	3 mean Gyr	1.43	1	0.01	1.01	1.28	1.53	3.82
	4 std Gyr	0.73	0.52	0.01	0.48	0.61	0.77	1.95
3	1 mean Acc	5.75	4.61	0.06	3.75	4.64	6.17	16
	2 std Acc	1.89	1.2	0.02	1.56	1.77	2.06	5.05
	3 mean Gyr	2.11	1.12	0.01	1.88	2.3	2.84	3.99
	4 std Gyr	0.97	0.48	0.01	0.85	1.03	1.25	1.9
4	1 mean Acc	6.09	4.14	0.14	4.63	5.52	6.77	15.11
	2 std Acc	2.01	1.12	0.09	1.62	1.9	2.35	4.4
	3 mean Gyr	2.18	0.99	0.07	2.19	2.5	2.85	3.46
	4 std Gyr	1.03	0.39	0.06	0.96	1.12	1.3	1.63
5	1 mean Acc	6.68	5.2	0.11	3.35	5.58	7.63	19.43
	2 std Acc	2.04	0.84	0.06	1.51	2.06	2.56	4.64
	3 mean Gyr	2.29	1.17	0.06	1.71	2.3	3.06	4.68
	4 std Gyr	1.06	0.44	0.05	0.77	0.99	1.43	2.07
6	1 mean Acc	5.24	4.91	0.14	2.92	3.81	4.43	16.6
	2 std Acc	1.78	1.14	0.02	1.23	1.42	2.12	4.4
	3 mean Gyr	1.46	0.94	0.01	1.1	1.34	1.58	3.52
	4 std Gyr	0.74	0.42	0	0.55	0.69	0.83	1.75
7	1 mean Acc	5.11	4.25	0.12	3.45	4.11	4.69	14.46
	2 std Acc	1.97	1.23	0.06	1.43	1.81	2.34	4.37
	3 mean Gyr	1.28	0.72	0.02	0.98	1.3	1.68	2.76
	4 std Gyr	0.67	0.34	0.02	0.54	0.65	0.86	1.39
8	1 mean Acc	4.99	4.46	0.07	3.17	3.63	4.25	15.87
	2 std Acc	1.56	1.03	0.03	0.95	1.27	1.53	4.22
	3 mean Gyr	1.60	0.92	0.01	1.28	1.53	1.86	3.64
	4 std Gyr	0.89	0.41	0.01	0.66	0.85	1.02	2
9	1 mean Acc	6.58	4.95	0.11	4.18	5.96	7.19	18.33
	2 std Acc	2.27	1.27	0.02	1.72	2.23	3.18	5.22
	3 mean Gyr	2.63	1.48	0	2.09	2.91	3.65	6.24
	4 std Gyr	1.32	0.74	0	1.1	1.36	1.66	3.91
10	1 mean Acc	5.31	4.99	0.12	2.74	3.77	4.96	18.23
	2 std Acc	1.99	1.35	0.03	1.15	1.53	2.71	6.62
	3 mean Gyr	1.65	0.97	0.01	1.16	1.67	2.2	3.64
	4 std Gyr	0.82	0.43	0.01	0.6	0.79	1.13	1.82
11	1 mean Acc	4.17	3.67	0.09	2.55	3.11	3.81	13.23
	2 std Acc	1.6	1.06	0.02	1.2	1.39	1.75	4.54
	3 mean Gyr	1.11	0.68	0	0.86	1.08	1.27	2.81
	4 std Gyr	0.56	0.29	0	0.44	0.58	0.72	1.45

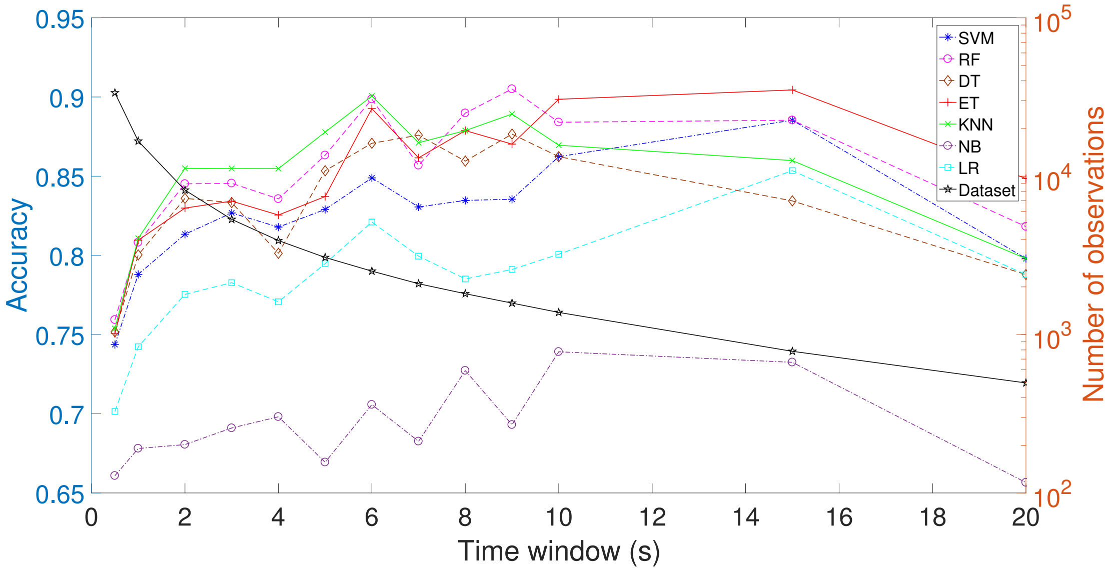

Figure 6.4: Predictive performances of various algorithms for different time window sizes.

time window is tiny, it usually can not contain useful information about the executed activity.

Table 6.6 summarizes the characteristics of all the features of the computed datasets for the time window size tests.

Table 6.6: Datasets summary for the time window size tests.

Tw	Feature no & name	Mean	Std	Min	25%	50%	75%	Max
0.5	1 mean Acc	5.43	4.67	0.04	2.5	4.14	6.51	23.88
	2 std Acc	1.71	1.26	0.01	0.86	1.45	2.38	10.29
	3 mean Gyr	1.70	1.17	0	0.85	1.6	2.52	10.44
	4 std Gyr	0.73	0.54	0	0.33	0.66	1.08	8.52
1	1 mean Acc	5.43	4.65	0.04	2.55	4.13	6.4	21.38
	2 std Acc	1.77	1.24	0.01	0.98	1.54	2.38	9.02
	3 mean Gyr	1.7	1.14	0	0.9	1.6	2.51	8.78
	4 std Gyr	0.78	0.53	0	0.4	0.73	1.12	7.13

Continued on the next page

Table 6.6 – continued from the previous page.

Tw	Feature no & name	Mean	Std	Min	25%	50%	75%	Max
2	1 mean Acc	5.42	4.63	0.05	2.62	4.1	6.35	20.17
	2 std Acc	1.81	1.21	0.01	1.05	1.59	2.38	8.55
	3 mean Gyr	1.7	1.12	0	0.93	1.58	2.51	7.84
	4 std Gyr	0.81	0.51	0	0.46	0.76	1.14	5.52
3	1 mean Acc	5.41	4.62	0.05	2.67	4.08	6.28	20.16
	2 std Acc	1.83	1.2	0.02	1.08	1.62	2.4	7.61
	3 mean Gyr	1.7	1.11	0	0.96	1.57	2.49	6.95
	4 std Gyr	0.83	0.5	0	0.5	0.78	1.15	4.62
4	1 mean Acc	5.4	4.62	0.06	2.67	4.09	6.26	19.82
	2 std Acc	1.84	1.19	0.02	1.09	1.65	2.42	7
	3 mean Gyr	1.7	1.1	0	0.98	1.56	2.5	6.53
	4 std Gyr	0.84	0.5	0	0.53	0.78	1.16	4.16
5	1 mean Acc	5.39	4.59	0.06	2.7	4.09	6.19	18.93
	2 std Acc	1.85	1.18	0.02	1.1	1.66	2.42	6.62
	3 mean Gyr	1.7	1.1	0	1	1.55	2.49	6.22
	4 std Gyr	0.85	0.49	0	0.54	0.79	1.17	3.88
6	1 mean Acc	5.4	4.62	0.06	2.72	4.06	6.22	19.13
	2 std Acc	1.86	1.18	0.02	1.1	1.67	2.44	6.33
	3 mean Gyr	1.7	1.09	0	1	1.54	2.49	5.99
	4 std Gyr	0.85	0.49	0	0.55	0.79	1.16	3.65
7	1 mean Acc	5.39	4.59	0.07	2.72	4.07	6.19	18.87
	2 std Acc	1.87	1.17	0.02	1.1	1.67	2.42	6.18
	3 mean Gyr	1.7	1.09	0	1.01	1.54	2.47	5.75
	4 std Gyr	0.8	0.48	0	0.56	0.79	1.17	3.46
8	1 mean Acc	5.4	4.62	0.07	2.69	4.07	6.19	19.03
	2 std Acc	1.87	1.18	0.02	1.11	1.68	2.41	5.99
	3 mean Gyr	1.7	1.09	0	1	1.54	2.48	5.71
	4 std Gyr	0.86	0.48	0	0.56	0.79	1.17	3.3

Continued on the next page

Table 6.6 – continued from the previous page.

Tw	Feature no & name	Mean	Std	Min	25%	50%	75%	Max
9	1 mean Acc	5.40	4.63	0.08	2.68	4.07	6.20	18.90
	2 std Acc	1.87	1.18	0.02	1.11	1.68	2.43	5.84
	3 mean Gyr	1.7	1.09	0	0.99	1.53	2.48	5.57
	4 std Gyr	0.86	0.48	0	0.56	0.8	1.17	3.14
10	1 mean Acc	5.4	4.61	0.07	2.75	4.08	6.18	18.37
	2 std Acc	1.88	1.17	0.02	1.1	1.67	2.45	5.71
	3 mean Gyr	1.7	1.09	0	0.99	1.54	2.48	5.62
	4 std Gyr	0.87	0.47	0	0.57	0.79	1.17	3.22
15	1 mean Acc	5.27	4.51	0.08	2.72	4.04	6.02	17.4
	2 std Acc	1.87	1.15	0.02	1.11	1.67	2.37	5.37
	3 mean Gyr	1.68	1.07	0	1.01	1.52	2.43	5.34
	4 std Gyr	0.87	0.46	0.01	0.57	0.79	1.15	2.82
20	1 mean Acc	5.30	4.49	0.09	2.78	4.03	5.95	17.22
	2 std Acc	1.9	1.13	0.04	1.15	1.7	2.37	5.21
	3 mean Gyr	1.69	1.05	0.01	1.02	1.51	2.39	5.26
	4 std Gyr	0.88	0.44	0.01	0.59	0.81	1.14	2.6

6.5.3 Sensors and Features

In this part, we have experimented with multiple machine learning algorithms, and with various combination of the input features. For each classifier, we tried all possible combinations of the groups of features of Table 6.1 from both available sensors. The results are presented in Table 6.7. Figure 6.5 presents the box plot for all trained classifiers for the user-independent tests using all available features from both sensors.

Table 6.8 summarizes the characteristics of all the features that were computed over a time window of 5s with a step size of 1s. By using this segmentation, 7506 instances were computed.

Table 6.7: Predictive performances of various algorithms for various input feature combinations.

Feature groups	**Sensors**	**Algorithms**						
		LR	SVM	RF	DT	NB	ET	KNN
1	A	79.7%	80.2%	83.3%	83.7%	71.2%	81.9%	83.4%
	G	54%	53.9%	61.9%	61.9%	51.5%	60%	60.6%
	A+G	81.2%	84.3%	87.6%	86.8%	68.8%	85.8%	90.7%
2	A	84%	84.9%	85.8%	85.6%	73%	83.8%	87%
	G	62.8%	65.2%	73.6%	72.2%	53.9%	70.1%	75%
	A+G	88.1%	89.3%	89.5%	90.3%	70.4%	88%	93.9%
3	A	77%	80.6%	89.1%	91.4%	76.7%	87.4%	88.8%
	G	75.5%	82.9%	91.1%	89.2%	72%	88.1%	90.3%
	A+G	89%	92.8%	93.3%	95.1%	80%	94.1%	93.7%
1+2	A	83.7%	84.6%	86.3%	86.2%	74.4%	84.4%	87.7%
	G	63.1%	66%	75%	71.3%	53.6%	71.7%	77%
	A+G	87.8%	89.8%	90%	90.7%	70.6%	89.8%	94.3%
1+3	A	86.4%	90.6%	93.7%	94.7%	79.4%	91.5%	92.5%
	G	78.4%	84.8%	91.9%	91.1%	70.4%	89%	91.5%
	A+G	92.6%	96.4%	95.3%	94.9%	80.6%	94.8%	95.5%
2+3	A	90%	92.8%	92.4%	94%	82%	90.5%	93.9%
	G	82%	88.2%	92.8%	91%	70.8%	90.4%	93.6%
	A+G	94.9%	97.3%	94.7%	94.5%	81.7%	94.7%	96.3%
1+2+3	A	90%	92.8%	93.5%	93.5%	81.3%	90.9%	94.1%
	G	82.6%	88.6%	93.1%	90.9%	70.3%	90%	94%
	A+G	95.2%	97.3%	95.5%	95%	81.3%	95.3%	97.1%

Table 6.8: Dataset summary for features computed over a time window of 5s with a step size of 1s.

Feature no & name	Mean	Std	Min	25%	50%	75%	Max
1 mean Acc	5.39	4.60	0.06	2.69	4.06	6.21	19.43
2 std Acc	1.85	1.78	0.02	1.1	1.65	2.41	6.62
3 median Acc	5.32	4.64	0.05	2.58	4.02	6.23	19.39
4 skewness Acc	0.58	1.11	-0.89	-0.11	0.21	0.85	8.84
5 kurtosis Acc	4.62	6.89	1.75	2.38	2.73	3.79	104.74
6 p25 Acc	4.04	3.8	0.04	1.73	2.9	4.61	16.12
7 p75 Acc	6.66	5.49	0.07	3.48	5.15	7.85	23.19
8 sumsq25 Acc	1539.72	2704.5	0.05	118.84	343.89	927.83	14692
9 sumsq75 Acc	9217.96	14158.54	0.58	1166.2	2863.85	6911.68	71353
10 mean Gyr	1.7	1.09	0	1.01	1.54	2.49	6.24
11 std Gyr	0.84	0.49	0	0.54	0.79	1.16	3.91
12 median Gyr	1.62	1.1	0	0.89	1.48	2.46	5.4
13 skewness Gyr	0.8	0.92	-0.59	0.21	0.54	1.1	13.12
14 kurtosis Gyr	4.22	4.6	1.54	2.28	2.85	4.29	199.51
15 p25 Gyr	1.05	0.75	0	0.55	0.96	1.59	4.24
16 p75 Gyr	2.24	1.44	0	1.34	2.08	3.3	6.93
17 sumsq25 Gyr	66.13	88.49	0	11.63	34.37	94.45	882.68
18 sumsq75 Gyr	670.03	733.4	0	145.59	383.66	1050	5425.7
19 maxFreqX Acc	2	1.19	0.12	1.41	1.88	2.11	15.94
20 sum5HzX Acc	33.43	26.04	0.16	18.8	26.08	43.27	115.61
21 numPeaksX Acc	3.12	1.81	0	2	4	4	7
22 maxFreqY Acc	1.8	1.43	0.12	0.94	1.05	2.7	13.01
23 sum5HzY Acc	30.66	26.71	0.21	13.58	21.2	35.14	117.68
24 numPeaksY Acc	2.64	1.76	0	2	2	4	7
25 maxFreqZ Acc	1.76	1.53	0.12	0.94	1.05	2.58	11.84
26 sum5HzZ Acc	19.43	15.44	0.19	9.77	14.99	26.08	93.17
27 numPeaksZ Acc	2.09	1.54	0	1	2	3	7
28 maxFreqX Gyr	1.49	1.24	0.12	0.47	1.29	1.99	7.97
29 sum5HzX Gyr	10.96	7.74	0.03	6.7	9.88	13.67	77.96
30 numPeaksX Gyr	1.46	1.35	0	0	1	2	6
31 maxFreqY Gyr	1.08	0.65	0.12	0.82	0.94	1.29	7.5
32 sum5HzY Gyr	6.98	5.53	0.01	3	5.41	9.48	33.4
33 numPeaksY Gyr	0.75	0.93	0	0	0	1	5
34 maxFreqZ Gyr	1.13	0.6	0.12	0.82	0.94	1.29	7.73
35 sum5HzZ Gyr	10.45	6.61	0	5.65	10.09	14.86	36.98
36 numPeaksZ Gyr	1.33	0.98	0	0	2	2	6

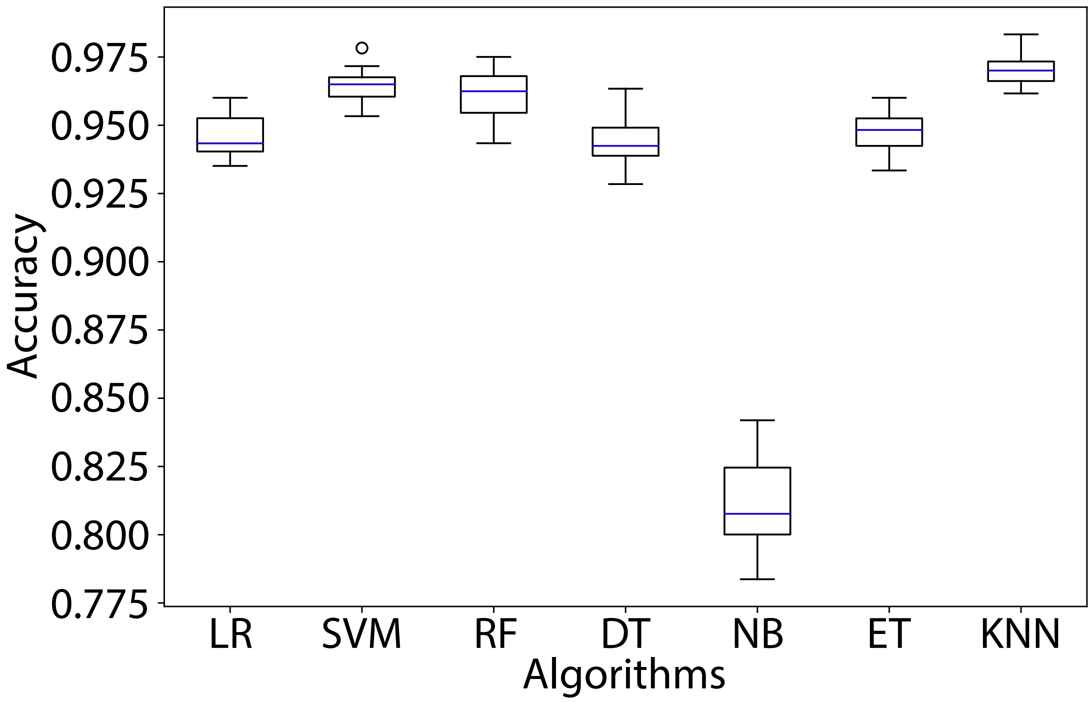

Figure 6.5: Box plot of multiple person-independent classifiers using all available features from both sensors.

The designer of an AR system usually has to decide on a single classifier. This choice is usually not only based on the performance of the model. Some classifiers tend to be more computationally complex, and sometimes their superiority in performance may not justify the difference in their requirements concerning computing power. In our tests, for the last test case, the SVM and the kNN classifiers achieved predictive performances of 97.3% and 97.1%. However, the former is more computationally demanding than the latter and may not be suitable for a resource-restricted environment.

Moreover, the more features we are feeding a classifier, the more computationally complex it becomes. Therefore, it is a good practice to apply feature selection techniques to reduce the feature set to the most significant features. In our case, all automatic feature selection techniques ranked highest the features extracted from accelerometer data. This result was expected, given that according to the results of

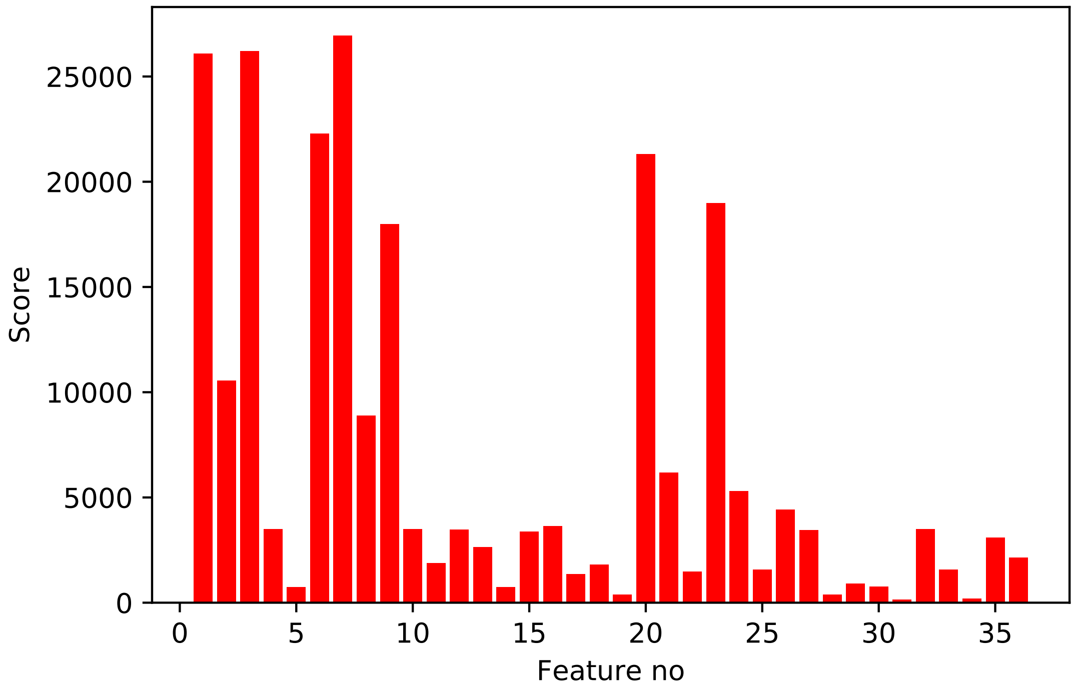

Figure 6.6: Feature selection scores.

Table 6.7, accelerometer features performed better than the equivalent gyroscope ones in most cases when used independently.

The fact that the features extracted from accelerometer data contained more informative information than the ones from the gyroscope was also proved with a univariate feature selection test. To select the features that have the strongest relationship with the output value, the ANOVA F-value was computed for the full feature set, and the scores are presented in Figure 6.6. The 12 best features (7, 3, 1, 6, 20, 23, 9, 2, 8, 21, 24, 26) were all computed with accelerometer data.

6.6 Conclusion

This chapter demonstrates different considerations for the design of an inertial sensor-based activity recognition system. After discussing many parameters of such a system, we presented an experiment of a smartwatch-based AR system. Depending on the selection of the different parameters of the AR system, the predictive performance of it changes. Although all design stages depend on each other and require joint optimization, we have evaluated stages separately and discussed their overall impact on the predictive performance of the AR system. In general, person-dependent models can be trained to achieve higher accuracy than person-dependent ones, but the latter can be optimized to generalize well by providing them with more training data from multiple users. Accelerometers provide in general more information than gyroscopes when those sensors are used alone, but fusing data from multiple sensors can further improve the accuracy. After several tests with multiple features and different machine learning algorithms, a predictive performance of up to 97.3% for a person-independent model was achieved.

As with any real-world application, there were some limitations regarding the collected dataset. For the study, participants were asked to perform the required activities in their work environment where conditions were not ideal. For example, when going up or down the staircase, the participants had to make a sharp turn to continue to the next chunk of stairs and some time up to 1s was spent walking during this transition. Moreover, the participants were all men, and of a small age range, so the data are skewed regarding these aspects. Last but not least, the collected dataset was limited in terms of the number of different activities of interest and the length of the recordings, but the considerations that were presented are relevant for more extensive experiments.

7 Gait Recognition Using a Smartwatch Assisting Postoperative Physiotherapy

7.1 Chapter Abstract

Postoperative rehabilitation is a vital program that re-establishes joint motion and strengthens the muscles around the joint after orthopedic surgery. This kind of rehabilitation is led by physiotherapists who assess each situation and prescribe appropriate exercises. Modern smart devices have affected every aspect of human life. Newly developed technologies have revolutionized the way various industries operate, including the healthcare one. Extensive research has been carried out on how smartphone inertial sensors can be used for activity recognition. However, there are very few studies on systems that monitor patients and detect different gait patterns in order to assist the work of physiotherapists during the said rehabilitation phase, even outside the time-limited physiotherapy sessions, and therefore literature on this topic is still in its infancy. In this chapter, we are presenting a gait recognition system that was developed to detect different gait patterns, including walking with crutches with various levels of weight-bearing, walking with different frames, limping,

This work has been published as: *'Gait Recognition with Smart Devices Assisting Postoperative Rehabilitation in a Clinical Setting'*, Kyritsis, A.I., Willems, G., Deriaz, M. and Konstantas, D., In IEEE International Conference on Artificial Intelligence for Industries (ai4i 2018), Laguna Hills, CA, USA, September 2018. This publication was granted the IEEE Computer Society Best Paper Award.

An extended and revised version was published as: *'Gait Pattern Recognition Using a Smartwatch Assisting Postoperative Physiotherapy'*, Kyritsis, A.I., Willems, G., Deriaz, M. and Konstantas, D., In International Journal of Semantic Computing, Vol. 13, No. 2, p. 245–257, 2019.

and walking normally. The proposed system was trained, tested, and validated with data of people who have undergone lower body orthopedic surgery, recorded by Hirslanden Clinique La Colline, an orthopedic clinic in Geneva, Switzerland. Professional physiotherapists labeled nine different gait classes. After extracting both time and frequency domain features from the time series data, several machine learning models were tested including a fully connected neural network. Raw time series data were also fed into a convolutional neural network.

7.2 Introduction and Related Work

The role of physiotherapy following orthopedic surgery is to assist the patient return to normal activities of daily living. Doctors and physiotherapists help the patient achieve this by prescribing suitable exercises that will establish the rehabilitation goals. There is a significant body of evidence coming from systematic reviews and controlled trials that dictate the best practices in physiotherapy [159]. Proper evaluation guarantees the effectiveness of physiotherapy [160] for a wide variety of medical conditions, including recovering after a lower-body orthopedic operation.

Gait refers to a person's manner of walking and is influenced by age, personality, mood, and sociocultural factors [161]. Several reasons, including a lower body operation, may lead to either temporary or permanent gait disorders. Any such disorder is typically thoroughly investigated by the physiotherapist who then suggests a specific treatment to the patient. There are various tools at the disposal of the physiotherapists, and many robotic solutions are being created in order to help people walk or to act as an aid during a physiotherapy session [162]. These robot-assisted gait solutions may be used as an excellent companion to conventional therapy and improve the independence and the gait capacity of the patient [163].

Activity recognition (AR) has emerged as a key research domain in computer science. The approaches for AR can be roughly divided into two categories: the camera-based ones [125], where gestures and activities are inferred from still images or videos using computer vision techniques, and the inertial sensor-based ones, where one or more body-worn sensors are used [147]. Any AR system includes many variables such as the definition of the classes of interest, the experiment design, the sensors, the data handling procedure, and the performance evaluation. These variable components can be implemented in a variety of ways [13] having an impact on the overall performance of the system.

The increased availability of inertial sensors due to the omnipresence of smartphones, and particularly smartwatches, has enabled AR to become an essential context-awareness tool for mobile and ubiquitous computing. Sensors in modern consumer electronics provide reasonably accurate recordings when compared to research monitors [164]. This is why these devices prove to have clinical utility, although they continue to be underutilized in the healthcare industry [165].

Besides recognizing daily activities, inertial sensors have been used in gait pattern analysis. In most studies, accelerometers are attached to the legs or feet, but gait patterns can also be extracted from data recorded by sensors attached to the upper body [166]. Common smartphone accelerometers have been used to detect different gait events [167]. In a similar manner, smartwatches that contain inertial sensors can be used for gait recognition. Unlike smartphones, smartwatches tend to be worn in the same location and the same orientation. They can be even used for gait-based biometrics based on the accelerometer and the gyroscope data [168].

Various recovery programs have been developed to improve the recovery time after surgery [169]. Wireless monitoring of mobility after a major operation has the potential of improving services provided by healthcare professionals [170]. With the proposed system, we incorporate smartwatches into the routine care of patients who have undergone a lower body operation in order to monitor their gait patterns. Doing so will enhance the patient-physiotherapist relationship, respect the patients' autonomy regarding their healthcare, and provide a remote monitoring solution to the physiotherapist in charge. The research question of this chapter is, "How can an

Table 7.1: Classification of gait patterns for recognition.

Category	Class
No aid	Limping Walking
Crutches	Unladen Rolled out Laden 10kg Laden 20kg According to pain
Frame	Without wheels With wheels

activity recognition system assist physiotherapists to monitor patients during the rehabilitation phase of someone that has undergone a lower body operation?"

The rest of the chapter is organized as follows. In Section 7.3 we discuss the system that we have developed. We present the data acquisition tools and the data preprocessing procedure. In Section 7.4 we present the experiment that we have conducted and we evaluate the performance of the overall system by training machine learning models and neural networks. Finally, we conclude our work in Section 7.5.

7.3 System Overview

7.3.1 Gait Classification

The physiotherapists of Hirslanden Clinique La Colline, an orthopedic clinic in Geneva, Switzerland, compiled a list of the gait patterns of interest to our system. The patterns include walking with crutches with various levels of weight-bearing, walking with different frames, limping, and walking normally. Table 7.1 includes the list of all the 9 gait patterns that our system should detect.

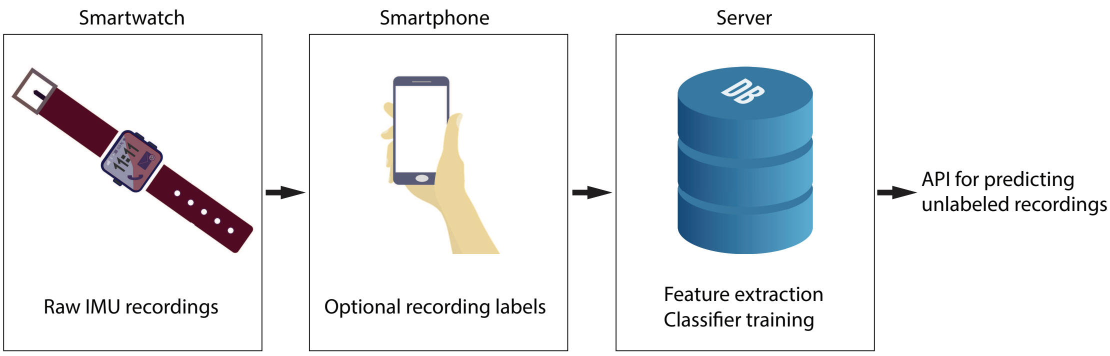

Figure 7.1: Summary of the implementation of the gait recognition system.

7.3.2 Workflow

The developed system comprises three components, the smartwatch, the smartphone, and the web server. Figure 7.1 presents the flow of data in the proposed system. The system is meant to be used during the rehabilitation phase, that is the time that the patient is undergoing physiotherapy after a lower-body surgery. During physiotherapy sessions in the clinic, any patient is walking while wearing a smartwatch that tracks wrist movements. At the same time, the physiotherapist is labeling on a smartphone any physiotherapy session with the observed gait pattern of the patient. All these data from multiple patients and physiotherapy sessions are uploaded to the web server, where a user-independent machine learning model is trained.

During everyday life, through the rehabilitation phase, the patient is wearing a given smartwatch. Throughout the day, the smartwatch is passively recording gait sessions of unknown gait patterns when the patient is moving. These recordings are uploaded from the smartwatch to the web server. Using the trained machine learning model, those new recordings are classified into the predefined gait patterns. Using the web server, the physiotherapists can monitor how each patient's gait pattern is evolving, even between physiotherapy sessions.

7.3.3 System Implementation

Wrist movements of the patients are recorded using the three-axis accelerometer and the three-axis gyroscope of an Android smartwatch running Wear OS. The accelerometer sensor provides a three-dimensional vector representing acceleration along each device axis, excluding gravity. The gyroscope sensor measures the angular velocity of each axis of the device. Recordings can be made either on-demand during a physiotherapy session when the physiotherapist can provide the ground truth with the observed gait pattern, or by transparently monitoring the movement of the user throughout the day and saving only sessions where prolonged movement or steps are identified.

At the end of every on-demand recording, sensor data are sent from the smartwatch to the connected Android smartphone. The smartphone is used by physiotherapists to label each recording during a physiotherapy session with the identified gait pattern. The recordings that are produced during the monitoring phase of the system during the whole rehabilitation program, naturally have no ground truth label and are directly sent from the patient's smartwatch to the web server.

Every recording is saved to the web server. On every upload of a new recording, the web server is extracting the features that will be later used for machine learning. Training of the selected user-independent machine learning classifier is run periodically when enough new labeled recordings from multiple users have been obtained. On the other hand, the server exposes an API with which the unknown gait patterns of the non labeled recordings can be predicted. The physiotherapist can query the server in order to monitor what is the dominant detected gait pattern of a specific time and how it evolves during the rehabilitation program.

7.3.4 Data Preprocessing

The accelerometer and the gyroscope sensors of the smartwatch that we have used did not provide a constant sampling rate throughout the recordings. This is why the raw sensor data were resampled with a constant sampling frequency of 60 Hz, as we also did in the previous chapter. This frequency was selected for this study as it is higher than the 20 Hz commonly required to assess daily living [153] and also lower than what typical off-the-shelf inertial measurement unit components can achieve. Features forming the feature vector used for machine learning were derived from these time series data, and these were the raw data fed to the convolutional neural network.

7.4 Experiment and Evaluation

Physiotherapists of the Hirslanden Clinique La Colline recorded wrist movements of patients walking soon after they have undergone a lower body orthopedic surgery, either in the hip, the knee, the ankle or the foot. During all recordings, the physiotherapist was in close proximity to the patient, in order to guarantee the correct ground truth annotation and the cleanliness of the data. In total, 48 recordings from 33 different patients were made over a period of 4 months. The recordings amount to a total time of 155 minutes of labeled gait patterns.

7.4.1 Feature Engineering

The same features that were described in the previous chapter were also computed for the needs of this work. Both time and frequency domain features were computed for both sensors over a selected time window. The time domain features include the mean, the standard deviation, the median, the skewness, the kurtosis, the 25th and the 75th percentile, and the squared sum of the components under the 25th and the

Table 7.2: Extracted features per sensor used in machine learning.

Domain	Features	No of features
Time (resultant vector)	Mean Standard deviation Median Skewness Kurtosis 25th percentile 75th percentile Sq. sum of < 25th perc. Sq. sum of < 75th perc.	9
Frequency (per axis)	Maximum frequency Sum of 5 Hz Number of peaks	9

75th percentile. Those were derived from the resultant vector computed by the three, x, y, and z components that each sensor provides.

For the frequency domain features, a Fast Fourier Transform (FFT) was performed after normalization on the windows, and the features were computed per axis. Those features include the maximum frequency, the sum of heights of frequency components below 5 Hz, and the number of peaks in the spectrum below 5 Hz, as it was noticed that most of the signal strength lied between 0-5 Hz. The selection of the features was based on a feature importance analysis presented in a work of ours [13]. All the features extracted for this study are summarized in Table 7.2.

7.4.2 Machine Learning

The classifiers that we have evaluated are Light Gradient Boosting Machine (LGBM) [171], Logistic Regression (LR) [154], Support Vector Machines (SVM) [95], Random Forest (RF) [92], Decision Tree (DT) [155], Extra Trees (ET) [157], and k-Nearest Neighbours (kNN) [158]. Each recording is segmented into multiple time windows. The features were computed over a time window of five seconds with a step size of one second, so there was a four-second overlap between consecutive windows. This value for the time window was identified in a work of ours [13] as a good candidate since it is large enough to contain useful information regarding the activity and small enough to increase the number of the produced segments during segmentation. The segmentation of any given recording is depicted in Figure 7.2. The constructed dataset contained 9089 observations in total. Table 7.3 summarizes the characteristics of all the features of the computed dataset. Accelerometers measure linear acceleration in meters per second squared (m/s^2), and gyroscopes measure rotational motion in radian per second (rad/s). The mean, min, and max columns are self-explanatory. Note that there were no null values. The std column shows the standard deviation that measures how dispersed the values are. The 25%, 50%, and 75% columns show the corresponding percentiles. A percentile indicates the value below which a certain percentage of observations among the whole set of observations fall. The value of the 50%, therefore, corresponds to the median. Indicatively, the mean and the standard deviation of the time domain features of the observations are presented in Table 7.4 on a per-class manner.

The acquired dataset was imbalanced. The reasons were either lack of availability of patients with a gait pattern belonging to one of the minority classes, or no consent from the patient. Figure 7.3 presents the observation count of the available dataset. This is problematic since training on an imbalanced dataset will lead to a biased model. To cope with the problem of the imbalanced dataset and to optimize the performance of the classification algorithms, the random minority over-sampling with replacement method was used [172].

We have used Matlab for feature extraction and Python and the Scikit-learn module [152] for machine learning. To evaluate the performance of our system, we split the available dataset into a training set (80%) and a test set (20%) in a stratified fashion. The minority classes of the training set were randomly over-sampled with replacement. The 10-fold cross-validation scheme was used on the training set to train the model, the performance of which was evaluated on the test set. Figure 7.4 presents the box plot for all trained classifiers. Different classifiers naturally perform differently. This is due to the nature of the problem, the characteristics of the dataset, and the capacity

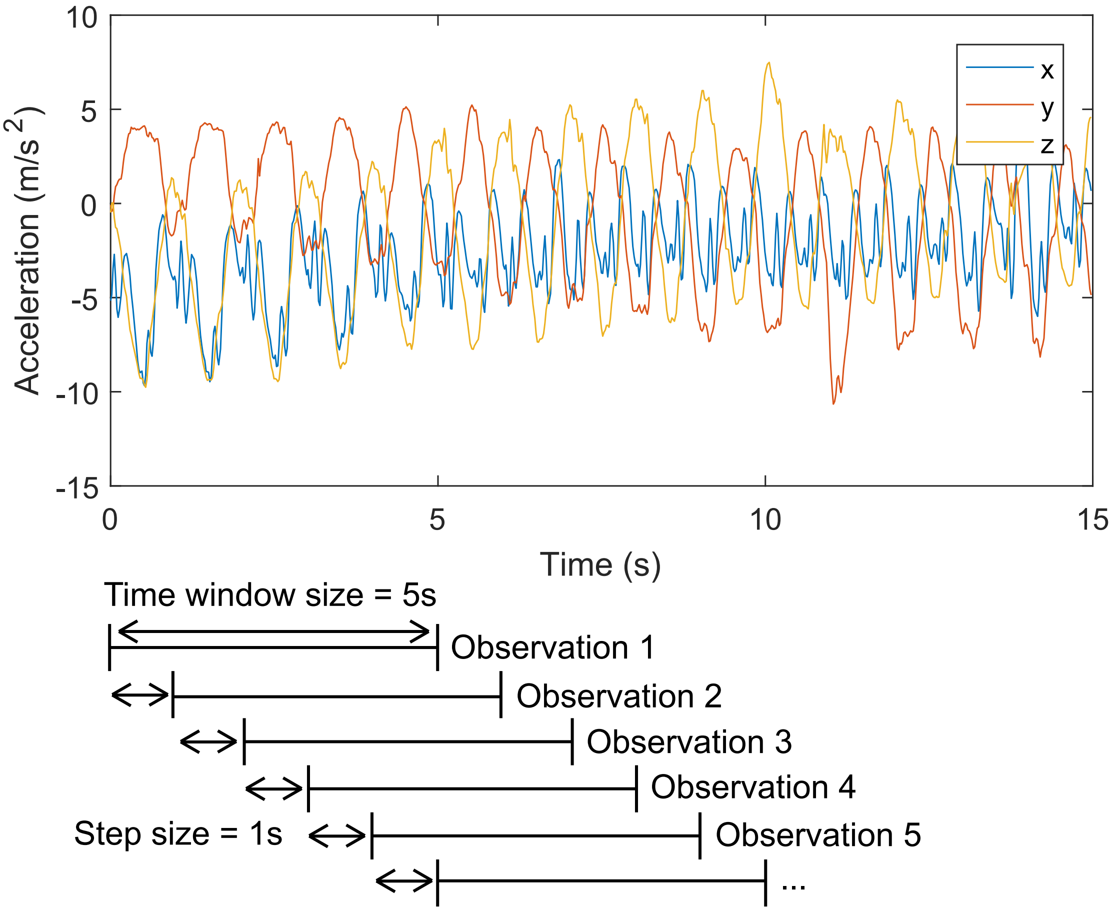

Figure 7.2: An example of the segmentation of a data recording.

Table 7.3: Dataset summary for all the computed features.

Feature no & name	Mean	Std	Min	25%	50%	75%	Max
1 mean Acc	1.75	1.15	0.09	1.14	1.58	2.05	15.38
2 std Acc	1.43	0.72	0.05	0.94	1.38	1.82	5.32
3 median Acc	1.38	1.13	0.08	0.83	1.2	1.58	15.93
4 skewness Acc	3.08	1.71	-0.71	1.87	2.97	4.16	10.42
5 kurtosis Acc	20.44	17.27	1.64	7.55	15.77	28.16	133.84
6 p25 Acc	0.91	0.89	0.05	0.52	0.76	1.02	12.89
7 p75 Acc	2.12	1.41	0.11	1.37	1.9	2.46	17.97
8 sumsq25 Acc	74.47	387.23	0.12	11.48	22.17	40.96	9239.8
9 sumsq75 Acc	596.13	2056.3	1.08	138.58	272.82	466.89	46655
10 mean Gyr	0.69	0.49	0.01	0.4	0.56	0.84	4.33
11 std Gyr	0.5	0.35	0.01	0.31	0.42	0.56	4.6
12 median Gyr	0.54	0.47	0.01	0.27	0.42	0.64	4.54
13 skewness Gyr	1.48	0.79	-0.6	1.02	1.38	1.82	6.77
14 kurtosis Gyr	5.95	4.86	1.58	3.31	4.52	6.73	63.96
15 p25 Gyr	0.34	0.29	0	0.17	0.27	0.42	3.16
16 p75 Gyr	0.92	0.67	0.01	0.51	0.72	1.12	6.32
17 sumsq25 Gyr	7.76	18.76	0	1.07	2.84	8.16	342.29
18 sumsq75 Gyr	95.56	220.5	0.01	16.16	35.68	85.45	3350.9
19 maxFreqX Acc	2	1.64	0.12	1.17	1.64	2.11	14.65
20 sum5HzX Acc	16.41	8.08	0.63	10.82	15.5	21.43	56.51
21 numPeaksX Acc	1.44	1.49	0	0	1	2	6
22 maxFreqY Acc	2.64	2.41	0.12	0.94	1.52	3.87	15.35
23 sum5HzY Acc	16.99	8.1	0.68	11.86	16.4	20.69	73.06
24 numPeaksY Acc	1.24	1.4	0	0	1	2	6
25 maxFreqZ Acc	2.36	1.81	0.12	1.05	1.88	3.16	15.12
26 sum5HzZ Acc	13.93	7.12	0.7	8.87	13.13	17.73	59.61
27 numPeaksZ Acc	0.97	1.27	0	0	0	2	6
28 maxFreqX Gyr	1.3	1.23	0.12	0.47	0.82	1.64	7.97
29 sum5HzX Gyr	5.93	4.4	0.12	3.63	4.78	6.73	69.57
30 numPeaksX Gyr	0.23	0.68	0	0	0	0	5
31 maxFreqY Gyr	0.85	0.81	0.12	0.47	0.59	0.94	9.96
32 sum5HzY Gyr	4.76	3.33	0.01	2.78	3.84	5.72	48.55
33 numPeaksY Gyr	0.19	0.53	0	0	0	0	7
34 maxFreqZ Gyr	0.69	0.7	0.12	0.35	0.59	0.82	11.02
35 sum5HzZ Gyr	5.27	3.61	0.08	3.09	4.36	6.35	40
36 numPeaksZ Gyr	0.35	0.61	0	0	0	1	6

Table 7.4: Dataset summary per class.

Class	Feature no & name	Mean	Std	Min	25%	50%	75%	Max
Limping	1 mean Acc	2.33	1.26	0.24	1.47	1.93	2.82	6.97
	2 std Acc	0.99	0.52	0.09	0.66	0.77	1.34	2.72
	10 mean Gyr	1.21	0.55	0.06	0.84	1.13	1.42	2.96
	11 std Gyr	0.67	0.35	0.06	0.43	0.62	0.87	1.96
Walking	1 mean Acc	3.72	3.65	0.17	0.94	2.07	6.35	15.38
	2 std Acc	1.28	0.96	0.11	0.44	1.15	1.82	4.48
	10 mean Gyr	1.42	1.15	0.04	0.32	1.20	2.54	4.33
	11 std Gyr	0.74	0.53	0.03	0.18	0.70	1.17	2.52
Crutches Unladen	1 mean Acc	2.53	0.57	0.65	2.32	2.53	2.87	3.78
	2 std Acc	2.04	0.51	0.43	1.81	2.04	2.29	3.25
	10 mean Gyr	0.87	0.24	0.13	0.76	0.88	1.01	1.59
	11 std Gyr	0.68	0.23	0.13	0.55	0.63	0.81	1.60
Crutches Rolled out	1 mean Acc	1.97	0.61	0.18	1.86	2.13	2.32	2.89
	2 std Acc	1.46	0.47	0.07	1.30	1.46	1.70	2.83
	10 mean Gyr	0.64	0.23	0.03	0.49	0.72	0.79	1.18
	11 std Gyr	0.43	0.16	0.01	0.36	0.44	0.49	0.99
Crutches Laden 10kg	1 mean Acc	1.79	0.68	0.32	1.33	1.65	2.35	3.49
	2 std Acc	1.57	0.57	0.07	1.21	1.61	1.96	3.05
	10 mean Gyr	0.56	0.29	0.03	0.34	0.45	0.84	1.16
	11 std Gyr	0.42	0.16	0.02	0.31	0.43	0.54	1.01
Crutches Laden 20kg	1 mean Acc	1.37	0.47	0.09	1.07	1.32	1.62	3.71
	2 std Acc	1.32	0.53	0.06	1.00	1.29	1.58	5.07
	10 mean Gyr	0.50	0.24	0.01	0.35	0.45	0.61	3.01
	11 std Gyr	0.39	0.25	0.01	0.29	0.36	0.44	4.37
Crutches According to pain	1 mean Acc	1.66	0.72	0.15	1.18	1.65	2.01	4.95
	2 std Acc	1.58	0.77	0.05	1.10	1.52	1.95	5.32
	10 mean Gyr	0.66	0.38	0.02	0.43	0.57	0.81	4.14
	11 std Gyr	0.52	0.38	0.01	0.33	0.43	0.57	4.60
Frame Without wheels	1 mean Acc	1.34	0.42	0.40	1.12	1.36	1.49	3.70
	2 std Acc	1.08	0.37	0.18	0.93	1.07	1.20	2.85
	10 mean Gyr	0.50	0.18	0.09	0.43	0.50	0.53	1.43
	11 std Gyr	0.34	0.25	0.06	0.26	0.30	0.33	1.53
Frame With wheels	1 mean Acc	0.57	0.09	0.32	0.52	0.56	0.61	0.84
	2 std Acc	0.36	0.11	0.17	0.29	0.33	0.40	0.70
	10 mean Gyr	0.24	0.08	0.13	0.19	0.21	0.24	0.61
	11 std Gyr	0.12	0.04	0.08	0.10	0.11	0.12	0.29

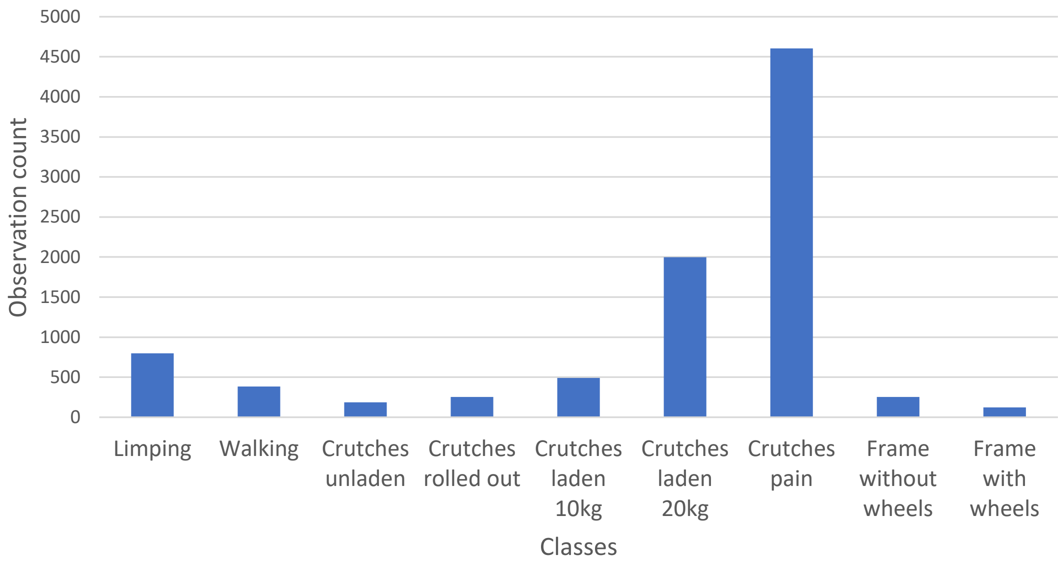

Figure 7.3: Observation count of the available dataset of all gait pattern classes.

Table 7.5: Confusion matrix of the LGBM classifier.

True class	**Predicted class**								
	L	W	CU	CRU	CL10kg	CL20kg	CP	FN	FW
Limping (L)	155	0	0	0	0	0	5	0	0
Walking (W)	3	66	0	0	0	0	8	0	0
Crutches unladen (CU)	0	0	36	0	0	1	0	0	0
Crutches rolled out (CRU)	0	0	0	43	1	2	5	0	0
Crutches laden 10kg (CL10kg)	0	0	0	0	81	3	14	0	0
Crutches laden 20kg (CL20kg)	0	0	0	0	0	377	23	0	0
Crutches pain (CP)	0	2	0	0	1	19	899	0	0
Frame without wheels (FN)	0	0	0	0	0	0	6	44	0
Frame with wheels (FW)	0	0	0	0	0	0	0	0	24

of each classifier in terms of the variety of functions it can fit. Table 7.5 presents the confusion matrix for the LGBM classifier, the best performing classifier and Table 7.6 presents the model's performance metrics.

We have achieved an accuracy of 94.9% with the LGBM classifier on the previously unseen test set. From the confusion matrix, it is worth noting that the misclassified observations belonging to one of the crutches classes were most of the times predicted to belong to another crutches class. Although misclassified per se, these kinds of

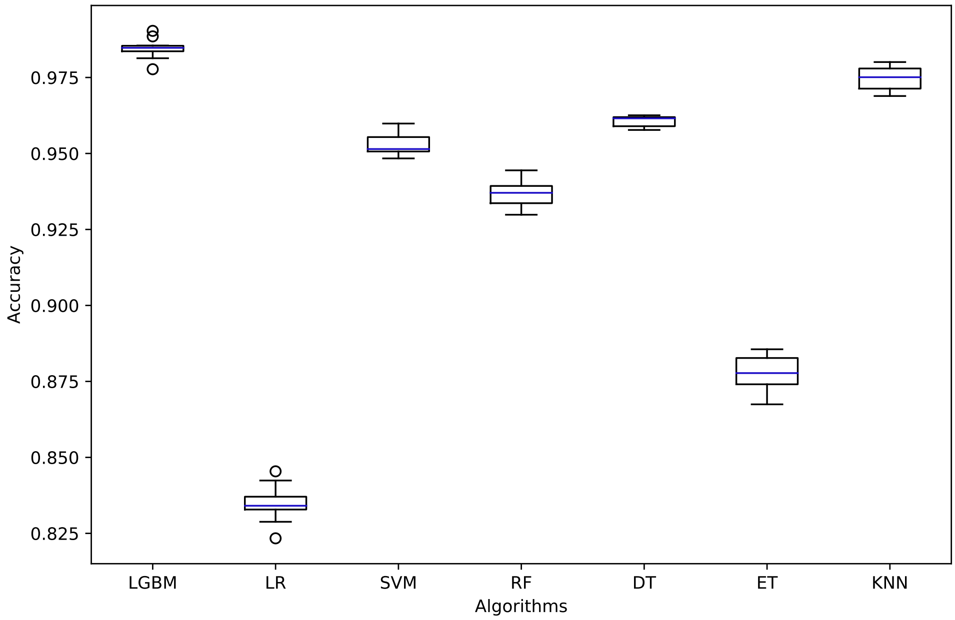

Figure 7.4: Box plot of multiple classifiers trained for gait recognition.

Table 7.6: Performance metrics of the LGBM classifier.

True class	**Precision**	**Recall**	**F1-score**
Limping (L)	0.981	0.969	0.975
Walking (W)	0.971	0.857	0.91
Crutches unladen (CU)	1	0.973	0.986
Crutches rolled out (CRU)	1	0.843	0.915
Crutches laden 10kg (CL10kg)	0.976	0.827	0.895
Crutches laden 20kg (CL20kg)	0.938	0.943	0.94
Crutches pain (CP)	0.937	0.976	0.956
Frame without wheels (FN)	1	0.88	0.936
Frame with wheels (FW)	1	1	1

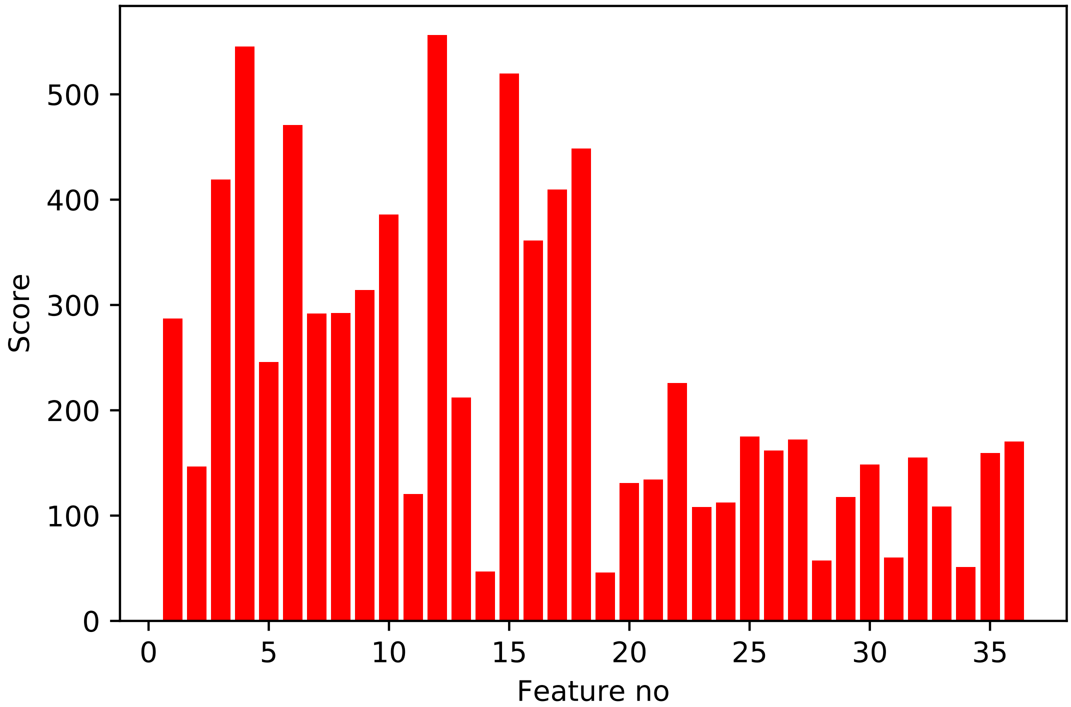

Figure 7.5: Feature selection scores.

observations may still provide physiotherapists useful information regarding the gait patterns of the patients.

To select the features that have the strongest relationship with the output value, a univariate feature selection test was conducted, and the ANOVA F-value was computed for the full feature set. The scores are presented in Figure 7.5. The 12 best features (12, 4, 15, 6, 18, 3, 17, 10, 16, 9, 8, 7) were not exclusively computed with accelerometer data this time, as in the previous chapter.

7.4.3 Fully Connected Neural Network

The same dataset that was constructed earlier was fed into a fully connected neural network. To evaluate the performance of our system, we split the available dataset again into a training set (80%) and a test set (20%) and the 10-fold cross-validation scheme was used on the training set to fine-tune the model. The network consists of

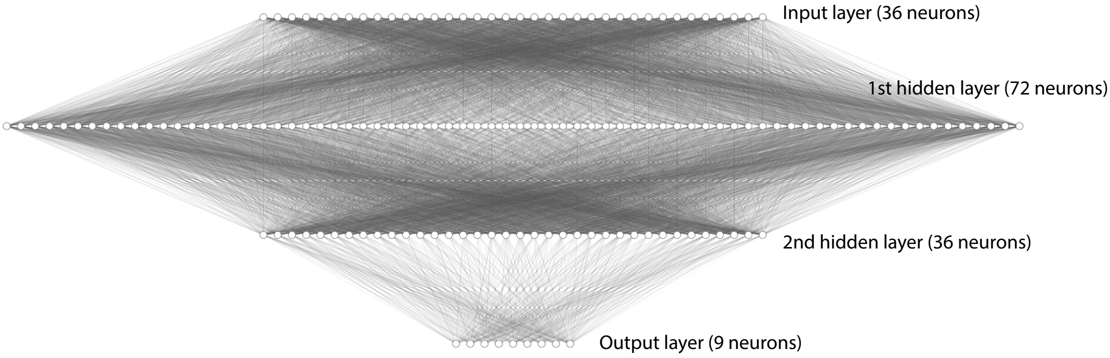

Figure 7.6: Visual representation of the fully connected neural network in use.

two hidden layers, the first with 72 neurons and the second with 36 ones, both applying the Rectified Linear Unit (ReLU) activation function. Figure 7.6 visually represents the network that we have used for this test. We have used the Adam optimization algorithm [173] with its default values, and since we have a problem of multiclass classification, we used the categorical cross-entropy as the loss function. Last but not least, the batch size was set to ten. We have used Python and the Keras API for the neural networks.

We achieved a performance of 90.9% on the test set. We notice that using the same dataset, the fully connected neural network that we trained performs worse than the LGBM classifier of the previous test.

7.4.4 Convolutional Neural Networks

For the models we have used so far, we manually engineered features from the time series data based on a fixed time window. However, there are deep learning methods such as recurrent neural networks and one-dimensional convolutional neural networks, that provide competent results with minimal or no feature engineering efforts.

We are exploring how one-dimensional convolutional neural networks perform for our problem of gait pattern recognition. The data that we will feed the neural networks with are the raw time-series ones, as these were produced after the universal resampling step with the sampling frequency of 60Hz. We have used the same time window of five seconds with a step size of one second. So for every time window, we end up with 300 values per axis per sensor. We were using two sensors, and each one of those had three axes, so we end up with six features in total. So each row of data contains 1800 elements. This is 50 times more than the features we manually engineered for the previous tests, so it is very likely that there are some redundant data.

The network that we have built consists of two one-dimensional convolutional neural network layers, both with a standard configuration of 64 feature maps and a kernel size of three. We have added a dropout layer [174] with a value of 0.2 for regularization and to prevent overfitting by slowing down the learning process. Then we have added a pooling layer that reduces the learned features to half of their size in order to avoid overfitting, and accelerate the training procedure. After the convolutional network, the learned features are flattened and passed through a fully connected layer of size 100 before the output layer. Last but not least, we are using again the Adam optimizer to optimize the network, and the categorical cross-entropy loss function for our multi-class classification problem.

Due to the stochastic nature of neural networks, we repeat the evaluation of the model 20 times and then summarize the performance of the model across those runs. We are training several one-dimensional convolutional neural networks, and we discuss how important parameter tuning is when creating such models.

7.4.4.1 Standardization

One possible transformation that we can apply on the available dataset is to standardize the input before training the model. By standardizing a variable, its distribution is shifted so that it has zero mean and a standard deviation of one. We

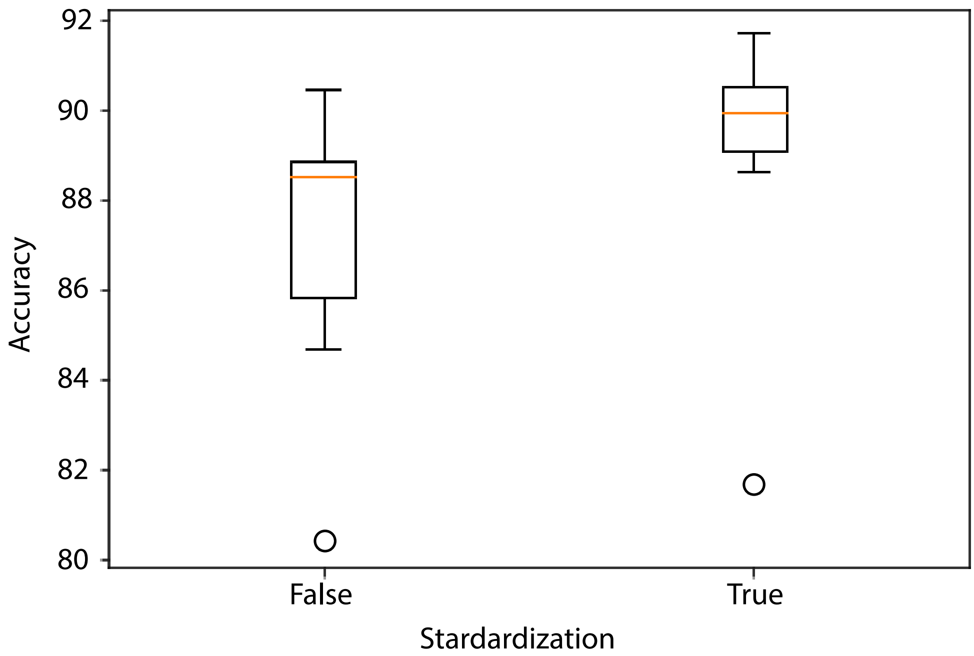

Figure 7.7: One-dimensional convolutional neural networks with and without standardization.

evaluated our model both with and without standardization, and the results are presented in Figure 7.7.

We can notice that by standardizing our dataset, we can easily lift the predictive performance. Without standardization, the accuracy was 87.3%, while with standardization that accuracy increased to 89.3%. For this reason, for the next tests, we are standardizing the input before training the neural networks. The performance of the convolutional network at this point did not surpass the fully connected one we tested before since the former did not have the learning capacity to do so as the latter had with the already engineered features. With the next tests, we are tuning the hyperparameters of the convolutional network in order to achieve higher accuracy.

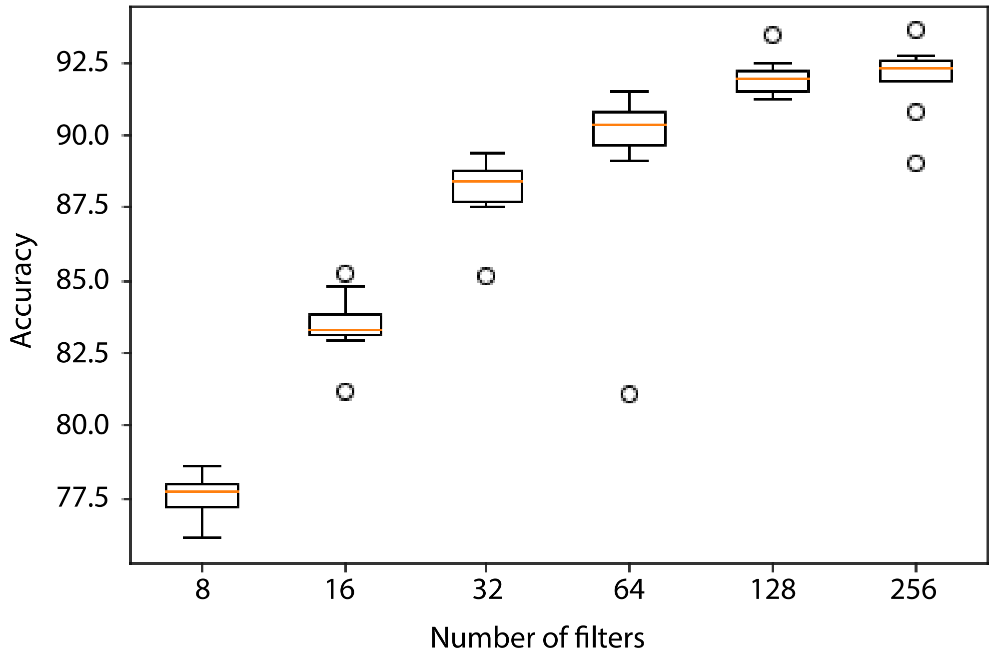

Figure 7.8: One-dimensional convolutional neural networks with different number of filter maps.

7.4.4.2 Number of Filters

In this test, we are exploring how modifying an important hyperparameter of a convolutional neural network such as the number of filters, has an impact on the overall predictive performance of the model. Specifically, we tested the following values for the number of filters: 8, 16, 32, 64, 128, and 256. Figure 7.8 presents the predictive accuracies for the different number of filter maps.

The bigger the number of filters we use, the better the accuracy gets. We have managed to achieve an accuracy of up to 92% using 128 filters. However, the more filters we use, the more computationally demanding fitting the model gets.

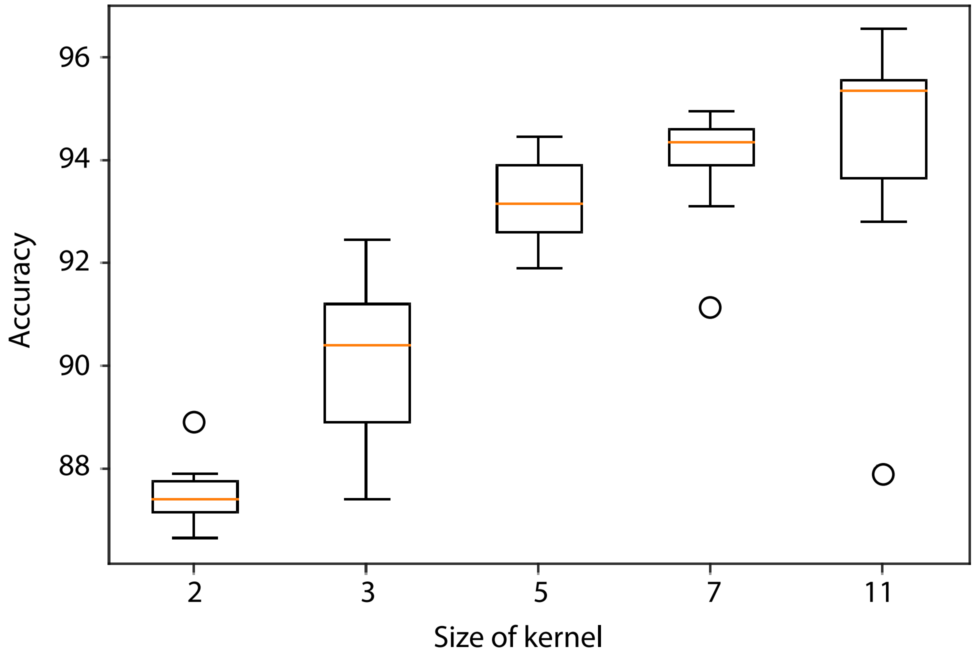

Figure 7.9: One-dimensional convolutional neural networks with different kernel sizes.

7.4.4.3 Size of Kernel

Another important hyperparameter of the one-dimensional convolutional neural network is the size of the kernel. This basically controls the number of time steps that are taken into account from the input data on each read. For our test, we used the following values for the kernel size: 2, 3, 5, 7, and 11. Figure 7.9 presents the predictive accuracies for the different values of the kernel size that were tested.

We achieve the best accuracy for a kernel size of 11 (94.3%). However, the kernel size of 7 provides a better balance between low variance and good performance (94%) and might be a better choice for our case.

7.4.4.4 Discussion

At this point, we can notice that even without any feature engineering, we can achieve very good results with a one-dimensional convolutional neural network by optimizing its hyperparameters. Tradeoffs exist, however. It is computationally more demanding to train a neural network compared to a gradient boosting model. However, in the long run, when new labeled data might become available, having an already trained neural network to retrain might be easier and more efficient than having another machine learning model that will need retraining from scratch.

7.5 Conclusion

This chapter presented a machine learning based, gait recognition system that assists physiotherapists with the postoperative rehabilitation phase of patients who have undergone a lower body operation. The architecture of the system comprising a smartwatch, a smartphone, and a web server was presented. The performance of the system was validated with labeled data recorded by physiotherapists of the Hirslanden Clinique La Colline, an orthopedic clinic in Geneva, Switzerland. Gait patterns of patients were recorded soon after they have undergone various types of lower body operation.

After engineering time and frequency domain parameters, several machine learning models, and a fully connected neural network were tested. The predicted performance of the system reached an accuracy of 94.9% with the best performing classifier among nine different gait classes. One-dimensional convolutional networks were also trained with the raw time-series data of the sensors. After the hyperparameter tuning, a predictive performance of 94.3% was achieved. The innovation of the proposed system lies in the fact that it enables physiotherapists to monitor the evolution of the gait pattern of a patient under rehabilitation, throughout the day, and not only during the defined and time-limited physiotherapy sessions.

There are, however, limitations of the presented work. The collected dataset was imbalanced, and therefore, the trained models are inherently constrained by their ability to predict instances belonging to the underrepresented classes. The trained models, although put in production, were not evaluated in the wild because of data ownership restrictions posed by the partners of the project.

8 Summary of Thesis Achievements

The goal of this thesis was to introduce new methods and applications that support healthy aging via the use of modern smart devices and support postoperative rehabilitation via recognizing activities with different machine learning techniques. The contributions of the thesis start with a user requirement analysis aiming to identify the older adults needs and fears when using modern application and connected devices. Three applications using smart devices are then presented, the first to position the user indoors with the use of Bluetooth Low Energy beacons, the second to detect mood and more specifically stress via monitoring smartphone usage, and the third to detect abnormal behavior through the use of a tablet application. Machine learning methodologies are then presented for activity recognition, first by researching the challenges of designing such an activity recognition system, and then by applying this knowledge to create an online gait recognition system to assist physiotherapists with their existent postoperative rehabilitation protocol.

The proposed innovations have been published to relevant conferences according to the domain of each work. The works have also been validated via the implementation of them in several Swiss national and European projects. In this chapter, the author of the thesis is concluding the work by discussing the contributions, the results, and the importance of each part of the thesis.

User Requirements

All the presented applications are used to enhance wellbeing and thus are obviously meant to be used by humans. It is necessary first to understand people's requirements and take their needs into account before designing and developing any such application. For this reason, we began our research by conducting a questionnaire study in which we targeted the elderly population, which supposedly has been more resistant to new technologies. This study provides useful insights to developers who are building user-centric applications and want to appeal to users of all age groups.

The results of the survey are presented in Chapter 2. In terms of familiarity and interest in using technology, there seem to be different subgroups among older adults. The older someone is, the harder it seems to be to keep up with the technological advances and be willing to use them. There is an age technology gap, but most participants were already using either a computer or a smartphone. Those were also more likely to use new types of devices, such as wearable ones to measure steps and resting time, or connected blood pressure monitors. However, there was still a lack of awareness in several aspects, such as what data sharing means. Social contact is essential to most of the participants, and a strong incentive to use technology is to be able to get in touch with others. Older adults seem to enjoy mental exercises and would enjoy such video games. They would also enjoy guidance towards improving their fitness with proper monitoring and encouragement through a mobile application.

A limitation of this study is that the questionnaires that were used were not validated from other studies. They were developed to guide the development of a tablet application, and therefore the drawn conclusions may not apply to every other case. The number of respondents was also limited.

This questionnaire study provided an insightful view of how elderly feel towards modern technology, health and physical activity tracking, social interaction using technology, and playing games. Older adults feel that in case they get proper explanations, they would be willing to use new kinds of technologies and devices.

This is why developers and researchers should put effort into accurately presenting their applications that promote independent living.

Applications Supporting Healthy Aging Using Smart Devices

During the last decades and especially after the mass adoption of the internet, new technologies have been created that revolutionized every aspect of human life. Modern technologies and applications have been developed to improve human life and to support healthy aging. The thesis contributions in this domain include an indoor localization system with room-level accuracy that is easy to deploy and use, a stress and mood detection system running on a smartphone, and a way to infer abnormal behavior from the usage of a tablet launcher application.

In Chapter 3, a low-cost, easy to use and deploy indoor localization system was presented. The system offers room level accuracy and uses Bluetooth Low Energy beacons. Each beacon is attached to the ceiling in the center of every room, and the algorithm we are proposing takes into account the geometry of each room to create thresholds that ultimately characterize the RSSI readings from the beacons. The proposed system was put into the test in two different locations, an office, and a house environment, where it improved the localization performance for most measured locations.

The limitation of the presented indoor positioning system is the number of different living scenarios that it was put into the test. Its usability was evaluated in a house and an office environment, but there is an infinite number of room arrangements and items that can deteriorate signal propagation and reception. Proving that this approach works, in any case, would be physically impossible.

We are presenting a stress and mood detection system we built in Chapter 4. An Android application was used to monitor user-smartphone interactions. The data the app was monitoring include activity information, steps, number of calls and text messages, and other human-smartphone interactions. During the study that we organized, the participants were giving us feedback about their subjective perception of their psychological state regarding their activeness, happiness, and stress levels. This personal feedback formed the ground truth that we then used for our supervised machine learning models. Several models were tested, and we managed to achieve a lift of predictive performance for all psychological states compared against the corresponding naive models.

The main limitation of the stress and mood detection system comes from the dataset that was used to train the system. There was a limited number of people participating in the experiment belonging to a narrow age group, and therefore the results may not generalize well to the whole population. The collected ground truth may also not be entirely reliable since it was self-reported, but the non-invasive character of the project imposed the way of collecting it.

We continued with a similar contribution in Chapter 5, the last chapter of this part, and we presented techniques to detect abnormal behavior from mobile app usage data. This time a launcher tablet application was used, and all different aspects of the app were recorded. The daily activity of each user was also monitored by measuring the steps and the resting time through a connected wearable device. In contrast to the previous part of our research, we had no direct feedback from the users regarding the normality of their behavior, and so we employed several unsupervised anomaly detection techniques on the available dataset to identify abnormal behavior. We demonstrated that clustering could be of use to detect anomalies in the dataset that can then translate to abnormal behavior and be the trigger of appropriate actions, like informing professional caregivers or family members.

The limitation of this part of the presented work was that an assumption was taken together with professional caregivers regarding abnormal instances in our dataset. This was a necessary step to have a performance evaluation with the available unlabeled dataset. Again the size of the experiment in terms of the participants and the participation time was limited.

In the era of ubiquitous computing, the presented contributions can be used to provide additional contextual information. Indoor localization information can be useful information for navigational purposes, or even critical in accessibility and emergency situations. So the presented low-cost IPS can be of great benefit to anyone wanting to deploy it in a specific area or even a building. Detecting stress, mood, and abnormal behavior through the use of a smartphone or a tablet can be innovative in the way such situations can be dealt with. Due to the omnipresence of those devices, remote caregivers can be notified in order to take necessary actions and handle alarming cases.

Machine Learning for Activity Recognition Supporting Postoperative Rehabilitation

In the frame of building context-aware applications, there has been an increase in the interest for activity recognition systems. More and more applications would benefit from knowing such contextual information to optimize the offered services. We have focused on such systems based on inertial sensors, and our contribution in this domain is twofold. First, all the parameters that affect the design and the implementation of an inertial sensor based activity recognition system are presented, and their impact is evaluated. Then, the design and the optimization of an online activity recognition system is presented. This system was deployed and used in a medical environment to provide useful information to physiotherapists.

In Chapter 6, we have presented research challenges in human activity recognition using inertial sensors. We have identified every single parameter that impacts the design of an AR system, and we have analyzed the significance of all of them. Some parameters have to do with the characteristics of the data, and other challenges arise from the available dataset. Data handling should be treated carefully as it is an essential step in all applications that involve machine learning. Lastly, some parameters are specific to each application. All those parameters have a direct impact on the design, implementation, testing, and evaluation phase of every AR system. To exemplify this impact, we have presented an experiment we organized and included several activities to be identified. Many machine learning models were trained, each time by tinkering different parameters. The predictive accuracy is greatly fluctuating among those models, a fact that highlights the importance of critically and correctly deciding on all the parameters as mentioned above.

As with any real-world application, the experiment was limited in terms of the selected activities of interest, and the collected dataset. The dataset was limited in terms of participants and activities durations. Despite that, the presented considerations remain relevant to more complex situations.

We are extending our previous work in Chapter 7, and we are presenting the design of an activity recognition system able to detect different gait patterns including walking with crutches, walking with different frames, limping, and walking without any aid. The proposed system was trained, tested, and validated with data of people who have undergone lower body orthopedic surgery, recorded by Hirslanden Clinique La Colline, an orthopedic clinic in Geneva. We also evaluated the use of a fully connected as well as a convolutional neural network in order to make the proposed system more dynamic and able to adapt to different scenarios. The system is meant to be used by physiotherapists during the rehabilitation phase of a patient. Using the online system we are proposing, physiotherapists are able to monitor the evolving gait pattern of patients during everyday life, and are not restricted by the time limited physiotherapy sessions.

A limitation of the presented online system was the imbalanced dataset that was used for its initial training. This may initially constrain the system's abilities to predict instances belonging to the underrepresented classes. It was also impossible to evaluate the system in the wild because of data ownership restrictions posed by the partners of the project.

Both our contributions in this part of the thesis provide useful guidance to any activity recognition system designer. The first contribution is meant to be used by designers of all levels and can serve as an introduction on designing an AR system while guaranteeing that all critical aspects of the design of it are taken into account. The second contribution, less abstract and more application oriented, presented a real-world use for an AR system and presented the whole infrastructure supporting it.

9 Future Directions

The landscape of technology is growing rapidly, and its pace of growth is not slowing down. A considerable number of new items and devices, from the most common ones to the most unexpected ones, including ovens and shoes, are connected to the internet and are becoming smart. All these connected objects form the Internet of Things (IoT) and enable people to remotely control or monitor various parameters in order to make our lives easier and more comfortable.

The Internet has already and will bring even more changes in our lives. Like smartphones, more IoT devices will become necessities. New concepts, like the ones of smart homes and smart cities, have recently emerged. Future projections include self-driving connected smart cars, with artificial intelligence playing an essential role in vehicle diagnostics and security. Smart connected cities will consist of smart parking systems and smart traffic lights with real-time information to ameliorate daily commuting.

IoT technologies have the potential to revolutionize entire industries. The energy industry can benefit from the smart grid and produce and distribute electricity more efficiently. The healthcare industry can benefit by constantly monitoring several human parameters, by diagnosing many diseases earlier than usual, and by further improving human living conditions. Future predictions expect the global IoT in the healthcare market to reach a value of 534.3 billion USD by the year 2025 [175].

Indoor Positioning

There has been a great effort during the last two decades in improving positioning indoors. Several indoor positioning systems have been developed using different technologies. Each technology has different advantages, but oftentimes brings several restrictions, in terms of accuracy, cost, complexity, and performance. The Ultra-wideband (UWB) technology is very robust and precise but requires expensive installation and units [176]. Potential future integration of UWB with mobile smart devices would make the use of this technology widespread.

Other technologies that are coming soon can be used to position people indoors or to assist in providing indoor localization services, even though they have a different core goal. The European GNNS named Galileo is expected to be fully in service by 2020 and is currently at a functional state already from 2016. Galileo, when combined with other GNSS like the GPS or GLONASS, significantly improves accuracy in challenging environments, including when indoors or in urban canyons [177]. The fifth generation wireless technology, also known as 5G, is at the deployment stage in several countries, including Switzerland, where the technology will launch in 2020. Indoor positioning techniques have already been proposed to position the user using the emerging dense 5G networks [178].

Stress Detection

Detecting stress in daily life remains an open challenge for researchers. Human physiology poses a core restriction. Different emotions can result in similar physiological signals like heart rate and respiration rate [179]. So clearly discriminating between stress and other emotions is an open research problem.

When detecting stress outside a laboratory setting, the stimuli are practically unrestricted. Acquiring all information about a person at all times is impossible. It is, therefore, useful to collect contextual and environmental information about each user in order to improve the stress detection system's performance.

Since for the development of stress detection systems, we rely on subjective user reports and questionnaires, the labels are bound to be unreliable. Ground truth is subjective, and it might be impossible for the system to generalize well. In one experiment [56], in the same situation and even with similar physiological signals, different people may either report stressed or not stressed at all. To tackle this challenge, a personalized user subjectivity test may be performed in controlled laboratory settings in order to calibrate to the user's stress evaluations.

Another challenge of daily life experiments comes from the limited battery life of smartphones and wearable devices. A typical modern smartwatch merely lasts around 4 hours when all its sensors are active. Research should also be done towards optimizing the use of sensors in a way that would not decrease the stress detection accuracy by much.

Abnormal Behavior Detection

The definition of abnormal behavior has some degree of ambiguity and usually depends on the application. Behaviors and activities that can be detected from a smartphone and a smartwatch are complex and can have a great variety in a non constrained environment. Abnormal behaviors may also have a spatiotemporal dependence, in the sense that the same activity might be either normal or abnormal depending on the time and the location of the user. Attempts to model normal behaviors, because of the abundance of them in the dataset, are easier than starting with abnormal events as the baseline. However, detecting an unseen behavior does not necessarily mean that it is an abnormal one since it could just be an unforeseen instance of normal behavior.

Future researches should also focus on aggregating datasets suitable to be used as benchmarks for evaluating abnormal behavior detection algorithms. Doing so in real life is very challenging because of the practically infinite variety of abnormal behaviors that can exist and because of the scarcity of them.

Activity Recognition

Most of the current activity recognition systems are developed and tested in a laboratory setting. Deploying such a system in real-world conditions is challenging, and the performance of the system is doomed to worsen if the system can not adapt to new environments. Commercially available systems like Google's Activity Recognition API [90] have so far restricted functionality, in the sense that they include a limited and relatively small set of activities that can be detected. New systems should expand the set of activities that can be detected and should be tested in noisy real-world conditions.

Open activity recognition datasets would greatly benefit the community. By using shared data, the community can better focus on building optimal models and comparing the results among them. The most well known open activity recognition dataset using smartphone sensors [135] was shared with the community in 2012. Unlike other public datasets in different domains, e.g., available handwriting recognition datasets, the activity recognition one is far from massive and consists of only recordings from 30 subjects. So an extra effort towards creating datasets the whole community can benefit from should be encouraged.

Another possibility is the organization of activity recognition contests, similar to the ones held by conferences for different problems requiring machine learning. In these contests, different teams compete to achieve the best recognition on a given labeled dataset. An online data science platform like Kaggle makes it very easy to start such a competition nowadays, given the availability of a suitable dataset.

The task of defining a set of activities to recognize can significantly vary and depends on the researchers' perceptions. It would be useful to focus on creating taxonomies of activities based on common sense. These taxonomies may form a tree where specific activities are a subset of others, like for example playing basketball being a subset of doing sports. Valuable research contributions can also be made towards assigning probabilities surrounding activities. These probabilities can be either contextual based (e.g., sleeping at midnight is more probable than hiking), or based on simple prior knowledge (e.g., drinking water is more probable than bungee jumping).

Depending on the defined activities and the complexity of the developed activity recognition systems, future AR systems can explore the detection of concurrent activities. Humans are spontaneous and tend to do several activities at the same time or interleaved. Examples may include talking on the phone while walking and drinking water while eating. Depending on the required discriminatory granularity of the system, uncertainty should be handled to avoid ambiguous behavior interpretations.

A Questionnaire for the User Requirements Study

In this appendix, the full questionnaire used in the user requirements study is presented. It consists of eight subsections, one for the demographics of the participants, and one for each of the seven groups of questions of Table A.1.

Table A.1: Number of questions per thematic category of the questionnaire.

Group	No of questions
Games in general	5
Computer games	5
Competition and consumer behavior	6
Social background	7
Health	8
Local opportunities	3
Technology acceptance	7

A.1 Demographics

To give a better insight regarding the participants in our study, Table A.2 includes more information regarding the demographics of our sample.

Table A.2: Demographics of the participants.

Variable	Value	n	n/N(%)
Gender	Female	68	51.1
	Male	65	48.9

Continued on the next page

Table A.2 – continued from the previous page.

Variable	Value	n	n/N(%)
Age	Less than 60	12	9
	60-64	9	6.8
	65-69	27	20.3
	70-74	27	20.3
	75-79	34	25.6
	80-84	12	9
	85-89	4	3
	90+	6	4.5
	No answer	2	1.5
Nationality	Switzerland	78	58.6
	United Kingdom	52	39.1
	Other	3	2.3
Qualification	University	46	34.6
	College	23	17.3
	Work-based	32	24.1
	18 year old (A levels)	23	17.3
	16 year old (O levels)	5	3.8
	None	2	1.5
	No answer	2	1.5
Accommodation	Own home	111	83.5
	Sheltered/Warden accommodation	4	3
	Care village	13	9.8
	With family	3	2.3
	Residential home	1	0.8
	Nursing home	1	0.8
Mobility (multiple choice)	Walk unaided	101	75.9
	Drive car, motorbike or bicycle	100	75.2
	Public transport	73	54.9
	Walking frame or stick	15	11.3
	Wheelchair	3	2.3

Continued on the next page

Table A.2 – continued from the previous page.

Variable	Value	n	n/N(%)
	No answer	1	0.8
Domestic support	Not at all	94	70.7
	Rarely	15	11.3
	Once or more per month	6	4.5
	Once or more per week	11	8.3
	Once or more per day	4	3
	No answer	3	2.3
Memory problems	Not at all	26	19.5
	Very occasionally	62	46.6
	Rarely	25	18.8
	Sometimes	16	12
	Often	3	2.3
	No answer	1	0.8
Anxiety or depression	Not at all	79	59.4
	Occasionally	33	24.8
	Rarely	10	7.5
	Sometimes	5	3.8
	Often	3	2.3
	All the time	1	0.8
	No answer	2	1.5
Medication	None	34	25.6
	One drug	26	19.5
	Two different drugs	21	15.8
	Three different drugs	14	10.5
	Four different drugs	13	9.8
	Five or more drugs	23	17.3
	No answer	2	1.5

A.2 Games in General

Table A.3 includes all information regarding the responses to the questions about games in general.

Table A.3: Questions about games in general.

Question	Answer	n	n/N(%)
How often do you play some sort of game?	Daily	21	15.8
	More than once a week	21	15.8
	Weekly	11	8.3
	Monthly	23	17.3
	Several times a year	30	22.6
	Never	13	9.8
	No answer	14	10.5
How often do you play table or board games?	Daily	15	11.3
	More than once a week	15	11.3
	Weekly	13	9.8
	Monthly	26	19.5
	Several times a year	40	30.1
	Never	23	17.3
	No answer	1	0.8
Why don't you play games more frequently? (multiple choice)	No time	36	27.1
	No partners to play with	28	21.1
	Not inspired by playing	40	30.1
	Do not know what games to play	3	2.3
	Do not like available types of games	3	2.3
	Games too complicated or complex	3	2.3
	Physical impairments (e.g. eyesight)	3	2.3
	No answer	17	12.8
Why do you play games? (multiple choice)	Good for my brain	74	55.6
	Social contact/inclusion	53	39.8
	Because it is fun	80	60.2

Continued on the next page

Table A.3 – continued from the previous page.

Question	Answer	n	n/N(%)
	To pass the time	39	29.3
	Nothing better to do	3	2.3
	I like to win things	11	8.3
	To gain more general knowledge	13	9.8
	I do not play games	20	15
	No answer	1	0.8
What types of multiplayer games do you like? (multiple choice)	Playing on my own	50	37.6
	Lots of individual players	36	27.1
	Play in pairs	42	31.6
	Play in a team	37	27.8
	Play against a computer	23	17.3
	I do not like playing games	16	12
	No answer	18	13.5

A.3 Computer Games

Table A.4 includes all information regarding the responses to the questions about computer and tablet games.

Table A.4: Questions about computer games.

Question	Answer	n	n/N(%)
Would you like to play games on a computer or tablet?	Yes, would love to be able to	30	22.6
	Yes, would quite like to	30	22.6
	Maybe yes	12	9
	Not very likely	22	16.5
	Definitely not	29	21.8
	I am not sure	10	7.5

Continued on the next page

Table A.4 – continued from the previous page.

Question	Answer	n	n/N(%)
Who would you play games with? (multiple choice)	Friends	75	56.4
	Relatives	52	39.1
	Club members	8	6
	Anyone I know	30	22.6
	Strangers	7	5.3
	I only play games alone	36	27.1
	I do not like playing games	12	9
	No answer	2	1.5
How interested are you in competing against others on a tablet?	Yes, would love to be able to	6	4.5
	Yes, would quite like to	17	12.8
	Maybe yes	13	9.8
	Not very likely	24	18
	Definitely not	44	33.1
	I am not sure	24	18
	No answer	5	3.8
Which of the following statement reflects your attitude most?	Only the prize makes winning matter	1	0.8
	Would quite like to get a prize	10	7.5
	Nice to get a prize	24	18
	Do not mind much about the prize	23	17.3
	Happy just to take part	45	33.8
	Do not want to take part in games	29	21.8
	No answer	1	0.8
Which prizes might inspire you to participate? (multiple choice)	Old-time music/cinema stars videos	8	6
	Recipes from your chosen chef	4	3
	Pictures of famous artworks	3	2.3
	Stories of important world events	6	4.5
	Stories of important national events	7	5.3
	Photos and videos from your family	11	8.3
	Famous poetry	2	1.5
	Digital postage stamps / trophies	4	3

Continued on the next page

Table A.4 – continued from the previous page.

Question	Answer	n	n/N(%)
	Free cup of tea at a local place	17	12.8
	Cash prize or discount voucher	34	25.6
	None of the above	57	42.9
	No answer	18	13.5

A.4 Competition and Consumer Behavior

Table A.5 includes all information regarding the responses to the questions about the users' attitude towards competition and consumer behavior.

Table A.5: Questions about competition and consumer behavior.

Question	Answer	n	n/N(%)
What kind of incentives influence your shopping decisions? (multiple choice)	Collecting points on loyalty cards	53	39.8
	Getting benefits customer cards	6	4.5
	Trophy for the customer of the year	1	0.8
	Challenge to have something special	9	6.8
	Consumer needs	70	52.6
	Peer-group pressure	1	0.8
	Advertising	11	8.3
	Just love going shopping	22	16.5
	No answer	14	10.5
How often do you receive points or use loyalty cards in your purchasing?	All the time	36	27.1
	Most of the time	37	27.8
	Sometimes	32	24.1
	Rarely	15	11.3
	Very rarely	5	3.8
	Never	8	6

Continued on the next page

Table A.5 – continued from the previous page.

Question	Answer	n	n/N(%)
Do you choose where to shop or what to buy to benefit from loyalty rewards?	All the time	8	6
	Most of the time	24	18
	Sometimes	39	29.3
	Rarely	30	22.6
	Very rarely	14	10.5
	Never	17	12.8
	No answer	1	0.8
How often do you play the lottery or participate in competitions?	Daily	3	2.3
	More than once a week	11	8.3
	Weekly	7	5.3
	More than once a week	10	7.5
	Rarer	35	26.3
	Never	65	48.9
	No answer	2	1.5
How often do you watch competitive sports either live or on TV?	Daily	9	6.8
	More than once a week	21	15.8
	Weekly	25	18.8
	More than once a month	21	15.8
	Rarer	27	20.3
	Never	30	22.6
Which of the following statements regarding winning do you agree with most?	I like to win	25	18.8
	I am happy for others to win	14	10.5
	I like to just take part	70	52.6
	I like to help others win	1	0.8
	I do not like to compete	22	16.5
	No answer	1	0.8

A.5 Social Background

Table A.6 includes all information regarding the responses to the questions about the social background of the participants.

Table A.6: Questions about social background.

Question	Answer	n	n/N(%)
How important is social contact for you?	Very important	63	47.4
	Important	45	33.8
	Maybe important	9	6.8
	Not important	5	3.8
	Definitely not important	11	8.3
Can you imagine engaging with more people using technology?	Definitely yes	21	15.8
	Yes	32	24.1
	Maybe	17	12.8
	Hardly	30	22.6
	Definitely not	24	18
	I am not sure	9	6.8
How often do you have regular contact with your family through visits or phone calls?	Every day	32	24.1
	More than once a week	49	36.8
	Weekly	21	15.8
	More than once a month	20	15
	Rarer	6	4.5
	Not at all	5	3.8
How often do you go out and about?	Daily	65	48.9
	More than once a week	46	34.6
	Weekly	13	9.8
	More than once a month	6	4.5
	Rarer	3	2.3
How often do you see people?	Daily	54	40.6
	More than once a week	45	33.8
	Weekly	18	13.5

Continued on the next page

Table A.6 – continued from the previous page.

Question	Answer	n	n/N(%)
	More than once a month	11	8.3
	Rarer	4	3
	No answer	1	0.8
Would you be willing to allow certain family members see how you are getting on with using the technology?	Yes, that would be fine	30	22.6
	Yes, if I knew exactly who	18	13.5
	Yes, if they check with me	26	19.5
	No, I don't think so	31	23.3
	I am not sure	26	19.5
	No answer	2	1.5
Would you be willing to allow certain healthcare professionals see how you are getting on with using the technology?	Yes, that would be fine	18	13.5
	Yes, if I knew exactly who	29	21.8
	Yes, if they check with me	32	24.1
	No, I don't think so	32	24.1
	I am not sure	19	14.3
	No answer	3	2.3

A.6 Health

Table A.7 includes all information regarding the responses to the questions about health monitoring and tracking.

Table A.7: Questions about health monitoring and tracking.

Question	Answer	n	n/N(%)
Would you like to improve your fitness?	Would love to	34	25.6
	Would quite like to	60	45.1
	Maybe yes	16	12
	Maybe no	4	3

Continued on the next page

Table A.7 – continued from the previous page.

Question	Answer	n	n/N(%)
	Not very likely	7	5.3
	Definitely not	8	6
	No answer	4	3
Would you enjoy mental exercise?	Would love to	37	27.8
	Would quite like to	59	44.4
	Maybe yes	15	11.3
	Maybe no	11	8.3
	Not very likely	2	1.5
	Definitely not	7	5.3
	No answer	2	1.5
Would you like encouragement from others?	Would love to	14	10.5
	Would quite like to	27	20.3
	Maybe yes	30	22.6
	Maybe no	20	15
	Not very likely	19	14.3
	Definitely not	17	12.8
	No answer	6	4.5
Would you be interested in wearing a measuring device?	Would love to	9	6.8
	Would quite like to	28	21.1
	Maybe yes	35	26.3
	Maybe no	16	12
	Not very likely	12	9
	Definitely not	29	21.8
	No answer	4	3
Would you like certain family members to see how you are getting on?	Would love to	7	5.3
	Would quite like to	18	13.5
	Maybe yes	32	24.1
	Maybe no	15	11.3
	Not very likely	16	12
	Definitely not	38	28.6

Continued on the next page

Table A.7 – continued from the previous page.

Question	Answer	n	n/N(%)
	No answer	7	5.3
Which of the following monitoring devices would you be willing to test? (multiple choice)	Bracelet to measure steps taken	43	32.3
	Clip-on pedometer (steps)	55	41.4
	Blood pressure monitor	60	45.1
	Weight scale	34	25.6
	Blood glucose monitor	35	26.3
	Passive home monitoring	23	17.3
	Panic alarm	46	34.6
	Home safety (e.g. smoke, intruder)	31	23.3
	No answer	25	18.8
Would you like to be able to see a health report?	Yes	51	38.3
	No	47	35.3
	Do not know	27	20.3
	No answer	8	6
Which of these aspects of health improvement training are appealing to you? (multiple choice)	Train with other people	39	29.3
	Feedback/reward to continue exercise	23	17.3
	Get a clear goal what should be done	48	36.1
	Have a competition with others	1	0.8
	Pass different levels of difficulty	25	18.8
	Points for every successful exercise	12	9
	Limited time for doing the exercise	25	18.8
	No answer	33	24.8

A.7 Local Opportunities and Volunteering

Table A.8 includes all information regarding the responses to the questions about local opportunities and volunteering.

Table A.8: Questions about local opportunities and volunteering.

Question	Answer	n	n/N(%)
Would you be interested in seeing volunteering opportunities in your area?	Yes	60	45.1
	No	38	28.6
	Do not know	24	18
	No answer	11	8.3
Would you be interested in seeing volunteering opportunities on your tablet?	Yes	46	34.6
	No	43	32.3
	Not sure	22	16.5
	No answer	22	16.5
Would you be willing for local organisations to know about the skills you are willing to volunteer on a tablet?	Yes, would love to	9	6.8
	Yes, would quite like to	16	12
	Maybe	25	18.8
	Not very likely	19	14.3
	Definitely not	34	25.6
	Not sure	29	21.8
	No answer	1	0.8

A.8 Technology Acceptance

Table A.9 includes all information regarding the previous experience and the attitude of the participants towards technology.

Table A.9: Questions about technology acceptance.

Question	Answer	n	n/N(%)
Which internet connection do you have?	High speed broadband at home	28	21.1
	Ordinary broadband at home	59	44.4
	Internet from my phone	14	10.5

Continued on the next page

Table A.9 – continued from the previous page.

Question	Answer	n	n/N(%)
	Other people's houses or places	1	0.8
	When at family members houses	3	2.3
	No internet connection	24	18
	No answer	4	3
Which of the following statements describes your use of a computer, tablet, or smartphone?	I have one, and use it all the time	70	52.6
	I have one, and use it regularly	24	18
	I have one, and use it sometimes	7	5.3
	I have one, but I rarely use it	9	6.8
	I have one, but I never use it	1	0.8
	Do not own any computers	20	15
	No answer	2	1.5
What is the most prevalent reason for not using a smartphone, tablet, or computer?	Do not understand technology	8	6
	Do not need to use technology	4	3
	Do not trust technological devices	2	1.5
	Do not know	3	2.3
	I already do use technology	41	30.8
	No internet access	5	3.8
	No answer	70	52.2
How interested are you in learning to use a tablet?	Already do	57	42.9
	Very interested in	14	10.5
	Interested in	21	15.8
	Maybe interested in	16	12
	Less interested in	5	3.8
	Definitely not interested in	18	13.5
	No answer	2	1.5
If interested, in which way would you like to learn it? (multiple choice)	Can learn it by myself	46	34.6
	Need a written instruction	20	15
	Need a video instruction	4	3
	Need individual/practical help	24	18
	Not sure	7	5.3

Continued on the next page

Table A.9 – continued from the previous page.

Question	Answer	n	n/N(%)
	No answer	32	24.1
Why do you think you would (or currently do) use technology? (multiple choice)	Staying in touch with people	100	75.2
	Staying up-to-date (e.g. news)	96	72.2
	For my work	33	24.8
	Maintain a hobby	49	36.8
	Shopping	43	32.3
	Music, video, TV	35	26.3
	Games	43	32.3
	Learning new things	66	49.6
	Not interested	1	0.8
	No answer	13	9.8
Which of the following statements regarding data security do you agree with most?	Not worried, shows trust	40	30.1
	I am worried, but happy	53	39.8
	Worried, asking family for advice	21	15.8
	Very worried about stolen data	9	6.8
	I do not know	2	1.5
	I do not think about data security	3	2.3
	No answer	13	9.8

Bibliography

[1] AAL, “Aal programme - active assisted living programme - ageing well,” http://www.aal-europe.eu/about/, 2019, [Online; accessed Feb 27, 2019].

[2] StayActive, “The stayactive project,” http://tam.unige.ch/projects/stayactive/, 2016, [Online; accessed Feb 27, 2019].

[3] SmartHeat, “The smartheat project,” http://tam.unige.ch/projects/smartheat/, 2018, [Online; accessed Feb 27, 2019].

[4] EDLAH2, “The edlah2 project,” http://tam.unige.ch/projects/edlah2/, 2018, [Online; accessed Feb 27, 2019].

[5] Innosuisse, “Innosuisse - swiss innovation agency,” https://www.innosuisse.ch/inno/en/home/ueber-uns/auftrag.html, 2019, [Online; accessed Feb 27, 2019].

[6] F2D, “The f2d project,” http://tam.unige.ch/projects/f2d/, 2015, [Online; accessed Feb 27, 2019].

[7] RecoverAtHome, “The recoverathome project,” http://tam.unige.ch/projects/recoverathome/, 2018, [Online; accessed Feb 27, 2019].

[8] FNS, “Mission statement - swiss national science foundation - fns,” http://www.snf.ch/en/theSNSF/profile/mission-statement/Pages/default.aspx, 2019, [Online; accessed Feb 27, 2019].

[9] FLTL, “The from lab to life: Cognitive aging revisited project,” https://cigev.unige.ch/index.php?cID=883, 2019, [Online; accessed Feb 27, 2019].

[10] A. I. Kyritsis, J. Nuss, L. Holding, P. Rogers, M. O'Connor, P. Kostopoulos, M. Suffield, M. Deriaz, and D. Konstantas, "User requirement analysis for the design of a gamified ambient assisted living application," in *International Conference on Computers Helping People with Special Needs*. Springer, 2018, pp. 426–433.

[11] A. I. Kyritsis, P. Kostopoulos, M. Deriaz, and D. Konstantas, "A ble-based probabilistic room-level localization method," in *2016 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, 2016, pp. 1–6.

[12] A. I. Kyritsis, M. Deriaz, and D. Konstantas, "Anomaly detection techniques in mobile app usage data among older adults," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2018, pp. 1–6.

[13] ——, "Considerations for the design of an activity recognition system using inertial sensors," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2018, pp. 1–8.

[14] A. I. Kyritsis, G. Willems, M. Deriaz, and D. Konstantas, "Gait recognition with smart devices assisting postoperative rehabilitation in a clinical setting," in *2018 IEEE 1st International Conference on Artificial Intelligence for Industries (ai4i)*. IEEE, 2018, pp. 1–5.

[15] ——, "Gait pattern recognition using a smartwatch assisting postoperative physiotherapy," *International Journal of Semantic Computing*, 2019.

[16] Eurostat, "Population structure and ageing," http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing, June 2017, [Online; accessed Nov 13, 2017].

[17] United Nations, Department of Economic and Social Affairs, Population Division, "Population ageing and sustainable development no. 2017/1,"

http://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFacts_2017-1.pdf, June 2017, [Online; accessed Nov 13, 2017].

[18] N. Charness and W. R. Boot, "Aging and information technology use: Potential and barriers," *Current Directions in Psychological Science*, vol. 18, no. 5, pp. 253–258, 2009.

[19] Y.-h. Wu, S. Damnée, H. Kerhervé, C. Ware, and A.-S. Rigaud, "Bridging the digital divide in older adults: a study from an initiative to inform older adults about new technologies," *Clinical interventions in aging*, vol. 10, p. 193, 2015.

[20] N. Wagner, K. Hassanein, and M. Head, "Computer use by older adults: A multi-disciplinary review," *Computers in human behavior*, vol. 26, no. 5, pp. 870–882, 2010.

[21] World Health Organization, "Proposed working definition of an older person in africa for the mds project," https://www.who.int/healthinfo/survey/ageingdefnolder/en/, 2019, [Online; accessed Sep 09, 2019].

[22] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE journal of biomedical and health informatics*, vol. 17, no. 3, pp. 579–590, 2013.

[23] M. Memon, S. R. Wagner, C. F. Pedersen, F. H. A. Beevi, and F. O. Hansen, "Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes," *Sensors*, vol. 14, no. 3, pp. 4312–4341, 2014.

[24] R. M. Baecker, K. Moffatt, and M. Massimi, "Technologies for aging gracefully," *interactions*, vol. 19, no. 3, pp. 32–36, 2012.

[25] N. Shapira, A. Barak, and I. Gal, "Promoting older adults' well-being through internet training and use," 2007.

[26] N. Marangunić and A. Granić, "Technology acceptance model: a literature review from 1986 to 2013," *Universal Access in the Information Society*, vol. 14, no. 1, pp. 81–95, 2015.

[27] N. Selwyn, "The information aged: A qualitative study of older adults' use of information and communications technology," *Journal of Aging studies*, vol. 18, no. 4, pp. 369–384, 2004.

[28] F. de Vette, M. Tabak, M. Dekker-van Weering, and M. Vollenbroek-Hutten, "Engaging elderly people in telemedicine through gamification," *JMIR serious games*, vol. 3, no. 2, 2015.

[29] D. L. Kappen, L. E. Nacke, K. M. Gerling, and L. E. Tsotsos, "Design strategies for gamified physical activity applications for older adults," in *System Sciences (HICSS), 2016 49th Hawaii International Conference on.* IEEE, 2016, pp. 1309–1318.

[30] B. Lee, Y. Chen, and L. Hewitt, "Age differences in constraints encountered by seniors in their use of computers and the internet," *Computers in Human Behavior*, vol. 27, no. 3, pp. 1231–1237, 2011.

[31] T. Broady, A. Chan, and P. Caputi, "Comparison of older and younger adults' attitudes towards and abilities with computers: Implications for training and learning," *British Journal of Educational Technology*, vol. 41, no. 3, pp. 473–485, 2010.

[32] G. Nimrod, "Seniors' online communities: A quantitative content analysis," *The Gerontologist*, vol. 50, no. 3, pp. 382–392, 2009.

[33] V. P. Gonçalves, V. P. de Almeida Neris, S. Seraphini, T. C. Dias, G. Pessin, T. Johnson, and J. Ueyama, "Providing adaptive smartphone interfaces targeted at elderly people: an approach that takes into account diversity among the elderly," *Universal Access in the Information Society*, vol. 16, no. 1, pp. 129–149, 2017.

[34] B. W. Parkinson, "The global positioning system (navstar)," *Bulletin géodésique*, vol. 53, no. 2, pp. 89–108, 1979.

[35] L. Fang, P. Antsaklis, L. Montestruque, M. McMickell, M. Lemmon, Y. Sun, H. Fang, I. Koutroulis, M. Haenggi, M. Xie, and X. Xie, “Design of a wireless assisted pedestrian dead reckoning system - the navmote experience,” *Instrumentation and Measurement, IEEE Transactions on*, vol. 54, no. 6, pp. 2342–2358, Dec 2005.

[36] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’00. New York, NY, USA: ACM, 2000, pp. 32–43. [Online]. Available: http://doi.acm.org/10.1145/345910.345917

[37] G. G. Anagnostopoulos and M. Deriaz, “Automatic switching between indoor and outdoor position providers,” in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, Oct 2015, pp. 1–6.

[38] Y. Li, P. Zhang, H. Lan, Y. Zhuang, X. Niu, and N. El-Sheimy, “A modularized real-time indoor navigation algorithm on smartphones,” in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, Oct 2015, pp. 1–7.

[39] A. Bekkelien, M. Deriaz, and S. Marchand-Maillet, “Bluetooth indoor positioning,” *Master’s thesis, University of Geneva*, 2012.

[40] P. Kostopoulos, T. Nunes, K. Salvi, M. Deriaz, and J. Torrent, “F2d: A fall detection system tested with real data from daily life of elderly people,” in *e-Health Networking, Applications and Services (Healthcom), 2015 IEEE 17th International Conference on*, Oct 2015.

[41] C. Papamanthou, F. P. Preparata, and R. Tamassia, “Algorithms for location estimation based on rssi sampling,” in *Algorithmic Aspects of Wireless Sensor Networks*. Springer, 2008, pp. 72–86.

[42] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, vol. 10, no. 1, pp. 91–102, Jan. 1992. [Online]. Available: http://doi.acm.org/10.1145/128756.128759

[43] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ser. MobiCom '99. New York, NY, USA: ACM, 1999, pp. 59–68. [Online]. Available: http://doi.acm.org/10.1145/313451.313476

[44] P. Bahl and V. N. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, 2000, pp. 775–784 vol.2.

[45] C. Wu, Z. Yang, Y. Liu, and W. Xi, "Will: Wireless indoor localization without site survey," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 839–848, April 2013.

[46] K. D'hoe, G. Ottoy, J. P. Goemaere, and L. De Strycker, "Indoor room location estimation," *Advances in electrical and computer engineering*, vol. 8, no. 2, pp. 78–81, 2008.

[47] J. Hallberg, M. Nilsson, and K. Synnes, "Positioning with bluetooth," in *Telecommunications, 2003. ICT 2003. 10th International Conference on*, vol. 2, Feb 2003, pp. 954–958 vol.2.

[48] R. Faragher and R. Harle, "Location fingerprinting with bluetooth low energy beacons," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, Nov 2015.

[49] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, Jul 1993.

[50] Bluetooth SIG, "Bluetooth core specification v4.2," https://www.bluetooth.com/specifications/adopted-specifications, Dec 2014, [Online; accessed Feb 21, 2016].

[51] R. A. Fisher, "On the interpretation of χ 2 from contingency tables, and the calculation of p," *Journal of the Royal Statistical Society*, vol. 85, no. 1, pp. 87–94, 1922.

[52] T. S. Langner and S. T. Michael, "Life stress and mental health: Ii. the midtown manhattan study." 1963.

[53] R. W. Picard, "Affective computing-mit media laboratory perceptual computing section technical report no. 321," *Cambridge, MA*, vol. 2139, 1995.

[54] K. A. Moehle and E. E. Levitt, "The history of the concepts of fear and anxiety," in *Clinical Psychology*. Springer, 1991, pp. 159–182.

[55] Harvard Medical School, "Understanding the stress response," https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response/, 2018, [Online; accessed Feb 01, 2019].

[56] A. Liapis, C. Katsanos, D. Sotiropoulos, M. Xenos, and N. Karousos, "Stress recognition in human-computer interaction using physiological and self-reported data: a study of gender differences," in *Proceedings of the 19th Panhellenic Conference on Informatics*. ACM, 2015, pp. 323–328.

[57] J. G. Rodrigues, M. Kaiseler, A. Aguiar, J. P. S. Cunha, and J. Barros, "A mobile sensing approach to stress detection and memory activation for public bus drivers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3294–3303, 2015.

[58] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of biomedical informatics*, vol. 59, pp. 49–75, 2016.

[59] European Agency for Safety and Health at Work, “E-guide to managing stress and psychosocial risks,” https://osha.europa.eu/en/tools-and-publications/e-guide-managing-stress-and-psychosocial-risks/, 2019, [Online; accessed Feb 02, 2019].

[60] ——, “Report - research on work-related stress,” https://osha.europa.eu/en/publications/reports/203/view/, 2000, [Online; accessed Feb 02, 2019].

[61] ——, “Healthy workers, thriving companies - a practical guide to wellbeing at work,” https://osha.europa.eu/en/tools-and-publications/publications/healthy-workers-thriving-companies-practical-guide-wellbeing/view/, 2018, [Online; accessed Feb 03, 2019].

[62] R. W. Picard, “Automating the recognition of stress and emotion: from lab to real-world impact,” *IEEE MultiMedia*, vol. 23, no. 3, pp. 3–7, 2016.

[63] J. R. Stroop, “Studies of interference in serial verbal reactions.” *Journal of experimental psychology*, vol. 18, no. 6, p. 643, 1935.

[64] D. Lowery, R. B. Fillingim, and R. A. Wright, “Sex differences and incentive effects on perceptual and cardiovascular responses to cold pressor pain,” *Psychosomatic medicine*, vol. 65, no. 2, pp. 284–291, 2003.

[65] P. J Lang, M. M Bradley, and B. N & Cuthbert, “International affective picture system (iaps): Affective ratings of pictures and instruction manual (rep. no. a-8),” *Technical Report A-8*, 01 2008.

[66] S. Greene, H. Thapliyal, and A. Caban-Holt, “A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health,” *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 44–56, 2016.

[67] Apple, “Apple watch series 4 - health - apple,” https://www.apple.com/lae/apple-watch-series-4/health/, 2018, [Online; accessed Feb 04, 2019].

[68] K. S. Rahnuma, A. Wahab, N. Kamaruddin, and H. Majid, "Eeg analysis for understanding stress based on affective model basis function," in *2011 IEEE 15th International Symposium on Consumer Electronics (ISCE)*. IEEE, 2011, pp. 592–597.

[69] J. Wijsman, B. Grundlehner, J. Penders, and H. Hermens, "Trapezius muscle emg as predictor of mental stress," in *Wireless Health 2010*. ACM, 2010, pp. 155–163.

[70] J. Healey, R. W. Picard *et al.*, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[71] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation, space, and environmental medicine*, vol. 76, no. 6, pp. B172–B182, 2005.

[72] V. P. Patil, K. K. Nayak, and M. Saxena, "Voice stress detection," *International Journal of Electrical, Electronics and Computer Engineering*, vol. 2, no. 2, pp. 148–154, 2013.

[73] B. Arnrich, C. Setz, R. La Marca, G. Tröster, and U. Ehlert, "What does your chair know about your stress level?" *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 207–214, 2010.

[74] C. Epp, M. Lippold, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 715–724.

[75] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *Journal of Biomedical Informatics*, p. 103139, 2019.

[76] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *2011 Computing in Cardiology*. IEEE, 2011, pp. 85–88.

[77] J. Bakker, M. Pechenizkiy, and N. Sidorova, "What's your current stress level? detection of stress patterns from gsr sensor data," in *2011 IEEE 11th international conference on data mining workshops*. IEEE, 2011, pp. 573–580.

[78] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 2013, pp. 209–214.

[79] M. S. Kalas and B. Momin, "Stress detection and reduction using eeg signals," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016, pp. 471–475.

[80] J. Hernandez, R. R. Morris, and R. W. Picard, "Call center stress recognition with person-specific models," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 125–134.

[81] A. Ghaderi, J. Frounchi, and A. Farnam, "Machine learning-based signal processing using physiological signals for stress detection," in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2015, pp. 93–98.

[82] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 2014, pp. 3–14.

[83] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cstress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 2015, pp. 493–504.

[84] Ofcom, "The communications market 2018," https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr-2018/, 2018, [Online; accessed Jan 13, 2019].

[85] J. Oechssler, A. Roider, and P. W. Schmitz, "Cognitive abilities and behavioral biases," *Journal of Economic Behavior & Organization*, vol. 72, no. 1, pp. 147–152, 2009.

[86] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.

[87] A. Muaremi, B. Arnrich, and G. Tröster, "Towards measuring stress with smartphones and wearable devices during workday and sleep," *BioNanoScience*, vol. 3, no. 2, pp. 172–183, 2013.

[88] F. Funke and U.-D. Reips, "Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales," *Field methods*, vol. 24, no. 3, pp. 310–327, 2012.

[89] R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, 1932.

[90] Google, "Activity recognition api," https://developers.google.com/location-context/activity-recognition/, 2018, [Online; accessed Mar 05, 2019].

[91] A. M. Boonstra, H. R. S. Preuper, G. A. Balk, and R. E. Stewart, "Cut-off points for mild, moderate, and severe pain on the visual analogue scale for pain in patients with chronic musculoskeletal pain," *Pain®*, vol. 155, no. 12, pp. 2545–2550, 2014.

[92] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[93] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, pp. 1189–1232, 2001.

[94] P. Geladi and B. R. Kowalski, “Partial least-squares regression: a tutorial,” *Analytica chimica acta*, vol. 185, pp. 1–17, 1986.

[95] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[96] M. Kuhn *et al.*, “Building predictive models in r using the caret package,” *Journal of statistical software*, vol. 28, no. 5, pp. 1–26, 2008.

[97] H. Oh, C. Rizo, M. Enkin, and A. Jadad, “What is ehealth (3): a systematic review of published definitions,” *Journal of medical Internet research*, vol. 7, no. 1, 2005.

[98] B. Martínez-Pérez, I. De La Torre-Díez, and M. López-Coronado, “Mobile health applications for the most prevalent conditions by the world health organization: review and analysis,” *Journal of medical Internet research*, vol. 15, no. 6, 2013.

[99] R. S. Weinstein, A. M. Lopez, B. A. Joseph, K. A. Erps, M. Holcomb, G. P. Barker, and E. A. Krupinski, “Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers,” *The American journal of medicine*, vol. 127, no. 3, pp. 183–187, 2014.

[100] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Patel, and A. Haines, “The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis,” *PLoS medicine*, vol. 10, no. 1, p. e1001363, 2013.

[101] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, “Enabling real-time context-aware collaboration through 5g and mobile edge computing,” in *Information Technology-New Generations (ITNG), 2015 12th International Conference on.* IEEE, 2015, pp. 601–605.

[102] H. Alemdar and C. Ersoy, “Wireless sensor networks for healthcare: A survey,” *Computer networks*, vol. 54, no. 15, pp. 2688–2710, 2010.

[103] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *computers & security*, vol. 28, no. 1-2, pp. 18–28, 2009.

[104] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection using hidden markov model," *IEEE Transactions on dependable and secure computing*, vol. 5, no. 1, pp. 37–48, 2008.

[105] F. Jiang, Y. Wu, and A. K. Katsaggelos, "Detecting contextual anomalies of crowd motion in surveillance video," in *Image Processing (ICIP), 2009 16th IEEE International Conference on.* IEEE, 2009, pp. 1117–1120.

[106] S. Lee, J. Kim, S. Shin, P. Porras, and V. Yegneswaran, "Athena: A framework for scalable anomaly detection in software-defined networks," in *Dependable Systems and Networks (DSN), 2017 47th Annual IEEE/IFIP International Conference on.* IEEE, 2017, pp. 249–260.

[107] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[108] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[109] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[110] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognition—a review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 865–878, 2012.

[111] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded scene analysis: A survey," *IEEE transactions on circuits and systems for video technology*, vol. 25, no. 3, pp. 367–386, 2015.

[112] S. M. Mahmoud, A. Lotfi, and C. Langensiepen, "Abnormal behaviours identification for an elder's life activities using dissimilarity measurements," in *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments.* ACM, 2011, p. 25.

[113] A. Lotfi, C. Langensiepen, S. M. Mahmoud, and M. J. Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour," *Journal of ambient intelligence and humanized computing*, vol. 3, no. 3, pp. 205–218, 2012.

[114] R. Baeza-Yates, D. Jiang, F. Silvestri, and B. Harrison, "Predicting the next app that you are going to use," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining.* ACM, 2015, pp. 285–294.

[115] Y. Xu, M. Lin, H. Lu, G. Cardone, N. Lane, Z. Chen, A. Campbell, and T. Choudhury, "Preference, context and communities: a multi-faceted approach to predicting smartphone app usage patterns," in *Proceedings of the 2013 International Symposium on Wearable Computers.* ACM, 2013, pp. 69–76.

[116] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on neural networks*, vol. 16, no. 3, pp. 645–678, 2005.

[117] J. Pansiot, D. Stoyanov, D. McIlwraith, B. P. Lo, and G.-Z. Yang, "Ambient and wearable sensor fusion for activity recognition in healthcare monitoring systems," in *4th international workshop on wearable and implantable body sensor networks (BSN 2007).* Springer, 2007, pp. 208–212.

[118] F. Kosmalla, F. Daiber, and A. Krüger, "Climbsense: Automatic climbing route recognition using wrist-worn inertia measurement units," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 2015, pp. 2033–2042.

[119] K. K. Biswas and S. K. Basu, “Gesture recognition using microsoft kinect®,” in *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on.* IEEE, 2011, pp. 100–103.

[120] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster, “Wearable activity tracking in car manufacturing,” *IEEE Pervasive Computing*, vol. 7, no. 2, 2008.

[121] J. C. Lee, “Hacking the nintendo wii remote,” *IEEE pervasive computing*, vol. 7, no. 3, 2008.

[122] K. Tanaka, J. Parker, G. Baradoy, D. Sheehan, J. R. Holash, and L. Katz, “A comparison of exergaming interfaces for use in rehabilitation programs and research,” *Loading...*, vol. 6, no. 9, 2012.

[123] J. Egger, M. Gall, J. Wallner, P. Boechat, A. Hann, X. Li, X. Chen, and D. Schmalstieg, “Htc vive mevislab integration via openvr for medical applications,” *PloS one*, vol. 12, no. 3, p. e0173972, 2017.

[124] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, “Sensor-based activity recognition,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[125] R. Bodor, B. Jackson, and N. Papanikolopoulos, “Vision-based human tracking and activity recognition,” in *Proc. of the 11th Mediterranean Conf. on Control and Automation*, vol. 1, 2003.

[126] X. Wang and Q. Ji, “A hierarchical context model for event recognition in surveillance video,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2561–2568.

[127] A. Baak, M. Müller, G. Bharaj, H.-P. Seidel, and C. Theobalt, “A data-driven approach for real-time full body pose reconstruction from a depth camera,” in *Consumer Depth Cameras for Computer Vision.* Springer, 2013, pp. 71–98.

[128] L. Chen, C. D. Nugent, and H. Wang, “A knowledge-driven approach to activity recognition in smart homes,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 961–974, 2012.

[129] T. Huynh, M. Fritz, and B. Schiele, “Discovery of activity patterns using topic models,” in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 2008, pp. 10–19.

[130] M. Längkvist, L. Karlsson, and A. Loutfi, “A review of unsupervised feature learning and deep learning for time-series modeling,” *Pattern Recognition Letters*, vol. 42, pp. 11–24, 2014.

[131] B. E. Ainsworth, W. L. Haskell, A. S. Leon, J. D. Jacobs, H. J. Montoye, J. F. Sallis, and J. R. Paffenbarger, “Compendium of physical activities: classification of energy costs of human physical activities.” *Medicine and science in sports and exercise*, vol. 25, no. 1, pp. 71–80, 1993.

[132] M. Kranz, A. MöLler, N. Hammerla, S. Diewald, T. PlöTz, P. Olivier, and L. Roalter, “The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices,” *Pervasive and Mobile Computing*, vol. 9, no. 2, pp. 203–215, 2013.

[133] A. Fernández, V. López, M. Galar, M. J. del Jesus, and F. Herrera, “Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches,” *Knowledge-Based Systems*, vol. 42, pp. 97–110, 2013.

[134] M. Csikszentmihalyi and R. Larson, “Validity and reliability of the experience-sampling method,” in *Flow and the foundations of positive psychology*. Springer, 2014, pp. 35–54.

[135] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones.” in *ESANN*, 2013.

[136] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566–4578, 2016.

[137] A. Cereatti, D. Trojaniello, and U. Della Croce, "Accurately measuring human movement using magneto-inertial sensors: techniques and challenges," in *Inertial Sensors and Systems (ISISS), 2015 IEEE International Symposium on.* IEEE, 2015, pp. 1–4.

[138] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "A survey of online activity recognition using mobile phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, 2015.

[139] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, "Soundsense: scalable sound sensing for people-centric applications on mobile phones," in *Proceedings of the 7th international conference on Mobile systems, applications, and services.* ACM, 2009, pp. 165–178.

[140] Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy, "Activity recognition with evolving data streams: A review," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, p. 71, 2018.

[141] G. Okeyo, L. Chen, H. Wang, and R. Sterritt, "Dynamic sensor data segmentation for real-time knowledge-driven activity recognition," *Pervasive and Mobile Computing*, vol. 10, pp. 155–172, 2014.

[142] A. Zinnen, C. Wojek, and B. Schiele, "Multi activity recognition based on bodymodel-derived primitives," in *International Symposium on Location-and Context-Awareness.* Springer, 2009, pp. 1–18.

[143] A. Dutta, O. Ma, M. P. Buman, and D. W. Bliss, "Learning approach for classification of geneactiv accelerometer data for unique activity identification," in *Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on.* IEEE, 2016, pp. 359–364.

[144] S. Pirttikangas, K. Fujinami, and T. Nakajima, "Feature selection and activity recognition from wearable sensors," in *International Symposium on Ubiquitious Computing Systems*. Springer, 2006, pp. 516–527.

[145] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[146] K. P. Murphy and S. Russell, "Dynamic bayesian networks: representation, inference and learning (phd thesis)," *University of California, Berkeley*, 2002.

[147] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *International Conference on Pervasive Computing*. Springer, 2004, pp. 1–17.

[148] T. Huynh, U. Blanke, and B. Schiele, "Scalable recognition of daily activities with wearable sensors," in *LoCA*, vol. 7, 2007, pp. 50–67.

[149] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.

[150] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 2, pp. 281–286, 2002.

[151] J. A. Ward, P. Lukowicz, and H. W. Gellersen, "Performance metrics for activity recognition," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 1, p. 6, 2011.

[152] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[153] C. V. Bouten, K. T. 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 Transactions on Biomedical Engineering*, vol. 44, no. 3, pp. 136–147, 1997.

[154] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression.* John Wiley & Sons, 2013, vol. 398.

[155] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[156] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22. IBM New York, 2001, pp. 41–46.

[157] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[158] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[159] A. M. Moseley, R. D. Herbert, C. Sherrington, and C. G. Maher, "Evidence for physiotherapy practice: a survey of the physiotherapy evidence database (pedro)," *Australian Journal of Physiotherapy*, vol. 48, no. 1, pp. 43–49, 2002.

[160] C. G. Maher, C. Sherrington, R. D. Herbert, A. M. Moseley, and M. Elkins, "Reliability of the pedro scale for rating quality of randomized controlled trials," *Physical therapy*, vol. 83, no. 8, pp. 713–721, 2003.

[161] W. Pirker and R. Katzenschlager, "Gait disorders in adults and the elderly," *Wiener Klinische Wochenschrift*, vol. 129, no. 3-4, pp. 81–95, 2017.

[162] C. T. Valadao, F. Loterio, V. Cardoso, T. Bastos, A. Frizera-Neto, and R. Carelli, "Robotics as a tool for physiotherapy and rehabilitation sessions," *IFAC-PapersOnLine*, vol. 48, no. 19, pp. 148–153, 2015.

[163] M. Belas dos Santos, C. Barros de Oliveira, A. dos Santos, C. Garabello Pires, V. Dylewski, and R. M. Arida, "A comparative study of conventional physiotherapy versus robot-assisted gait training associated to physiotherapy in individuals with ataxia after stroke," *Behavioural neurology*, vol. 2018, 2018.

[164] Y. Bai, G. J. Welk, Y. H. Nam, J. A. Lee, J.-M. Lee, Y. Kim, N. F. Meier, and P. M. Dixon, "Comparison of consumer and research monitors under semistructured settings." *Medicine and science in sports and exercise*, vol. 48, no. 1, pp. 151–158, 2016.

[165] G. Appelboom, E. Camacho, M. E. Abraham, S. S. Bruce, E. L. Dumont, B. E. Zacharia, R. D'Amico, J. Slomian, J. Y. Reginster, O. Bruyère *et al.*, "Smart wearable body sensors for patient self-assessment and monitoring," *Archives of Public Health*, vol. 72, no. 1, p. 28, 2014.

[166] M. Yang, H. Zheng, H. Wang, S. McClean, and D. Newell, "igait: an interactive accelerometer based gait analysis system," *Computer methods and programs in biomedicine*, vol. 108, no. 2, pp. 715–723, 2012.

[167] H. K. Chan, H. Zheng, H. Wang, R. Gawley, M. Yang, and R. Sterritt, "Feasibility study on iphone accelerometer for gait detection," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*. IEEE, 2011, pp. 184–187.

[168] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE, 2015, pp. 1–6.

[169] A. Nicholson, M. Lowe, J. Parker, S. Lewis, P. Alderson, and A. Smith, "Systematic review and meta-analysis of enhanced recovery programmes in surgical patients," *BJS*, vol. 101, no. 3, pp. 172–188, 2014.

[170] D. J. Cook, J. E. Thompson, S. K. Prinsen, J. A. Dearani, and C. Deschamps, "Functional recovery in the elderly after major surgery: assessment of mobility

recovery using wireless technology," *The Annals of thoracic surgery*, vol. 96, no. 3, pp. 1057–1061, 2013.

[171] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.

[172] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017. [Online]. Available: http://jmlr.org/papers/v18/16-365

[173] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[174] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[175] Research and Markets, "Internet of things (iot) in healthcare market size, share & trends analysis report by end use, by component, by connectivity technology (cellular, wi-fi, satellite), by application, and segment forecasts, 2018 - 2025," Tech. Rep., nov 2018.

[176] L. Batistić and M. Tomic, "Overview of indoor positioning system technologies," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018, pp. 0473–0478.

[177] The European GNSS Agency, "The results are in: Galileo increases the accuracy of location based services," https://www.gsa.europa.eu/newsroom/news/results-are-galileo-increases-accuracy-location-based-services, 2014, [Online; accessed Mar 05, 2019].

[178] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen, and M. Valkama, "High-efficiency device positioning and location-aware communications in dense 5g networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 188–195, 2017.

[179] R. Sioni and L. Chittaro, "Stress detection using physiological sensors," *Computer*, vol. 48, no. 10, pp. 26–33, 2015.

List of Figures

List of Tables

Abbreviations

AAL	Active and Assisted Living
ADL	Activities of Daily Living
AR	Activity Recognition
BDS	BeiDou Navigation Satellite System
BLE	Bluetooth Low Energy
CTI	Swiss Commission for Technology and Innovation
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DT	Decision Tree
ECG	Electrocardiogram
EDA	Electrodermal Activity
EDLAH2	Enhanced Daily Living and Health 2
EEG	Electroencephalogram
EMG	Electromyogram
ET	Extra Trees
EU	European Union
FNS	Swiss National Science Foundation
FP	False Positive
GBM	Gradient Boosting Machines
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability

IAPS International Affective Picture System
IMU Inertial Measurement Unit
IoT Internet of Things
IPS Indoor Positioning System
IQR Interquartile Range

kNN k-Nearest Neighbors

LR Logistic Regression

NB Naive Bayes

PLS Partial Least Squares
PPV Positive Predictive Value
PR Precision-Recall

ReLU Rectified Linear Unit
RF Random Forest
ROC Receiver Operating Characteristic
RSSI Received Signal Strength Indication

SNR Signal-to-Noise Ratio
SVM Support Vector Machines

TOA Time Of Arrival
TP True Positive

UK United Kingdom
UN United Nations
UWB Ultra-wideband

VR Virtual Reality

WLAN Wireless Local Area Network