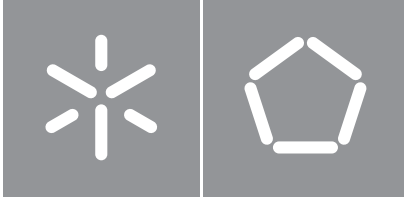
Universidade do Minho
Escola de Engenharia

Luís Miguel Marques Martins

Fall Risk Assessment: Towards Real-Time and Continuous Estimation



Universidade do Minho
Escola de Engenharia

Luís Miguel Marques Martins

Fall Risk Assessment: Towards Real-Time and Continuous Estimation

Master's Dissertation
Master's Degree in Biomedical Engineering
Medical Electronics

Work supervised by:
Professora Doutora Cristina P. Santos

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Acknowledgements

All the work done over the last year would not have been possible without the assistance, understanding, and support of many people who deserve to be here recognized.

First and foremost, I want to express my gratitude to my supervisor, Prof. Cristina Santos, who created a wonderful learning environment for me, providing me with several possibilities to broaden my horizons through significant research contacts. I am thankful to her for her support, and motivation.

I'm also grateful to Nuno, who was always there, willing to help and discuss the study, as well as make recommendations that aided in the development of this achievement. Thank you also for your encouragement and understanding at difficult times, and for always keeping me on track to achieve my best and beyond.

I'd also like to express my gratitude to everyone in the BiRDLab for the pleasant working environment they've built, specially to Raquel, Cristiana and Rafael for helping me with the waistband. Thank you as well to all of my fellow classmates and friends I've made. I also can't forget to thank all the unicorns and other mythological beings who, without a doubt, had a significant influence on me in order for me to complete this stage in the most pleasurable way possible, listening to all my prayers and concerns. I hope I left a positive impression on you, because you surely did that on me.

Maya, my girlfriend, deserves a heartfelt thank you for her unwavering support, devotion, and patience throughout the past year. She both inspired me to finish my dissertation while provided much-needed distractions. Even though she was far away, she accompanied me on this journey and is a big part of this success as well.

Finally, but certainly not least, I want to express my gratitude to my family, particularly my mother, father, and brother, for all their love and support, not only during my dissertation but throughout my entire life. Without them, none of this would be possible.

I'm grateful to all of you. With love and the utmost appreciation,

Luís Martins

Abstract

Fall Risk Assessment: Towards Real-Time and Continuous Estimation

Falls are a common cause of injuries and deaths in the senior population. Due to population ageing, more falls occur. As a result, the expense of treating fall-related injuries is rising as well. Thus, the development of strategies with the capability of real-time fall risk monitoring without user restriction is imperative. Due to their advantages, daily life items, such as waist straps, smart watches and smartphones, can be a solution to integrate fall-risk assessment systems. Summing up, it is crucial the development of strategies that can in real time evaluate and recognize risky situations based on several fall risk factors, in order to prevent falls, also allowing timely assistance to reduce its harmful consequences. This thesis aims to idealize a tool capable of assessing the fall risk of a subject in real time, taking into account several existing fall risk factors, focusing on continuous fall risk assessment (FRA) and recognition of daily activities performed by the elderly.

The strategy regarding the FRA consisted of a three-part modular architecture that allows real-time fall risk assessment using only a waist strap fitted with an inertial sensor and an eHealth platform. Each module must examine risk factors such as biological data, environmental changes, and even geriatric behavioral matters such as daily activities. The activity recognition module was validated in this dissertation.

For the recognition of 20 different daily activities and falls, a machine learning study was completed involving 10 feature selection methods and 5 different machine learning classifiers. Results have shown that the Ensemble Learning with the 65 first features ranked by PCA presented the best results with 96.53% of overall accuracy, while maintaining a low classification time, leaving the possibility of its usage in real-time scenarios open in the future. The same problem was also addressed via Deep Learning algorithms and 4 basic Neural Networks architectures were tested. Despite its outcomes not being as good as in the prior procedure, their potential was demonstrated (overall accuracy of 92.55% for BiLSTM), indicating that they could be a good option in the future with some changes and new testing.

Keywords: Fall, Fall Risk Assessment, Activities of the Daily Living, Machine Learning, Deep Learning.

Resumo

Análise de risco de queda: em direção a uma estimativa em tempo real

Quedas são uma das causa mais comuns de lesões na população idosa. Como resultado, custos com o tratamento destas têm também aumentado. Assim, é imperativo o desenvolvimento de ferramentas de Análise do Risco de Queda (ARQ) em tempo real, sem restrições para os utilizadores. Dispositivos como cintas e smartphones, podem hoje em dia ser uma solução para integrar sistemas de avaliação de risco de queda. Em suma, é fundamental o desenvolvimento de estratégias que possam avaliar e reconhecer situações de risco de queda com base em vários fatores, a fim de as prevenir, permitindo ainda uma assistência atempada por forma a reduzir as suas consequências. Assim, esta dissertação pretende dar os primeiros passos em direção a uma ferramenta de ARQ, tendo em conta diversos fatores de risco, com foco principal na ARQ e reconhecimento de atividades diárias realizadas por idosos.

A estratégia de ARQ consistiu em uma arquitetura modular de três partes para a avaliação do risco de queda em tempo real, usando uma cinta equipada com um sensor inercial e uma plataforma de *eHealth*. Cada módulo examina fatores de risco, como dados biológicos, mudanças ambientais e até mesmo questões comportamentais dos idosos, como as suas atividades diárias. O módulo de reconhecimento de atividades foi validado nesta dissertação.

Para o reconhecimento de 20 diferentes atividades diárias e quedas, foi realizado um estudo de Machine Learning envolvendo 10 métodos de seleção de features e 5 classificadores diferentes. O Ensemble Learning com as primeiras 65 features selecionadas pelo PCA apresentou os resultados mais promissores, com 96,53% de accuracy, mantendo um tempo de classificação baixo, deixando em aberto a possibilidade de seu uso em tempo real no futuro. O mesmo problema foi ainda solucionado através de algoritmos de Deep Learning, e 4 arquiteturas básicas de Redes Neurais foram testadas. Apesar de seus resultados não serem tão bons quanto no procedimento anterior (accuracy de 92,55% para BiLSTM), seu potencial foi demonstrado, indicando que podem ser uma solução de futuro, após modificações e novos testes.

Palavras-chave: Queda, Análise de risco de queda, Atividades Diárias, Machine Learning, Deep Learning.

Contents

List of Figures

List of Tables

Acronyms

ADL Activities of Daily Living

BBS Berg Balance Scale

CNN Convolutional Neural Networks

DL Deep Learning

DTs Decision Trees

FRA Fall Risk Assessment

KNN K-Nearest Neighbours

LDA Linear Discriminant Analysis

LSTM Long Short-Term Memory

ML Machine Learning

NN Neural Networks

PCA Principal Component Analysis

POMA Performance Oriented Mobility Assessment

QDA Quadratic Discriminant Analysis

TUG Timed Up and Go

Chapter 1

Introduction

This dissertation presents the research carried out in the scope of my Masters Degree in Biomedical Engineering, branch of Medical Electronics, at the Biomedical Robotic Devices Lab (BirdLab) included in the Center for Micro-Electro-Mechanical Systems (CMEMS), a research center of the Department of Industrial Electronics (DEI) in the University of Minho.

The focus of this dissertation centers around the idealization of an eHealth platform-based Fall Risk Assessment (FRA) strategy which, in a continuous and multifactorial manner, performs the FRA of an individual, through the provision of several types of data. This data types may be: real-time accelerometry and angular velocity data collected by an inertial sensor located in his lumbar region; clinical information such as age, weight, medications or diseases collected in medical appointments; and data regarding the surrounding ambient, such as temperature, humidity level or the brightness of the place where a subject is. This strategy aims to help seniors improve their quality of life (QoL), ensuring a feedback process that communicates with caregivers or health professionals in the event of a fall to provide timely support.

The project was divided into 3 main phases, namely: **i)** research and design of a modular strategy for multifactorial real-time FRA; and **ii)** research, development and validation of AI-based, Machine Learning (ML) and Deep Learning (DL) models, for the recognition of 16 Activities of the Daily Living (ADL) and 4 types of fall events, using information gathered from several public datasets; and **iii)** Enhacement of an adjustable waistband and delineation of data acquisition protocols of relevant ADLs in nursing home environments.

1.1 Motivation

People all over the world can now expect to live longer lives, with the majority living into their sixties and beyond. According to a *World Health Organization* (WHO) study, every country in the world is experiencing an increase in the size and proportion of older people within the population. By 2030, one in every six people on the planet will will have 60 years old or more, with the number of people aged 80 or older expected to triple between 2020 and 2050, reaching 426 million [1]. The aging process is accompanied by the rise of a number of health conditions known as geriatric syndromes, which are often the result of a number of underlying problems, such as weakness and falls [1].

Falls are among the most serious clinical issues that the elderly confront, and they are considered a public health concern, causing substantial rates of mortality, morbidity, immobility and reduced QoL, as well as their impact on the health system services [2–4]. It is estimated that at least one third of the population over the age of 65 years fall at least once per year and this proportion tends to increase strongly with age [5]. In 2006, falls constituted two-thirds of unintentional injuries, which was the fifth leading cause of death in older adults (after cardiovascular disease, cancer, stroke and pulmonary disorders) [4].

The high incidence together with a high susceptibility to injury of the elderly, make falls in this age group more dangerous than in others, where even a relatively mild fall can be particularly dangerous, because of a high prevalence of clinical diseases and age-related physiological changes in the older adults [4]. Fractures, joint dislocations, and head trauma are some of the serious injuries that can be caused by a fall. Moreover, psychological trauma such as fear of falling is another consequence of falls which may lead to self-imposed restrictions in activity and, consequently, loss of independence [6, 7]. Despite not being one of the European countries with the highest number of deaths caused by falls, according to WHO studies from 2014, there were 618 deaths in Portugal as a result of unintentional falls. This number has been increasing, with 815 deaths recorded in 2018 [8].

As a consequence of the current problem of falls among the elderly community, and its propensity to cause severe harm, the health care costs associated with falls from the elderly and fall-related injuries are high, as well as its social burden [2, 9–11]. According to *WHO*, for the elderly, the average health system cost per fall injury in Australia can be higher than US$ 1000 [2, 11]. In Finland, the average health-care expenditure for a single fall injury episode for adults 65 and older was US$ 3611 [12, 13]. Furthermore, the average cost of hospitalization for a fall-related injury ranges from US$ 6646 in Ireland to US$ 17 483 in the United States for those 65 and older [14, 15]. This expenditure is expected to rise to US$ 240 billion by 2040 [16].

Summarising, with an increasingly aging population, and the risks that it entails in terms of the probability of falls, as well as its consequences, research in fall prevention methods and FRA becomes imperative. The application of an adequate FRA method can thus prevent a fall and avoid its harmful outcomes, both those which affect the population at a physical and psychological level, as well as the negative social and

economic impacts [2, 10, 11, 17].

1.2 Problem Statement and Scope

Until today, the FRA in the elderly has been performed through occasional medical consultations, and it depends mostly on analysis of fall risk specific and functional mobility assessment tools, in the form of questionnaires, physical tests, gait analysis, and physical activity measurements, such as the Timed Up and Go (TUG) [18], Berg Balance Scale (BBS) [19], Tinetti Performance Oriented Mobility Assessment (POMA) [20], or the 5 times sit-to-stand (STS) [21], among others [10, 17]. ADLs are not addressed in these approaches, which solely evaluate motions that frequently do not correlate to those conducted in an aged person's daily life, making this sort of evaluation particularly prone to error. Furthermore, these occasional consultations may also occur on periods when a subject may be unprepared or overly prepared for the assessment of his fall risk.

Despite being the most used today, the referred tests can be influenced by other several factors, such as the subjectivity of the evaluation, the ambient conditions where the tests are carried, or even the physical and psychological state of the individuals when performing the tests [22]. Moreover, these types of testing can be time consuming [23] and require a team of professionals for their preparation and execution, which may become cumbersome for the elderly [24]. In addition, the results are usually not the most understandable by the persons being tested [25]. Clinical FRA tools have also not taken into account hazards found in the normal daily living environment of the elderly, which may also explain why the actual clinical tests fail at giving the fall risk a score with satisfying results [22, 26, 27].

In recent years, with advances in sensing with wearable technology, as well as the emergence of Artificial Intelligence (AI)-based algorithms, studies have attempted to make FRA less subjective and prone to error, by investigating the potential use of instrumented FRA tools based on features extracted from inertial sensors (i.e. accelerometers and gyroscopes) attached to the subject's body.

FRA based on wearable sensors can be performed in a prospective way, i.e., the risk of a subject to fall in a long-term future (usually 6 to 12 months), in which sensor data is used to predict subject's long-term fall risk based on clinical scale scores or to automatically distinguish between Fallers (F) and Non-Fallers (NF) [17, 22, 28]. Wearable sensor-based FRA can also be performed from a short-term or even immediate approach, where the acquired data is used to detect pre-fall/unbalanced situations in real-time and consequently identify fall risk events [29, 30], however, there is still a very limited number of works which perform this type of FRA [22, 31]. The first approach is significant since it assesses a subject's overall physical and psychic abilities and places these assessments in a fall risk context to better understand an aged person's behavior in future fall risk scenarios. The second approach, is also very crucial because of its ability to analyze in real-time risk situations, which can be triggered by slips or trips

caused by environmental hazards, or even the inherent fall risk that a given activity presents to the elderly, such as walking up or downstairs.

In order to design effective interventions for wearable-based fall prevention and FRA, it is crucial to identify the optimal combination of three factors: **i)** where to place the sensor(s); **ii)** which task(s) to be performed; and **iii)** which features should be extracted and analyzed [17, 27].

Also, falls happen at specific times and contexts [32], so, in order to prevent falls from happening, when assessing fall risk, the current context such as time of day, current health status, location, and other relevant context information needs to be included in the FRA as well [22, 27]. Likewise, the analysis of ADL reveals an extreme importance for the FRA, whether it is performed in an immediate or prospective manner. According to Lord et al. [33], the number of falls in the elderly is also linked to the activity that was being performed immediately before the fall, such as walking, climbing up or downstairs, or getting up from bed. As a result, real-time ADL identification will provide information regarding the performance of certain activities that may raise the risk of falling in a short-term/immediate perspective. This ADL recognition also allow for researchers to assess activity performance, see how it changes over larger periods of time, and see how gait characteristics and other health-related aspects get better or worse, allowing for correlations between the elderly's everyday activities and fluctuations in their risk of falling in the long term.

In light of the foregoing, it should be highlighted that there is still a long way to go regarding improvements in the way the FRA is conducted in the elderly. It is apparent that nowadays both prospective and immediate perspectives of FRA should be assessed, achieving the most accurate and least subjective assessment possible. Furthermore, for this purpose, it is crucial for FRA to have a context-aware property influenced by multiple fall risk factors, such as environment hazards that surround the elderly, as well as their physiological and psychological state. Finally, due importance must be given to the ADL that the subjects perform, as well as to the way they carry out said activities, studying characteristics such as the speed and the balance presented in order to be used in both prospective and immediate FRA approaches. The combination of this several factors is a problematic that has not been adequately addressed by existing wearable-based FRA strategies [22].

1.3 Objectives and Research Questions

From the scope shown in Section 1.2, in order to surpass the current drawbacks of FRA strategies, the present work aims to, together with the strong technological evolution of AI and wearable sensor technology, design a multifactorial tool capable of continuously assess the fall risk of elderly subjects, taking into account both prospective an immediate FRA considerations, and focusing on the recognition of daily activities performed by the elderly population, thus avoiding possible falls and its consequences.

In order to achieve this final objective, other objectives will have to be achieved, among which:

- **Goal 1**: Gather knowledge on methods used currently to fall risk assessment, by performing a state-of-the-art review on different methods. This will be resumed on Section 2.2.
- **Goal 2**: Gather knowledge on methods used currently for ADL recognition, by performing a literature review, focusing on Machine Learning and Neural Networks based approaches, and list the ADL that have to be considered for fall risk assessment. This will be resumed on Chapter 3.
- **Goal 3**: Design a custom strategy architecture for a multi-factorial fall risk assessment tool, to attain a fall risk from diverse types of subject (inertial, biological, psychological, etc) and ambient (temperature, humidity, luminosity, etc) data. This strategy will be described on chapter 4.
- **Goal 4**: Create a custom and dataset with inertial and demographic data by joining several team-owned and public datasets with data acquired from multiple healthy and unhealthy subjects in real and simulated domiciliary settings, while performing several ADLs using waist-located intertial sensors, in order to train and evaluate different ADL recognition algorithms offline - Chapter 5.
- **Goal 5**: Exploit Machine Learning and Neural Network models capable of recognising several ADLs with a big dataset made from diverse public datasets. The explored approaches should lead to, at least, 95% of accuracy for recognition of 20 classes of ADL, since literature analyzed shows a detecting a maximum of 14 ADL (in this case, with 80% accuracy) [34] for ML-based approaches and 96.7% accuracy for the recognition of 6 ADL with DL-based approaches. These steps will be described on chapter 5. Chapter 6 shows the obtained results.
- **Goal 6**: Delineate protocols with clinicians and caregivers to obtain inertial and demographic data, as well as data from clinical FRA tests, from multiple elderly subjects in multiple real homecare settings and domiciliary, while performing several ADLs using the waistband (Section 4.8). These will be used to train and evaluate the best framework in the medium term.
- **Goal 7**: Run experiments with the identified elders and start creating a custom dataset with multi-modal sensory information, including inertial data from ADL, as well as physiological data, that will be used to evaluate and train the best framework in the medium term (Section 4.8).

Consequently, in this dissertation the following Research Questions are expected to be answered:

- **RQ 1**: What is the best strategy to perform FRA in real-time and in a multi-factorial way?
- **RQ 2**: What is the best Machine Learning and Deep Learning-based strategy to implement for real-time ADL and fall events recognition?
- **RQ 3**: What are the most suitable features for recognizing ADL and fall events from Machine Learning models?

- **RQ 4**: Which inclusion and exclusion criteria, ADL and circuits should be included in a protocol to collect relevant, variable data for training Machine Learning models to recognize ADL and fall events from elderly in real-life environments?

1.4 Solution Requirements

To be effectively used as a daily FRA tool on the on community dwelling people and elderly in nurse homes, this FRA solution has to:

- Function properly using light and small-sized technological components, in order to be comfortable and able to be used for several days by the target audience;
- Obtain the fall risk in real-time and in a multi-factorial way, taking into account fall risk factors of different natures;
- Display the fall risk clearly to the target audience (elderly and health professionals) and activate feedback measures in scenarios whenever there is a high risk of falling.

In order to introduce ADL and fall events recognition into the FRA strategy, this recognition needs to:

- Detect a wide set of daily activities and fall events from the smallest possible number of sensors. With the analyzed state-of-the-art detecting a maximum of 14 ADL [34], this dissertation aims for the recognition of 20 different ADL from the data of a single waist-located inertial sensor;
- Be well validated using a large dataset and multiple validation processes, such as Hold-Out (HO) and k-folds Cross-Validation (CV), in line with what is most commonly used in the literature;
- Present the highest possible accuracy, sensitivity, specificity, precision, F1 Score and MCC, with the literature showing performances of 96.70%, 96.83%, 96.83%, for accuracy, precision, and sensitivity, respectively, in 6 class classification problems using deep learning algorithms [35];
- Run in real-time. Since in this dissertation ADL recognition will be performed offline, it is important that the selected model classifies in a short time considering the window size and overlap used, i.e., classification time should be lower than the time between to consecutive classification windows.

1.5 Contributions to Knowledge

The main contributions of this dissertation to the actual knowledge in this field are:

- The proposal of a FRA tool architecture capable of using several fall risk factors to compute the risk of falling in real-time.
- A dataset creation from the fusion and normalization of diverse public and non-public datasets, which allows a more robust and validated training and testing of Machine and Deep Learning models.
- Machine Learning classification models and Neural Networks tool which are able to recognise 16 Daily activities and postural events, as well as 4 types of falls.
- The enhacement of a team-owned instrumented waistband for lower trunk inertial data acquisition.
- Development of user-friendly smartphone applications, for automatic Bluetooth data acquisition and data labeling.
- Establish contacts and data acquisition protocols in nurse homes using the developed software and waistband.

1.6 Thesis Outline

The remainder of this dissertation is organized in 6 chapters:

Chapter 2 contains an introduction to the problem of falls, with its definition of falls as well as some epidemiological knowledge (Section 2.1). This is followed by a review of the literature regarding different FRA techniques, such as Prospective FRA, Immediate FRA and Multi-factorial FRA (Section 2.2). Finally, in Section 2.3, conclusions are drawn about the research carried out.

Chapter 3 shows the most recent Machine Learning and Neural Networks trends for the recognition of daily activities, followed by general conclusions regarding the results and limitations of what is currently done in the literature.

In Chapter 4, proposals for solutions to the drawbacks encountered in the literature are presented, idealizing the architecture of a multifactorial FRA strategy. Moreover, This chapter also shows the enhancement process of a team-own instrumented waistband for inertial data collection and the respective data acquisition protocols applied in nursing homes.

The Chapter 5 demonstrates the process of developing and evaluating ML and DL-based models for ADL and fall recognition, as well as selecting the classifier that best fits the ADL recognition topic and determining the most relevant set of features for this task. Furthermore, studies regarding window size

and classification time will be addressed in order to establish potential use of these models in real-time applications.

Chapter 6 presents multiple results obtained with the proposed methods for validation, evaluation and best model choosing for ADL and fall events recognition, followed by the results regarding how changes in window size can influence the final performance of the selected models and it's classification time. These will be discussed, and it's limitations analyzed, with connections to the outcomes found in the current literature.

Chapter 7 concludes the dissertation, offering a short analysis of the work developed, along with directions for future research and improvement.

Chapter 2

Falls and Fall Risk Assessment: A State-of-the-Art

2.1 Falls: Definition and Epidemiological Studies

A fall is usually defined as an event which results in a person coming to rest inadvertently on the ground, floor or other lower level. However, it is important to note that there is no universal consensus on the definition of a fall [11, 32]. As stated in Chapter 1, falls have become a public health problem over the years among the senior community. Various studies concerning this topic have been carried out over time, ranging from retrospective to prospective analysis, which allowed some epidemiological conclusions to be drawn regarding the likely causes and conditions for the occurrence and prevalence of falls.

Some of the most significant research show that the prevalence and consequences of falls are strongly linked to the individuals' gender, age, frailty degree, and even other medical disorders. Similar results were reported in other studies, in which around 30% of the elderly will experience at least one fall per year. This value tends to increase with age [36, 37]. It has also been shown that subjects with greater difficulty in performing daily activities or with physical conditions that affect posture and balance have a greater propensity to fall in the future [38–40]. The same conclusion was drawn regarding cognitive impairments - such as Alzheimer's or Parkinson's diseases [32].

According to Lord et al. [32], the risk of falling can also affected by where it occurs and the activity being performed at the time of the fall, as shown in Figure 1. It should be noted, however, that the relationship with the activity discovered in the study does not entirely imply an inherent risk of the activity, but rather the combination of this with other factors such as gender, age, and fragility of the subjects, as well as the frequency with which activities are carried out or the environment in which they occur (Figure 2).

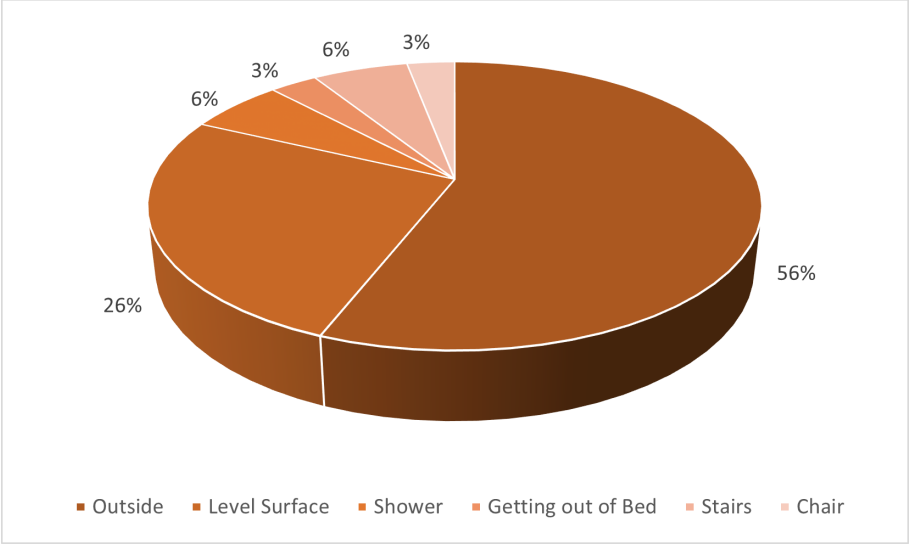


Figure 1: Location of falls and activities which led to it in community-dwelling women. 56% of falls occur outside the house, with the remainder (44%) occurring at various locations inside. Adapted from: [41].

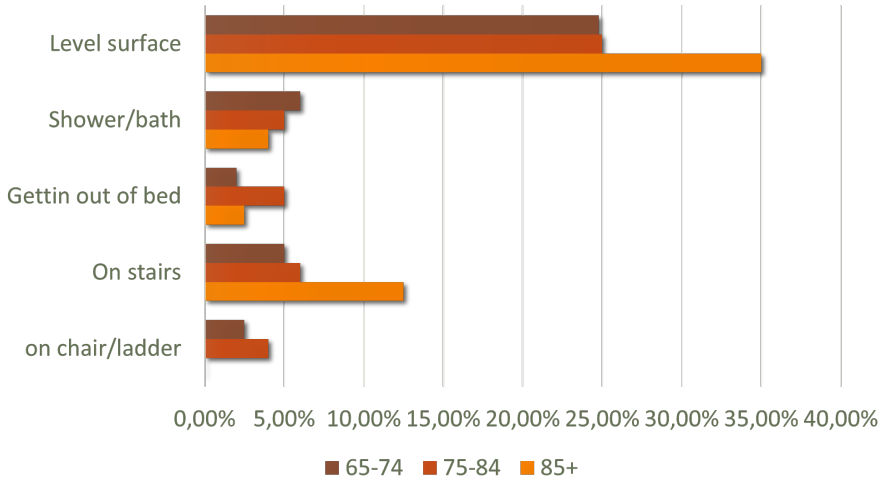


Figure 2: Indoor falls location according to age in community-dwelling women. Adapted from: [42].

2.1.1 Risk Factors

As demonstrated in the previous section, falls occur as a result of a complex interaction of risk factors. These show the existence of a great diversity of determinants that affect the well-being of the subjects and that, directly or indirectly, can cause falls [11]. According to [11], those can be categorized into four dimensions: biological, behavioural, environmental and socioeconomic factors. Figure 3 encapsulates the risk factors and the interactions of them on falls. The higher the exposure to these risk factors by the elderly, the greater becomes the fall risk for these people.

Fall risk factors can be: **i) Biological risk factors** - these factors embrace characteristics of the individuals that concern the human body. For instance, age, gender and race are non-modifiable biological

factors. Changes due to ageing such as the decline of physical and cognitive capacities are also recognised as Biological risk factors; **ii) Behavioural risk factors** - behavioural risk factors include those related human actions and daily choices. For example, the intake of multiple medications, excess alcohol use, and a sedentary lifestyle are considered risky behaviours which can unravel a fall. However, these are potentially modifiable; **iii) Environmental risk factors** - these factors encapsulate the interplay of individuals' physical conditions and the surrounding environment. Some environment hazards that pose a greater risk for injurious falls include uneven floors, narrow steps, slippery surfaces, or insufficient lighting; and **iv) Socioeconomic risk factors** - fall risk factors related to influence social conditions and economic status have on the elderly lifestyle. These factors include: inadequate housing, lack of social interaction, limited access to health and social care and lack of community resources, and are associated with a higher risk of falls.

Figure 3: Risk factor model for falls in older age. Adapted from [11].

2.1.2 Consequences

As indicated in Chapter 1, and from [33, 43, 44], falls can result in several devastating outcomes, either for the person who suffered the fall (physically, psychologically and socioeconomically), or for the society, as extra socioeconomic burdens for families and even overloading hospital emergencies.

Fractures, fear of falling, hospital admissions, and reduced quality of life are some of the main physical and psychological consequences. Falls can also result in dependence, loss of autonomy, confusion, immobilization, depression, and restrictions in daily activities post fall [43, 44]. Often, falls are considered

a reason for admission to a nursing home [33]. The diagram in Figure 4 illustrates some of the most common effects of falls nowadays.

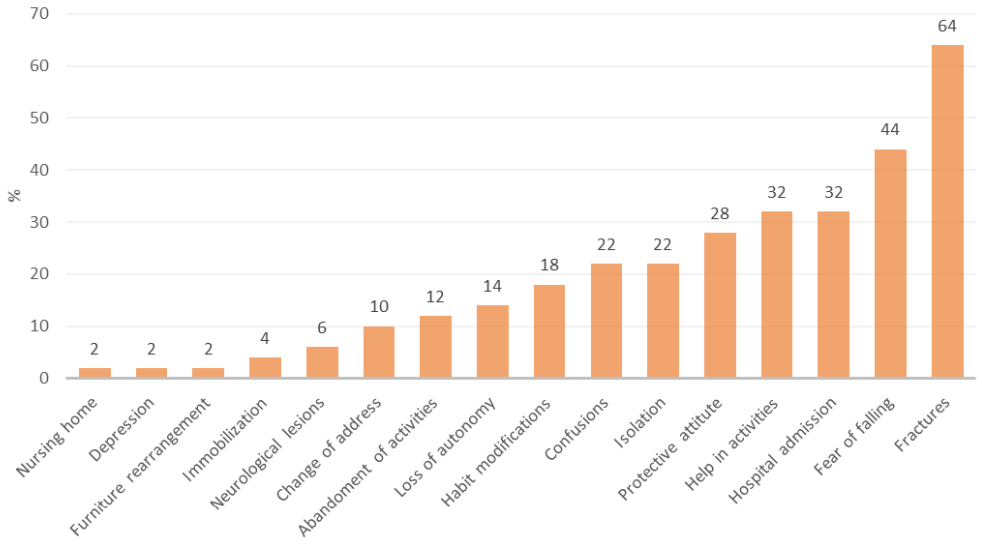


Figure 4: Consequences presented by older adults after falls (Adapted from [44]).

The economic impact of falls is critical to family, community, and society. Healthcare impacts and costs of falls in older age are significant and can be categorized into two aspects: Direct and indirect costs. The first encompass health care costs such as medications and adequate services. Indirect costs are related to societal productivity losses of activities in which family or other care givers would have been involved if no fall was sustained [11]. As an example, in Finland and Australia, the average health-care expenditure for a single fall injury episode for adults 65 and older was US$ 3611 and US$ 1049, respectively [12, 13]. Furthermore, the average cost of hospitalization for a fall-related injury ranges from US$ 6646 in Ireland to US$ 17 483 in the United States for those 65 and older [14, 15]. This expenditure is expected to rise to US$ 240 billion by 2040 [16].

2.1.3 Fall Prevention

Protective factors for falls in older age are related to avoiding or eliminating risk factors. The factors that are most easily modified in order to reduce the risk of falling have to do with behavioral changes and environmental modifications. Behavioral change to healthy lifestyle is a key ingredient to encourage healthy aging and avoid falls. Non-smoking, moderate alcohol consumption, maintaining weight within normal range in mid to older age can protect older people from falling. One example of the environmental modifications is home modification, preventing older persons from hidden fall hazards in daily activities at home. Installation of stairway protective devices and slip-resistant surfacing in the bathroom floor are some

modifications which can help in avoiding falls [2]. Another way to prevent falls in the elderly, as indicated in Section 1.1, is to carry out a constant analysis of the subjects' risk of falling, in order to understand its evolution and be able to act in a timely manner in order to reduce it. Thus, and taking into account the objectives outlined in Section 1.3, in Section 2.2 an analysis of the literature will be carried out within the scope of recent FRA strategies.

2.2 Fall Risk Assessment: A State-of-the-Art on Automatic Approaches

A literature review is presented in the following Section, seeking to find current trends in relation to the various steps performed when designing and testing FRA strategies, and the current main limitations. The research work that was done, as well as the outcomes, will be discussed in the following subsections. First and foremost, the search strategy will be discussed. Following that, the analysis of the chosen articles will be revealed. Finally, conclusions will be taken on the current state of the art and approaches for FRA in elderly people, as well as their limitations and scope for future enhancements on the subject.

2.2.1 Research Strategy and Eligibility Criteria

The literature search was divided into two phases, with the main goal of answering the following questions: **i)** Which are the main FRA methods using wearable sensors in literature?; **ii)** What types, number, and location of wearable sensors are most adopted?; **iii)** Which tasks are being performed in experimental protocols for data acquisition?; and **iv)** Which algorithms are being implemented for FRA?

In the first phase, several review articles ([10, 17, 27, 28, 45]) were investigated for a better contextualization and understanding of the stages that are generally evaluated during State of the Art research. The second stage involved searching for papers related to the issue in several online databases using keywords entered into search engines. Thus, several well-known keyword combinations related to the FRA theme in elderly people were initially investigated in renowned databases, such as Scopus, Web of Science, IEEEXplore and PubMed, in order to find a. The distribution of keywords by the various databases is shown in Table 1.

Table 1: Databases and respective keywords used for article research

Data Base	Key-words
Scopus	Aged OR Elderly OR Geriatric OR Old AND Wearable Sensor OR Wearable Device AND Fall Risk AND Gait OR Posture OR Walking
WoS	
PubBed	
IEEE	Aged OR Elderly OR Geriatric OR Old AND Fall Risk AND Wearable Sensor

After a preliminary search of these combinations, a total of 574 articles were found, as shown in Figure 5. Then, a screening and selection process was carried out until reaching the articles that will be analyzed in this Section. Before the screening process, a search for duplicate articles was made, after which 114 articles were excluded. Of the remaining 460, 427 were excluded based on the title and abstract, during the screening process. The remaining 33 articles were analyzed in full to proceed with a new, more refined filtering of articles, according to some eligibility criteria, such as: **i)** The use of elderly people in the tests performed; **ii)** Validation/Benchmarking method well defined; and **iii)** Information about the sensing system and protocol of tests performed. With the performance of this analysis, 16 articles were removed, and the rest of which are analyzed in the following sections of this document.

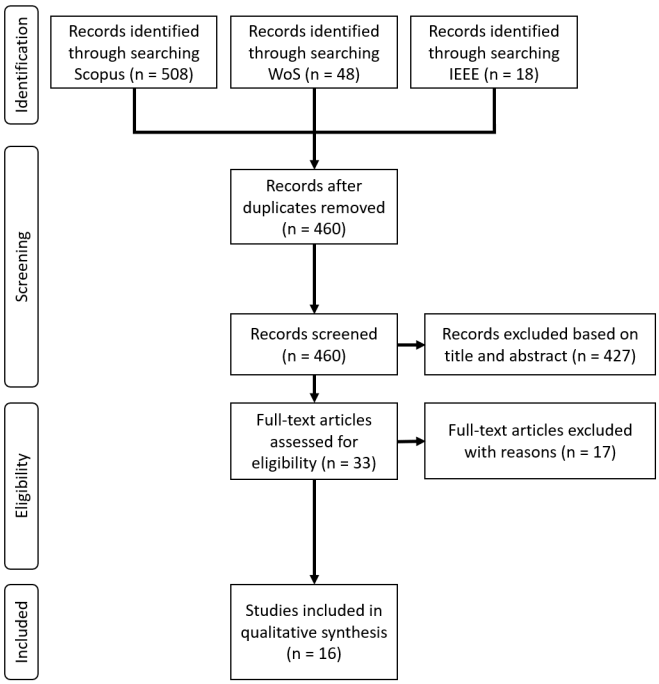


Figure 5: Adopted methodology for Fall Risk Assessment articles research. The flow chart illustrates the steps of the selection procedure.

2.2.2 Reviews Analysis

At an early stage, several reviews on the subject were analyzed. This study has two key objectives: The first step was to place the topic in perspective with the FRA problem. The second goal concerned the many stages that should be examined in the chosen papers in order to identify current trends and constraints.

Several studies assessed the state of the art in terms of the most widely used wearable sensor technologies and their combinations of locations on the human body [17, 27, 28]. Among them, accelerometers and gyroscopes are the most widespread technologies. Regarding sensor positioning, the trunk is the most used segment [10, 27, 28], yet, other approaches were also investigated as feasible locations, making finding a general rule a difficult task [17, 28].

For assessing fall risk or subject classification as a Faller or Non-Faller, several tasks and tests are performed. Quiet standing, Walking tasks and a sit-to-stand tests are usually adopted for static and dynamic evaluations, respectively [17, 28]. However, findings suggest the need for the definition of gold standards in terms of tasks to be performed to face the extremely high variety of proposed approaches [28, 45]. In general, these task give a good information regarding the subjects balance capabilities. However, due to the large number of different factors that can cause falls, FRA requires a multifactorial analysis, where both intrinsic and extrinsic factors must be considered to develop more reliable systems (Figure 3) [45].

Sensor-based features and algorithm investigations were also performed to determine which ones were most effective in distinguishing between Fallers and Non-Fallers [10, 17]. The results demonstrated that there are several factors which affect the choice of a machine learning algorithm, because of its task dependency. In addition, the best algorithms for a particular task depends on several factors, such as the feature set and dataset characteristics [10, 17, 28].

According to [10], different Threshold-based and Machine Learning algorithms (SVM, Decision Trees and Fuzzy Logic) have been employed for FRA, where Tilt features in combination with a decision tree algorithm present the best performance .

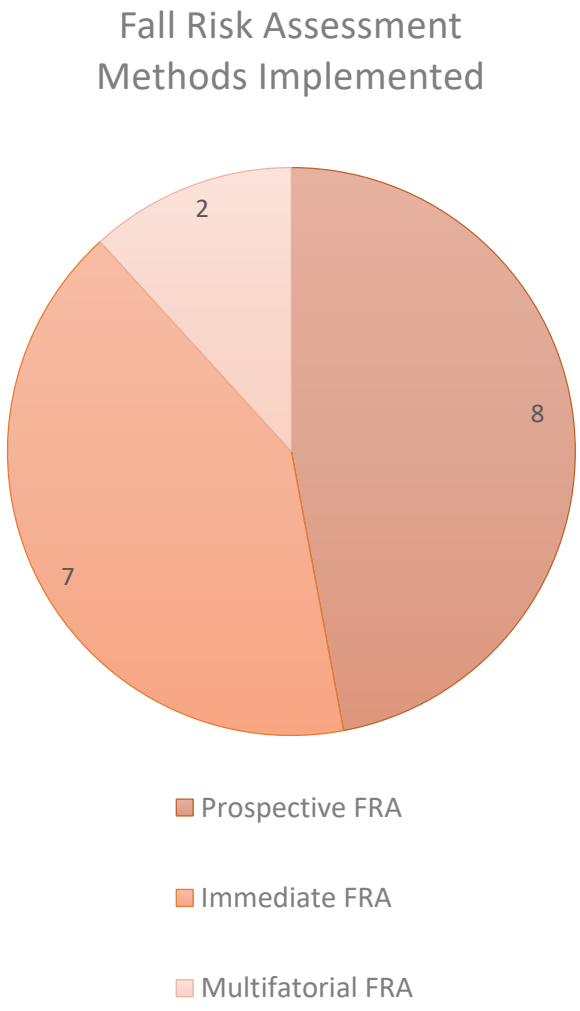
In short, from this initial analysis, it was concluded that the most relevant steps for creating a FRA method are: **i) The data acquisition protocol** - which includes the type, number and location of the sensors used; as well as the participants, the tasks performed and conditions of the tests; and **ii) The algorithms used and results obtained** - either for collected signals processing, collected features, validation methods and models used. Thus, in the next sections, each of the selected articles will be analyzed in relation to these three steps.

2.2.3 Prospective Vs Immediate Fall Risk Assessment

FRA can be performed for the long-term prevention of falls or for a more urgent prevention of falls. In the first one, in general, through clinical tests accompanied by health professionals (such as the Tinetti POMA [20], or TUG [18] tests), or fall history related questionnaires, people are classified as having a high or low risk of falling in a long-term perspective, such as 6 months or 1 year [22]. When the risk of falling is assessed in the short term, it is generally done by a qualitative or quantitative examination of changes in gait and balance characteristics [22, 31]. There is also the potential of determining the fall risk in a multifactorial manner, i.e., computing the fall risk analysis using information about numerous risk variables, such as environmental dangers, gait or balance problems, or even information received from surveys (history of falls, poly-pharmacy, fear of falling, etc.).

As a result, the studied articles were classified into three categories based on the criteria listed above (Figure 6). The first group of articles refers to FRA in a prospective way, the second group of articles refers to FRA in real-time, and the third group of articles refers to papers whose goal is the simultaneous

integration of both prospective and immediate FRA.



(a)

Author (Year)	FRA Method
Rivolta (2015)	Prospective FRA Methods
Rivolta (2019)	
Tang (2019)	
Shahzad (2017)	
Yang (2019)	
Saporito (2019)	
Buisseret (2020)	
Silva (2016)	
Rescio (2015)	Immediate FRA Methods
Parvaneh (2017)	
Cola (2015)	
Di Rosa (2017)	
Sansrimahachai (2017)	
Gagnon (2013)	
Mehmood (2021)	
Annese (2015)	Multifatorial FRA Methods
Greene (2019)	

(b)

Figure 6: (a) Number of studies from each fall risk assessment methods identified. (b) Fall risk assessment method adopted by each study.

2.2.4 Prospective Fall Risk Assessment

2.2.4.1 Sensor System Characteristics and Data Acquisition Protocols

Regarding the sensing methodology, Rivolta et al. [46, 47] proposed an automatic method of determining the Tinetti POMA test score (Figure 7), using a wearable accelerometer located in the subjects'

sternum. The sensor was positioned so that the 3 acceleration components corresponded with the vertical (V), mediolateral (ML) and anterior-posterior (AP) axes of the subject and data were acquired at a frequency of 50 Hz. With regard to the testing protocol performed on each article, it is possible to find some resemblances. In [46], the tests were performed under controlled conditions by group of 13 individuals over 60 years of age. The data acquisition in [47] was done in a similar form, where 90 participants were requested to participate, with an average age of 69.33 ± 16.76 years.

Balance assessment items		**Score**
I1	Sitting balance	0 to 1
I2	Arises from the chair	0 to 2
I3	Attempts to arise	0 to 2
I4	Immediate standing balance (1^{st} 5*s*)	0 to 2
I5	Standing balance (eyes closed)	0 to 2
I6	Romberg	0 to 2
I7	Standing balance (eyes opened)	0 to 1
I8	Turning 360 degrees continuously	0 to 2
I9	Sitting down	0 to 2
	Total Balance score	0 to 16
Gait assessment items		**Score**
I10	Initiation of gait	0 to 1
I11	Step length and height	0 to 4
I12	Step Symmetry	0 to 1
I13	Step Continuity	0 to 1
I14	Path related to the floor	0 to 2
I15	Trunk	0 to 2
I16	Walking stance	0 to 1
	Total Gait score	0 to 12

Figure 7: Items performed during the Tinetti POMA test in [46]. Taken from [46].

An objective approach to assess the Berg Balance Scale (BBS) and Mini Balance Evaluation System Test (miniBEST) using wearable sensors was proposed by Tang et al. [48]. This work used a shoe-based wearable sensor system (*SmartShoe*), which incorporates a total of 3 pressure sensors and 1 accelerometer in each shoe and a last accelerometer placed on the right side of the subject's hip, making a total of 9 sensors. The pressure and acceleration data of the various sensors were sampled at 400 Hz. Thirty elderly subjects were required for the data acquisition protocol. In controlled conditions, subject were asked to perform the BBS and miniBEST assessments with the wearable sensors attached. BBS and miniBEST tests were administered by two trained clinicians and subjects' performances were expressed as numeric scores according to the current clinical guidelines.

A similar approach was provided by Shahzad et al. [49], where an objective and unsupervised method to obtain functional balance and mobility assessment-based fall-risk of older adults was proposed. For data acquisition, a 3-axis accelerometer placed in the lower back of the subjects was used and these accelerations were collected at a rate of 41 Hz. Twenty three elderly volunteers carried out, under controlled conditions, the BBS assessment. Then, all the subjects were required to perform directed doutine tasks twice, which included TUG test, Five Times Sit-to-Stand test (FTSS) and Alternate Step Test (AST). It should

be noted that while the subjects performed these tests, an observer inserted the markers in the data stream to note important sections of the tests.

A smart insole placed at the subjects' feet, equipped with the 1 Inertial Measurment Unit (IMU) and a 16 pressure sensors array for data collection was developed by Yang et al. [50], who proposed four environment-adapting TUGs, designed to assess subjects' ability to adapt gait in complex environments. The sampling frequency for data acquisition from each sensor was 100 Hz. Data were collected from 10 healthy young adults. These subjects, under controlled conditions, were requested to perform several versions of the TUG clinical scale, which were: the Extended TUG; Incline TUG; Bypass TUG and Overpass TUG. Each of these activities was filmed to make a division of the different stages of each test and make a ground truth reference for the algorithms used.

A novel method to estimate TUG from activities recorded in 3-days free-living was also proposed by Saporito et al. [51], using for that a wearable pendant. The method was used to monitor changes in mobility following total hip arthroplasty. A Senior Mobility Monitor (SMM) containing a 3-axis accelerometers and a barometer was used on the chest of the individuals for the acquisition of data. The acquisition of the accelerometer data were made at a frequency of 50 Hz and the barometer at 25 Hz. Data collection was made to 239 community dwelling. All 239 participants underwent a standardized TUG test at baseline before being supplied with an SMM, which was used as a Benchmark validation. Following that, the 239 subjects wore the device for 7 days, and 15 study participants followed a 12-week home-based exercise program with video instructions on a tablet. In the follow-up period, participants were required to wear the senior mobility monitor (SMM) in their usual environment during their normal ADLs, with no restriction for indoor or outdoor activities.

The main goal of Buisseret et al. [52] was to show the possibility of designing a test which identified nursing home residents presenting fall risk, the test's accuracy was assessed by comparison to their actual falls during a six months period after the test. A secondary objective was to present three different ways of FRA: a standard clinical test (TUG), a standard clinical test augmented by sensor measurements (TUG+), and an AI algorithm based on sensor measurements only. The data acquisition stage was achieved through 3-axis accelerometer, gyroscope and magnetometer sensors. These were placed in the lumbar region of the individuals and the data of each sensor were collected at a sampling frequency of 100 Hz. A total of 73 healthy elderly people were classified as fallers and non-fallers, based on fall events in a 6-month period follow-up. Two tests were performed, separated for a period of 6 months (t1 and t2, respectively). At t1, with controlled conditions, the TUG test was performed in all participants. Participants then performed the six-minute walking test with the right to take short pauses as required.

Finally, Silva et al. [53] also proposed a methodology for a wearable inertial sensors-based analysis to the TUG test. The IMU of a smartphone was placed in the pocket, waist or leg of the subjects and collected acceleration and angular velocity data at a rate of 200 Hz. A total of 18 healthy dwelling subjects performed the tests, and were classified as high or low fall risk based on the standard FRA tools of Tinetti

POMA and TUG time. A medical history questionnaire was given to evaluate chronic conditions, audition and vision problems, other relevant medical conditions, and exercise habits. In contrast to the traditional TUG test, in this project, instead of keeping the walking distance of 3 meters fixed, auditory cues were used to instruct the person to stand up and walk forward and then turn, walk back, and sit down. Table 2 depicts the sensor characteristics and tasks adopted in the studies that performed FRA based on clinical scales analysis.

Table 2: Adopted sensor and task specifications for data acquisition in propective FRA studies, where: FS = sampling frequency, A= accelerometer, G= gyroscope, M= magnetometer, P= pressure sensors, B= barometer and NA = Not Available

Work	Sensors	Number	Location	FS	Participants	Tasks
Rivolta et al. [46]	A	1	Chest	50 Hz	13 subjects >65 years	POMA
Rivolta et al. [47]	A	1	Chest	50 Hz	90 subjects >65 years	POMA
Tang et al. [48]	A,P	3,6	Feet, Waist	400 Hz	30 Subjects >65 years	BBS miniBEST
Shahzad et al. [49]	A	1	Waist	41 Hz	23 subjects >65 years	TUG, AST FTSS
Yang et al. [50]	A,G,M,P	2,2,2,32	Feet	100 Hz	10 subjects	TUG
Saporito et al. [51]	A,B	1,1	Chest	50 and 25 Hz	15 subjects >65 years	TUG Free Living
Buisseret et al. [52]	A,G,M	1,1,1	Waist	100 Hz	73 subjects >65 years	TUG
Silva et al. [53]	A,G	1,1	Waist,Pocket, Leg	200 Hz	18 subjects >65 years	TUG POMA

2.2.4.2 Algorithms for Fall Risk Determination

For the automatic determination of the Tinetti test score, Rivolta et al. [46] calculated 8 features: i) tilt angle, ii) sit to stand time, iii) SampEN Eye Closed, iv) Romberg VM Power, v) SampEN Eye Opened, vi) SampEN 360° rotation, vii) Stand to Sit time, viii) Walking PSD Area. The extracted features were used with linear regressions to establish a relationship between them and the test results. LOO approach was used for validation of the algorithm and the results suggested that the automatic method might be a promising tool to assess the falling risk of older individuals.

Rivolta et al. [47] processed the acquired data using an High-Pass Filter (Butterworth, 3rd-order, 0.5 Hz cut-off frequency), to filter out the effect of respiration. Next, 21 features were extracted and used in standard linear regression (LM) and single hidden layer Artificial Neural Network (ANN) models after Least Absolute Shrinkage and Selection Operator (LASSO) feature selection, in order to classify subjects

with high or low risk of fall, according to Tinetti's clinical scale score. Results showed that the Pearson's correlation between the Tinetti score and a subset of 9 features was 0,71. The misclassification error of high-risk patient was 0,21 and 0,11, for LM and ANN models, respectively. Validation was done using the Hold-Out (HO) method, as well as the gold standard classification provided by health professionals in [47] as well as in [46].

The pressure and acceleration data acquired by Tang et al. [48] were firstly downsampled to 25 Hz. Two types of features were then determined: i) common features like standard deviation, entropy, number of mean crossings, mean absolute deviation, root mean square, etc; and ii) Activity Specific Features (ASFs), such as placing alternate foot on stool, walking, sit-to-stand posture transitions, retrieving objects from the ground, and standing. Some of these features were selected and used in an Support Vector Regression (SVR) algorithm, with a linear kernel, to determine the BBS and miniBEST scores. A Leave-One-Out (LOO) cross validation was used to evaluate the performance of the SVR. The evaluation of the algorithm's performance was made from the error between the actual scores to BBS and miniBEST tests from health professionals and the score predicted by the model. The results show that the wearable sensor system has a capability to estimate the BBS and miniBEST scores with absolute mean errors and standard deviations of 6.07±3.76 and 5.45±3.65, respectively, demonstrating high agreement with fall history-based risk assessment.

A total of 173 features were determined from the data acquired by Shahzad et al. [49], and several Machine Learning models tested to obtain the BBS score values: multiple linear regression (Linear Least Square - LLS), LASSO regularized linear regression, DTs and DTs ensemble models. Root-Mean-Square Error (RMSE), mean and standard deviation of absolute error, Pearson's correlation coefficient and p-value were used to evaluate the results obtained, which showed that the average of two BBS estimates based on test and retest yielded a strong correlation $p = 0.86$ with the standard BBS score. Also, high correlation ($p = 0.90$) and low RMSE (1.66) was observed.

Regarding the algorithms implemented by Yang et al. [50] for the segmentation and automatic analysis of the various TUG tests, the denoising of the pressure values obtained was done first, followed by the calibration and filtering of the data. Three types of gait features were extracted: spatiotemporal features (such as threshold, forefoot contact time, rear foot contact time, full contact time, gait cycle time, TUG time, stride length, etc), spatial feature (pitch, roll, and yaw angles), and pressure features (sole average pressure, COP location, etc). These enabled to make TUG phase recognition, without any information regarding the use of any regression or ML method. The proposed algorithm was capable of extracting gait related spatial-temporal features with all mean accuracies over 92%. The system also achieves a mean accuracy of 92.23% in segmenting the five TUG phases.

Regularized linear models were used to classify TUG test scores in the work developed by Saporito et al. [51]. These were trained with 6 mobility indicators, including the categories of walking quantity, walking quality, chair rise quality, and the duration of active and inactive periods. This algorithm was tested and

validated with a LOO approach. The results were evaluated with Spearman's correlation coefficient, AUC-ROC curve, and p-values. A strong correlation was observed between the remote TUG and the standardized TUG with Spearman's correlation coefficient (p=0.70), while the mean absolute error was 2.1 s. The remote TUG demonstrated high test-retest reliability in the model development population (ICC=0.94). The AUC-ROC curve for the classification of the participants with standardized TUG>10s was 0.89.

For the classification of subjects into 2 classes (high risk or low risk), Buisseret et al. [52] used two methodologies: threshold-based and AI-based methods. Features such as TUG time, standard deviation and fractal dimension obtained from the six-minute walking test were used in these algorithms. The validation method of this project was the ground-truth status assigned after the 6-month follow-up period. In the work carried out by Silva et al. [53], duration and the number of steps of each phase of TUG, statistical measures (number of times the magnitude signal crosses the mean value, energy, entropy, standard deviation, mean value, median deviation, Root Mean Square, etc) and frequency-based features obtained using the Fast Fourier Transform (FFT) of the accelerometer signals, were calculated. P-values were used to differentiate between the high and low fall risk classes. Results show that only features belonging to the component of first walking phase and first turning phase of TUG showed significant differences between the mean of high and low risk groups. Although the TUG test is used, it is not used as a benchmark to evaluate the performance of the wearable system.

Table 3 depicts the methods and validation approaches implemented by the works whose main goal was to assess subjects risk of falling through the automatic analysis of clinical scales.

Table 3: Methods and validation approaches adopted by the 8 selected articles, where FRA is made in a prospective, through clinical scales. In this table: ML = machine learning, Th = threshold-based, Acc = accuracy, Sens = sensitivity, Spec = specificity, CV = cross-validation, LLS = Linear Least Square Regression and LASSO = Least Absolute Shrinkage and Selection Operator regression and N\A = Not Available

Work	Model Used	Validation Method	Clinical Score	Results
Rivolta et al. [46]	Linear Regression Model	Leave-one-out CV	Tinetti POMA	Acc = 84.6% Sens = 85.7%; Spec = 83.3%
Rivolta et al. [47]	(ML) Linear Regression Model; (DL) Single Hidden Layer ANN	Holdout (60% training; 40% testing)	Tinetti POMA	Sens (ML) = 71% Spec (ML) = 81% Sens (DL) = 86%; Spec (DL) = 90%
Tang et al. [48]	(ML) Linear Kernel SVR	Leave-one-out CV	BBS miniBEST	Mean error: 6.07±3.76 (BBS); 5.45±3.65 (MiniBEST)
Shahzad et al. [49]	LLS and LASSO models	10-fold CV	BBS	Mean error: 1.9 ± 2.53 (LLS); 1.44 ± 1.98 (LASSO)
Yang et al. [50]	N\A	N\A	Video recordings from TUG	Acc (gait cycle count) = 100% Accu(segment TUG phases) = 92.23% Accu(spatial–temporal features) = 92%
Saporito et al. [51]	LLS and LASSO models	10-fold CV	TUG	Mean error: 1.9 ± 2.53 (LLS); 1.44 ± 1.98 (LASSO)
Buisseret et al. [52]	Regularised Linear Model	Leave-one-out CV	Tinetti POMA	Mean error: 2.1 ± 1.7s

2.2.5 Immediate Fall Risk Assessment

2.2.5.1 Sensor System Characteristics and Data Acquisition Protocols

The feasibility of a system for the detection of the fall risk was studied by Rescio et al. [54]. Aiming for the detection of a fall before the impact, EMG sensors were placed on the subject's lower limbs (tibialis and gastrocnemius muscles groups), making a total of 4 EMG sensor that acquired at a sampling frequency of 1000 Hz. Seven healthy subjects simulated, under controlled conditions, the following events: 1) idle in normal condition and in the presence of deviant auditory stimuli; 2) walking in normal condition and in the presence of deviant auditory stimuli; 3) ADLs (sitting down, standing up, bending and lying down); 4) situations of instability (typical of a fall event), using a tilting platform to bring the individuals beyond their

limits of stability, like in a fall event.

A postural transition quantifier using a chest-worn wearable sensor to identify frailty in community-dwelling older adults was explored by Parvaneh et al. [55]. A validated software package was used to identify body postures and postural transition between each activity since it was hypothesized that qualitative data from postural transitions could distinguish non-frail from frail elders. Each of the 120 healthy elderly subjects, 43 non-frail and 76 pre-frail/frail (according to Fried's criteria), worn the IMU for 48 hours in free-living conditions and the IMU data were acquired at sample rate of 50 Hz.

The method proposed by Cola et al. [56] was designed to achieve continuous monitoring of gait in a completely unsupervised fashion. The sensorization was achieved through the use of a 3-axis accelerometer (Shimmer 2R device) mounted on the waist of the subjects. The frequency of data collection was 51.2 Hz. The tests were performed on 30 healthy subjects and are ilustrated in Figure 8. The participants were asked to wrap Neo-G thigh support straps around their knees and performed gait-related experiments in a corridor 4 times. Four types of gait were collected - normal, mild knee condition, severe knee condition, and mild condition in both knees, using the support straps. Four users were also involved in uncontrolled experiments during 48 hours.

Figure 8: Three different types of impaired gait simulated in Cola et al. [56] data acquisition protocol: mild knee condition, severe knee condition, and mild condition in both knees. Taken from [56].

A Fall Risk Index based on gait pattern recognition was proposed by Di Rosa et al. [57], where a pair of electronic insoles was used to assess fall risk. The insoles contained 14 thin-film resistive pressure sensors, integrated with one accelerometer and gyroscope. No information was given regarding the acquisition frequency. Thirty-nine healthy elderly volunteers were enrolled across three research centres. Subjects used the insole system for two weeks during daily life and no instruction was provided for specific tasks. Participants completed a detailed diary reporting the hours of use and also underwent functional performance-based assessments of fall risk: Tinetti POMA; TUG; and Dynamic Gait Index (DGI), supervised and evaluated by trained physicians.

A novel FRA system able to dynamically perform gait analysis to detect the risk of falls in the elderly in real-time was proposed by Sansrimahachai et al. [58]. A mobile phone accelerometer was worn on the subject's waist to acquire acceleration data and remotely monitor gait parameters in a timely fashion. No information was given regarding the sampling frequency. The tests were carried by 12 elderly subjects, as they perform walking activity in several rounds.

Three FRA methods were proposed by Gagnon et al. [59]. The first uses a statistical model while the two others exploit ANNs. Non-invasive augmented shoes were instrumented with 4 FSR sensors and a variable bending resistor (Flex Sensor), totalling 5 sensors per shoe to acquire plantar pressure parameters. The sample frequency was 100 Hz. Four healthy young adults were asked in to walk along a corridor in a straight line four times in controlled conditions. During the tests, an auditory disturbance was applied to prevent the influence of heel strike sounds on the user gait and four visual conditions were explored: no visual disturbance, lightly obscured vision, obscured vision and highly obscured vision.

A proactive FRA mechanism for elderly was proposed by Mehmood et al. [60], through the performance of Muscle Fatigue Analysis on data collected using a system of wearable sensors to collect data of four activities in a controlled environment. The data were acquired through EMG sensors placed on the Tibialis Anterior and Gastrocnemius muscles of subject's dominant leg and the EMG data were acquired at a rate of 500 Hz. Twelve subjects (10 young adults and two elderly subjects) were asked to perform four different activities in a controlled environment such as: walking, running, sit on a chair and get up, ankle stretch in dorsiflexion and stretch out planter flexion. Figure 9 depicts the most used sensor technologies in the immediate FRA group of studies, as well as the several sensor locations used.

Figure 9: (a) Most used sensor technologies identified in the several immediate FRA works. (b) The several sensor locations found in the immediate FRA works: i) Tibialis [54, 60] ,ii) Gastrocnemius [54, 60],iii) Chest [55], iv) Waist (Lower back) [56, 58] and, v) Feet (Instrumented shoes) [57, 59].

2.2.5.2 Algorithms for Fall Risk Assessment

The acquired EMG signals were amplified by an instrumentation amplifier, then low-pass filtered, and A/D sampled in the work developed by Rescio et al. [54]. A band-pass filter using a 12th order FIR filter with cutting frequencies between 20 Hz and 450 Hz was also used to reduce artifacts and avoid signal aliasing. Then, the signals have been fully rectified and linear enveloped. Muscle Co-Contraction Indices (CCI) and Integrated EMG (IEMG) were some of the features calculated and used in a threshold-based approach to classify two classes: "instability event"and "normal event". The results show average values of sensitivity and specificity of about 70%. The average delay with respect to the start of the plane stimulus perturbation has been measured in about 200ms, considering all imbalance events simulated.

Parvaneh et al. [55] determined features like the total number of the following activities: i) ratio of cautious sitting; ii) transitions; iii) number of sit-to-stand; iv) number of sit-to-walk; v) number of stand-to-sit; vi) number of stand-to-walk; vii) number quick sitting; and viii) number walk-to-stand. The independent-sample *t*-test or χ^2 test was used to evaluate between-group differences in demographics and health parameters. Between-group comparisons for postural transitions parameters were done using general linear model tests. A Logistic regression model was used to identify independent predictors of frailty. In the end, subjects were classified as non-frail or pre-frail/frail. Results show that the total number of postural transitions, stand-to-walk, and walk-to-stand, were, respectively, 25.2, 30.2, and 30.6% lower in the frail group when compared to the non-frail group. Furthermore, the ratio of cautious sitting was significantly higher by 6.2% in frail compared to non-frail.

Acceleration signals acquired by Cola et al. [56] were low-pass filtered at 20 Hz using a second-order Butterworth filter. Then, features such as mean, median, Peak-to-Peak amplitude (P2P), RMS, standard deviation, and Zero Crossing Rate (ZCR) were calculated on the gait segmented samples, creating a gait instance. These were used in the anomaly detection algorithm: a binary classifier based on neural networks, random forest and k-Nearest Neighbours (k-NN) analysis, where each gait instance is either classified as abnormal or normal. Leave-one-out was used as validation method. The results show that, on average, 23% of the gait instances were classified as abnormal when the user was walking normally. Instead, when the user's gait pattern was abnormal, the rate of detected anomalies increased up to 84.5% in the mild condition, 96.1% in the severe condition, and 93.1% in the mild both conditions.

The algorithm proposed by Di Rosa et al. [57] started with three filtering operations carried out to isolate walking bouts. In a second step, several features were extracted from the force and inertial sensors data, and an optimization algorithm identified different weights for each parameter. For these parameters, two thresholds were identified: one describing the representative value for no fall risk and the other describing the representative value for a risk of fall. Normalized scales were built on each parameter's value contribute to the FRI, which ranges from 0 to 100. The results found showed that the FRI captures the risk of falls with a similar accuracy to the conventional performance-based tests of FRA. These preliminary findings

support the idea that the system may have clinical utility for long-term FRA at home and in the community setting.

No noise filtering was carried out in the study carried out by Sanrimahachai et al. [58]. Data were partitioned by using a sliding window of 30 seconds. This article's algorithm tried to differentiate between normal and abnormal gait through observational gait analysis and consisted of the determination of 2 indexes – Stability index and Symmetry index. To calculate these walking gait indexes, the Dynamic Time Warping algorithm (DTW) was used. DTW is commonly applied in the field of time series analysis to measure similarity (distance) between two temporal sequences [61]. This study used the accuracy to evaluate the algorithm's performance and the results (Accuracy = 83.8%) demonstrate that the system can differentiate between the walking pattern of the elderly with normal gait and that of the elderly with abnormal gait.

Four fall risk levels were assessed with the pressure data acquired from the walking tests carried out by Gagnon et al. [59]. A statistical model and 2 types of ANNs were used (ANN-RT and ANN-S). The ANN-RT model processed data at a rate of 100Hz, while the ANN-S relays on a windowing approach, namely for each window, several gait features are computed. Similar rates of 76.6% and 75.28% for respectively the ANN-RT and the ANN-S were obtained. Furthermore, results did show, for most of the participants, a corresponding increase of the average risks when a decrease visual perception was applied.

The calibrated data in the FRA system proposed by Mehmood et al. [60] contained a sequence of onsets which indicated the muscle activation time. After localizing the onset intervals, two time-domain features - Mean Absolute Value (MAV) and Root mean square (RMS) and two frequency domain features - Mean Frequency (MNF) and Median Frequency (MDF), were computed on each onset interval. If the two frequency domain features show the decreasing trends, and the two time domain features indicate increasing trends, then EMG signals must have the muscle fatigue. If level of muscle fatigue is found above a certain threshold value, then the FRA is computed, otherwise no warning alarms are generated. Results justify the claim that muscle fatigue in the two muscles assessed could lead to fall events in an elderly person.

Table 4 presents the methods and validation approaches implemented by the works whose main goal was to assess subjects risk of falling through the classification of fall risk or abnormal gait situations, as well as the respective results.

Table 4: Methods and validation approaches adopted by the 7 selected articles, where fall risk assessment is made in an immediate manner, through fall risk and imbalance situations detection. In this table: ML = machine learning, Th = threshold, Acc = accuracy, Sens = sensitivity, Spec = specificity, CV = cross-validation, DGI = Dynamic Gait Index, NN = Neural Networks, ANN = Artificial Neural Networks

Work	Model Used	Validation Method	Reference Measures for Classification	Results
Rescio et al. [54]	Th-Based Model	10-fold CV	Type of event (Instability or Normal event)	Sens: 70%; Spec: 70%
Cola et al. [56]	NN, RF K-NN	Leave-one-out CV	Type of gait (Normal or Abnormal)	Acc: 96.1% (Severe condition);
Di Rosa et al. [57]	Th-Based Model	N\A	Fall Risk Index (FRI) (3 Levels)	Acc: 69%; ROC: 0.73; FRI close to DGI (p = 0.001)
Sansrimahachai et al. [58]	Dynamic Time Warping Algorithm (DTW)	N\A	Type of gait (Normal or Abnormal)	Acc= 83.8% (Fall Detection)
Gagnon et al. [59]	Statistical Model; ANN-RT; ANN-S	N\A	Fall Risk (4 Levels)	Higher risk w/ visual impairments
Mehmood et al. [60]	Th-Based Model	N\A	3 Fall Risk Levels through Muscle Fatigue	Higher risk w/ muscle fatigue

2.2.6 Multifactorial Fall risk Assessment

2.2.6.1 Sensor System Characteristics and Data Acquisition Protocols

On the basis of the subject's clinical state, ambient factors, electromyographic co-contraction analysis, and electroencephalographic analysis, a novel approach for fall-risk on-line evaluation is provided by Annese et al. [62]. Nine healthy subjects participated in the data acquisition protocol carried out. However, very little information was given regarding the activities performed, only that the tests were made in controlled conditions. This fall-risk assessment approach was implemented by a complete cyberphysical system made up of EEG and EMG wearable recording systems. EMG sensors were placed in gastrocnemius and Tibialis muscles and 32 EEG Electrodes are used. EEG data were collected at a frequency of 500 Hz. The remaining data is acquired by surveys and remains constant during the experiments.

Greene et al. [63] discussed a digital FRA protocol. Wearable sensors' data were captured from 8521 elderly subjects during a TUG test, along with self-reported questionnaires on falls risk factors (Polypharmacy, mobility problems, vision problems, dizziness, fall history, etc), which were collected between 2014 and 2019. Two IMUs were used bellow both knees of the subjects who participated in the tests.

2.2.6.2 Algorithms for Fall Risk Classification

The Field Programmable Gate Arrays (FPGA) built by Annese et al. [62] synchronized both signal from EMG and EEG sensors. EMG data were band-pass filtered (from 0.1Hz to 200Hz) and down-sampled to match the EEG sampling frequency. The algorithm for determining the risk of fall is multifactorial, since it uses several factors (Figure 10): a baseline risk factor related to the clinical condition of the patient (BF); the risk associated to the environmental conditions (EnF); the EMG co-contraction-based fall risk (CF) and the risk factor considering the EEG data study (EF). BF and the EnF are considered constant in this algorithm, while CF and EF are re-calculated as soon as a new step is recognized during gait. A set of weights is given to each one of this risk factors and a fall risk value from 0 (very low risk) to 1 (very high risk) is given. Thus, the fall risk never can be lower than the baseline and environmental combination and it can be real-time calculated since the proposed procedure never uses future samples. A Detectable fall-risk increasing (+1.5%) when obstacles are overcome during the tests protocol was observed by the algorithms proposed. This analysis verified that co-contraction time can be used to evaluate the balance and the stability of the subject during gait.

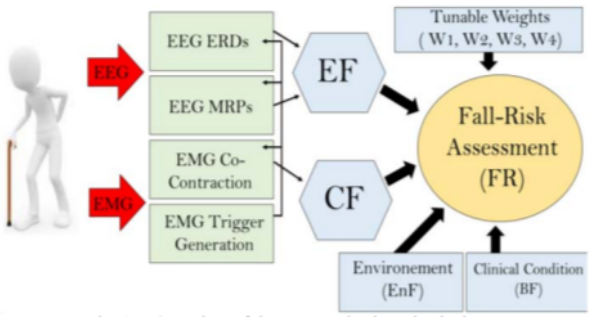


Figure 10: Overview of the cyber-physical system proposed in [62].

The frailty estimate algorithm proposed by Greene et al. [63] used IMU data and anthropomorphic data applied to a logistic regression model. The falls risk and frailty scores are essentially posterior probabilities and defined as low risk if less than 50%, 50–69% is considered medium risk, while high risk and very high risk are 70–89% and greater than 90%, respectively. The TUG test was used as a benchmarking technique for validation. Different mobility scores regarding the subjects fall risk were evaluated using p-value. Each mobility score was significantly associated with falls history. A large proportion of patients were predicted to be at high risk or very high risk for a measure of falls and frailty.

2.3 Discussion

The analysis carried out in the previous sections provided a variety of information regarding the most used methods nowadays for automatic FRA, as well as their main limitations. It is currently difficult to

verify a clear trend in each type of FRA analyzed, since several possibilities are being tested. However, it was possible to find some common points in each of the investigated groups of FRA.

2.3.1 Prospective Fall risk Assessment

It was possible to extract some common characteristics in the many stages investigated in studies that do prospective FRA through automatic analysis of clinical scales, despite no clear trend being detected, since there is a vast range of hypotheses to evaluate in each phase.

In terms of sensing modalities, the accelerometer was the most popular form of sensor, followed by gyroscopes and pressure sensors (Table 2). There is no apparent trend in terms of the number of sensors used, as some research use only one and others employ multiple sensors. In terms of placement, the lumbar region is the most common, followed by the sternum, legs, and the usage of instrumented shoes.

The testing protocol in this group of articles is usually achieved by replicating clinical tests' tasks, namely TUG, Tinetti POMA and BBS, or modified versions of these (Table 2). Most of the articles collect data from elderly subjects in controlled conditions, accompanied by health professionals who also perform a manual assessment of said tests, in order to evaluate the algorithms performance.

The algorithms carried out, generally apply statistical or regression models to differentiate between subjects with low or high fall risk, using features related to gait and time of testing and statistical metrics such as the p-value for assess the level of correlation of these features with fall risk (Table 3).

The use of small datasets is the main limitation found in this type of articles, as well as the fact that generally, the results obtained only indicate two levels of risk of falling (high or low), or identify people as fallers or non-fallers, with no reference to numerical scales or probabilities of falling.

2.3.2 Immediate Fall risk Assessment

Table 4 shows a summary of the analysis made to the papers which studied FRA associated with the balance and gait abnormalities detection, and similar conclusions to those obtained previously can be drawn for this group of articles as well.

As can be observed in Figure 9, accelerometers or inertial sensors were once again selected by the vast majority of articles as the type of sensor to be used for data collection. This group also confirms EMG and pressure sensors as viable alternatives to complete the acceleration data.

Walking was the most commonly accessed activity in data collecting protocols, and in most cases, this activity is carried out with varying degrees of abnormality, using visual or physical disturbances for the goal. Only few studies make this assessment in totally free-living situations.

There is a great variability of algorithms applied to compute the fall risk. Statistical models, threshold models, or even AI-based methods are used in these algorithms. However, most studies only make a binary distinction, between normal gait or abnormal gait, and there is no conversion to a risk level or scale (Table

4). In addition, other drawbacks, such as the aforementioned use of a reduced dataset and the lack of information regarding validation and benchmarking methods were found in these group of articles.

2.3.3 Multifactorial Fall risk Assessment

The amount of articles found that determine the risk of falling in real-time in a multifactorial way shows the novelty of this theme, as well as its margin of progression and need for investment in improvements in the near future. The addition of questionnaires on risk factors for falls, such as the subjects' medication, vision problems, or psychological disorders, appears to be the most adopted path for a more complete risk analysis, adding its results to other assessment methods, such as frailty or balance clinical scales, or the assessment of gait abnormalities, based on the analysis of the articles that perform a multifactorial FRA. Inertial and EMG data were still used to quantify fall risk. Furthermore, ambient data enabling a more in-depth investigation of additional fall risk factors unrelated to individuals was mentioned in these types of studies. Another difference between this type of article and those that calculate risk levels solely through clinical tests or gait quality analysis is weight attribution for the various risk calculation approaches.

Chapter 3

Daily Activity Recognition: Literature Review

Despite the progress toward real-time and multifactorial analysis of the risk of falling, as discussed in Chapter 2, there is still a need for further enhancements so that the fall risk analysis takes into account as many information as possible. One of the fall risk variables stated in Section 2.2 that was not found during the literature search is the awareness regarding the daily activities which are being conducted at the time of a fall event. Since this is an important factor to consider when developing a FRA system, the following sections will show and discuss the research made on the most recent trends in automatic recognition of ADLs and fall events based on wearable sensors use and the employment of AI-based models.

Initially, several reviews on activity recognition were initially analyzed with the purpose of having a general overview over the topic and answer the following questions: **i)** What types, number, and location of wearable sensors are most adopted?; **ii)** Which ADLs are performed in experimental protocols for data acquisition?; and **iii)** Which algorithms are being implemented for ADL recognition? Following that, various ML and DL-based ADL recognition works were analyzed in order to have a better understanding of the most recent trends in automatic recognition of ADLs and fall occurrences using wearable sensors.

3.1 Review's Analysis

Offline activity recognition is the most basic method of activity recognition [64]. Sensors are placed in the surroundings or directly on the subject in most applications, and data is gathered, labeled, and supplied to classification algorithms for activity recognition [64–66]. Plenty of studies on activity recognition used wearable sensing technologies [66]. Simpler wearable systems rely on a single accelerometer sensor, and others combine different sets of sensors on different parts of the body to fully capture the subject's movements [64, 65]. Moreover, researchers have created various benchmark datasets to evaluate human

activity recognition algorithms freely [65]. Once the raw data has been collected, must be labelled correctly in accordance wit the performed activities. Then, time and frequency-domain features are extracted from the acquired data [64].

Most offline approaches rely on supervised machine learning models for activity recognition, such as SVM, Decision Trees (DTs) and K-NN [64]. However, on several benchmark datasets used by machine learning-based solutions, conventional (CNN and LSTM) and hybrid (CNN-LSTM) deep learning models have lately shown improved and promising performances, cutting down on the time spent on data processing and feature extraction [64, 65].

Class imbalance of the datasets and the lack of postural transitions recognition were some of the limitations which should be addressed in the future [65, 67]. Furthermore, the datasets generated are person, time and sensor dependent. These factors lead to challenges when training and testing with person independent data [65, 68]. Also, concerns involving the validation methodologies that have been employed across literature were addressed by Shany et al. [69]. Future improvements were also given, namely: **i)** recruit a large enough dataset to ensure the training and test data is sufficient; **ii)** collect data strategically, specifically to manoeuvre around the issue; or **iii)** external validation across completely different samples of the same cohort of interest.

As a result of the initial review articles analysis, and also in accordance with what was done in Chapter 2, each of the selected ADL recognition papers was analyzed regarding three phases: **i) data acquisition** - which includes the type, number and location of the sensors used, as well as the participants, activities performed and its conditions; and **ii) algorithms and results** - either for signal processing, feature extraction, validation and evaluation models used.

3.2 Machine Learning-Based Approaches

3.2.1 Data Acquisition Methods

A framework for fall event detection was created by Kambhampati et al. [70] with the main goal of finding suitable classifiers to effectively classifying different types of fall and non-fall events. The data acquisition protocol was performed in 6 healthy subjects, where they performed several activities, such as sitting, standing, walking, bending, and running, with consecutive falls in different orientations, using a waist-mounted accelerometer with a sampling frequency of 40 Hz. The same sensor and location was also used by Gupta et al. [71], to classify six ADLs and transitional events, namely walking, jumping, running, sit-to-stand/stand-to-sit, stand-to-kneel-to-stand, and standing. Seven healthy young individuals (22-28 years old) were asked to wear the device and perform the activities for about 2–3 min each in controlled conditions. The subjects' data were collected at a sample rate of 126 Hz

A smart watch was used in the work carried out by Gomaa et al. [34] to collect data and classify 14 ADLs with different combinations of sensory data and a random-forest-based classifier. Data were acquired from female and male young adults at a sampling frequency of 50 Hz from accelerometer and gyroscope sensors, while the volunteers repeated each activity several times while wearing the smart watch on the right wrist.

Instrumented shoes capable of recording movement and foot loading data for devise an activity classification algorithm were introduced by Moufawad et al. [72]. Three different types of sensors were used in this design. An IMU, a barometer and a force sensing insole with 8 sensors, were used in the instrumented shoe, making it a total of 10 sensors, whose signals were acquired at a frequency of 200 Hz. The data acquisition protocol was performed in 10 elderly subjects. Each participant wore the instrumented shoes and, in semi-controlled conditions, a predefined track was followed to mimic ADLs, included walking, sit-to-stand and stand-to-sit transfers, sitting and standing bouts, uphill/downhill and upstairs/downstairs walking, and elevator use. An observer followed the participants and marked each period of activity.

Nguyen et al. [73] used an IGS-180 motion capture suit for data acquisition, and it was equipped with 17 IMUs with a sampling frequency of 60 Hz. Objects were strategically placed at different locations and heights, to induce a cleaning task. Participants were asked to navigate the apartment and collect and place these objects. Nine community dwelling older adults who were diagnosed with early stages of PD, including two females and seven males, were recruited for 3, 4, and 5 minutes trials, and perform the tasks in a simulated free-living environment.

An inertial sensors-based physical activity classification system was developed by Awais et al. [74], with older adults as the target population. In this project, 3 types of different sensors were used and placed in the chest (uSense, GoPro), wrist (Shimmer), lower back, and thigh (uSense) of the subjects, making a total of 5 sensors. Each of them acquired data at a frequency of 100 Hz. The dataset was collected in free-living conditions without placing constraints on the way and order that 16 elderly subjects performed their daily activities. A synchronization between the sensors and the camera unit was accomplished for labeling purposes, before attaching the sensors to the subjects.

3.2.2 ADL Recognition Algorithms and Results

Noise reduction and median filters were used by Kambhampati et al. [70] to process the acceleration data. The segementation of the processed signals into small windows came next and 2nd to 5th order cumulants were determined for each window. These features were used in a hierarchical Decision Trees (DTs) structure which implemented classification models such as DTs and SVMs in a sequential manner considering the complexity of activities. First distinguishes fall from non-fall activities. In the second level, the detected falls are classified into four classes. In the third level, the acceleration windows are used for determining different activities of daily living. The main results showed that 2nd and 5th order cumulants

and SVM classifier can achieve optimal accuracies of above 95%, as well as the lowest false positive rate of 1.03%. SVM-based methods had an overall computation time required for executing all the three levels of the system of 54.05 ms.

Auto-correlation functions were applied to the raw acceleration, rotation, and angular velocity signals in the work developed by Gomaa et al. [34], for feature extraction, making it a 33-dimensional feature space per signal as the input to a Random Forest (RF) classification algorithm. Hold-out data splitting (65% for the training Phase and the other 35% for the testing phase) and 5-fold cross-validation with five repetitions were used for the models validation and hyperparameters optimization. These authors showed that combining acceleration with angular velocity allow higher performances. In this work, accuracy reaches about 80% for the whole set of 14 activities. The smallest average sensitivity and specificity achieved by the algorithms proposed are 81% and 98%, respectively.

Two ML algorithms were implemented by Gupta et al. [71] for the classification of 6 classes of daily activities - Naive Bayes (NB) and k-NN. In these algorithms, the acquired signals were segmented in 6 second windows with 3 seconds of overlap, and a total of 31 time and frequency-domain features were extracted, such as Energy, Entropy, Mean, Variance, etc. Feature selection was performed using Relief-F and the wrapper-based method sequential forward floating search (SFFS). The LOO validation strategy was used to train and test the models. Therefore, data collected on six individuals was used to train the system and then the system was tested by classifying the data of the seventh individual accordingly. This was repeated seven times until data from all the individuals was classified. The overall accuracy of the system was 98% from both the classifiers and showed accuracies superior to 95% for all activities.

Regarding data processing, Moufawad et al. [72], low-pass filtered the elevation data (Butterworth order 10 filter, 0.1 Hz cut-off frequency) to remove high frequency noise, and gait cycles shorter than 0.75 s were removed and consecutive gait cycles less than 3 s apart were aggregated to form locomotion periods. From the collected and processed data, features such as angular velocity, body weight, elevation range and the ground slope were determined. These were then used in a Decision Trees (DT) ML algorithm to differentiate 9 classes (Figure 11) with an overall accuracy of 97% across all activities.

Non-linear transform and adaptive threshold algorithms were developed by Nguyen et al. [73] for ADL classification. Five 5-minute trials were randomly selected and used to train the algorithm to find the optimal set of sensor signals and parameters to accurately detect and segment activities during an unstructured cleaning task. The optimization process was used to determine the cutoff frequency, the adaptive threshold parameters, and the IMU signals needed to accurately detect and segment each activity. The optimal parameters were found with respect to the training data set. The algorithms were then validated using the 3- and 4-minute trials from all nine participants. Across all activities, classification was about 90% accurate (sensitivity=89.8%, specificity=97.9%). The segmentation of these activities was made within 350 ms of the ”gold standard”manual segmentation.

Several time and frequency-domain features were extracted by Awais et al. [74] from acceleration,

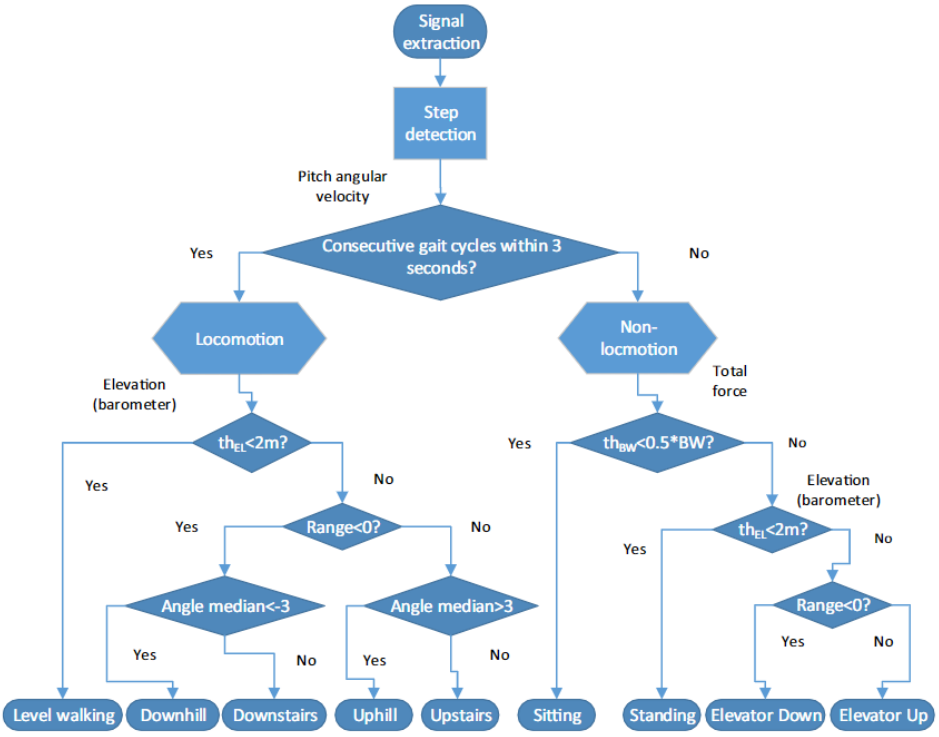


Figure 11: Algorithm flowchart for the DTs activity recognition algorithm from instrumented shoe signals used in [72].

angular velocity, and signals correlation, across a time window of 5 seconds with a 50% overlap. Correlation-based feature selection (CFS), fast correlation-based filter (FCBF), and Relief-F were feature selection methods implemented to eliminate redundant features. The selected features were inserted in an SVM classifier with RBF kernel to classify the 4 activity classes. The effect of class imbalance was compensated by using the weighted SVM, setting the weights of the different classes to the inverse ratio of the training classes sizes. To overcome bias in the training process, the leave-one-subject-out cross validation procedure was used to split the data in training and testing datasets. F-measure was analysed to evaluate the proposed algorithm's performance, and among all single-sensor solutions, the best result was accomplished by the sensor at the lower back (L5), with an F-measure of above 80%. It was noticed an improvement in the performance of 7.3% from single-sensor solution (L5) to two-sensor solution (Thigh+L5). Furthermore, improvement in the performance is almost negligible by increasing number of sensors from two to four.

Table 5 presents a summary of the approaches employed by the several articles which investigated the ADL and fall recognition using ML-based algorithms.

Table 5: Algorithm characteristics adopted by the 6 selected articles who perform ML-based algorithms for fall and activity recognition, where: ML = machine learning, Th = threshold, Acc = accuracy, Sens = sensitivity, Spec = specificity, LOO = Leave-One-Out, CV = cross-validation, NB = Naive Bayes , K-NN = K Nearest Neighbours, RBF-SVM = Radial Kernel Support Vector Machine

Work	Model Used	Validation Method	Type of Classification	Results
Kambhampati [70]	DT, SVM, NB	N\A	Fall vs Non-Fall Fall Class ADLs	Acc: 95%; FPR: 1.03%
Gomaa [34]	Random Forest	Hold-out CV 5-fold CV	14 ADL Recognition	Acc: 80%; Sens: 81%
Gupta [71]	NB K-NN	LOO CV	6 ADL Recognition	Acc: 98%
Moufawad [72]	ML Decision Trees	N\A	9 ADL Recognition	Acc: 97%
Nguyen [73]	Th-Based Model	Hold-out CV	7 ADL Recognition	Acc: 90% Sens: 90.8% Spec: 97.8%
Awais [74]	Weighted RBF-SVM	LOO CV	4 ADL Recognition	F1-Score: 80%; (1 sensor modality)

3.3 Deep Learning ADL and Fall Recognition Approaches

3.3.1 Data Acquisition Methods

The majority of data gathering and validation of Deep Learning-based algorithms for falls and ADL recognition is done using publicly available datasets, which is significant for benchmarking and comparing algorithm performance.

The recognition of six human activities was carried out by Altuve et al. [75]. Inertial signals available at the UCI-HAR public Machine Learning Repository from 30 subjects were used on bidirectional LSTM networks, without manually extracting features from the signals. UCI-HAR was also employed in the work developed by Murad et al. [35], where data from miscellaneous benchmark datasets were used in several types of Deep Recurrent Neural Networks for human activity recognition. Accelerometers, gyroscopes and magnetometers were the sensors used for data acquisition in these datasets.

In the work proposed by Wang et al. [67], different types of lightweight neural networks were explored to improve the fall detection results. Lower trunk acceleration data from SisFall, a large public dataset, was used in this work. With the same goal, Gil-Martín et al. [76] used the Physical Activity Monitoring Data Set (PAMAP2) for physical activity monitoring. This database contains recordings of 12 different physical activities, performed by 9 subjects wearing three IMUs.

Unlike previous DL-based ADL recognition works, Chung et al. [77] created a testbed with eight body-worn IMU sensors in various locations, as well as an Android mobile device for collecting activity data, in order to perform a sensor positioning and data acquisition study in 9 human activities. Data were acquired from 5 adult subjects in both real-world and controlled environments.

The properties of the data obtained by this group of articles are summarized in Table 6.

Table 6: Public datasets used by Deep Learning-based ADL recognition works, with description regarding sensing methods and location, sample frequency, participants and activities recorded. In this table: A = accelerometer, G = Gyroscope, M = magnetometer and PD = Parkinson's Disease

DataSet	Work(s)	Sensors	Location	Sample Frequency	Participants	ADLs Falls
Private Dataset	[77]	8 A,G,M	Various Locations	100 Hz	5 Subjects	9 ADL
SisFall [78]	[67]	A,G	Waist	200 Hz	23 subjects<30 years 15 subjects>60 years	19 ADL 15 Falls
PAMAP2 [79]	[76]	3 A,G,M	Wrist, Chest Ankle	9Hz	9 subjects	18 ADL
UCI HAR [80]	[35, 75]	A,G	Waist	50 Hz	30 subjects [19-48 years]	12 ADL
USC-HAD [81]	[35]	A,G	Front Right Hips	100 Hz	14 subjects (5 sessions)	18 ADL
Opportunity [82]	[35]	7 A,G,M	Various Locations	30 Hz	4 subjects (5 sessions)	18 ADL
Daphnet FOG [83]	[35]	3 A	Shank, Thigh Waist	64 Hz	10 subjects With PD	2 ADL (Normal vs Freeze Gait)
Skoda [84]	[35]	A	Rigth Hand	98 Hz	1 subjects (19 sessions)	11 ADL

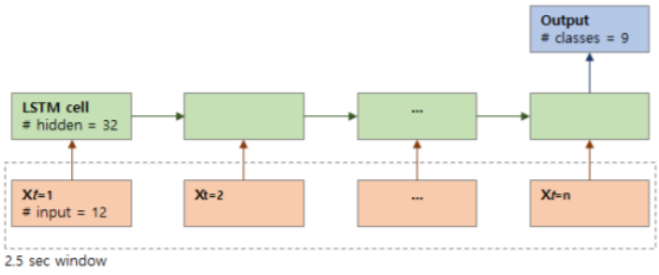
3.3.2 ADL Recognition Algorithms and Results

A slinding window technique was applied by Altuve et al. [75] to produce segments of 1,28s and 12.5% overlap. These were introduced to a first network to identify three dynamic activities and the static ones as a single class, with 98.86% accuracy. Then, the statics class data were introduced to a second network to classify three more ADLs, with an accuracy of 88.46%. In total, a 92.91% accuracy was obtained to classify the six human activities.

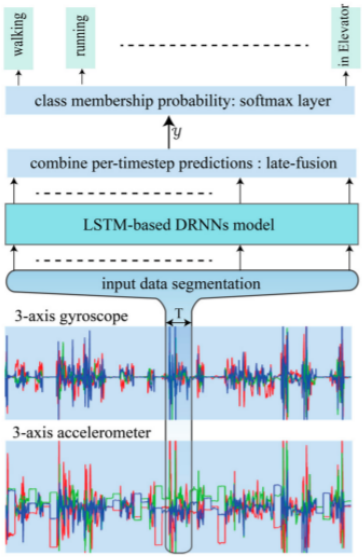
A two-fold cross-validation strategy was conducted by Wang et al. [67] on the SisFall dataset, which ensured that subjects data were always assessed with classifiers trained on different people, allowing for a realistic assessment. CNN, MLP, DAE and CAE were the neural networks explored in this work. Experiment results show that lightweight neural networks have obtained better results than machine learning methods previously used. The best result was attained with the supervised CNN, which achieved an accuracy beyond 99%. Sensitivity and specificity were other metrics used to evaluate the networks performance.

A Many-to-one LSTM network architecture (Figure 12) was used by Chung et al. [77] for activity recognition with nine classes. A 10-fold cross-validation was applied on the data to generalize the result to an independent dataset. Moreover data augmentation was used to perform LOO validation. From the experiment results, the authors found that activity data with sampling rate as low as 10 Hz from four sensors is sufficient for recognizing human daily activities with 93% of accuracy. Furthermore, classifier-level sensor fusion is a technique that can improve the classification performance.

In the work carried out by Murad et al. [35], unidirectional, bidirectional, and cascaded architectures based on long short-term memory (LSTM) DRNNs were created and their effectiveness was evaluated on several benchmark datasets (Figure 12a), where each one was split in two subsets, using 80% of the data for training and 20% for testing. Experimental results show that the proposed models outperformed conventional machine learning methods, as well as other deep learning techniques, such as CNNs. Overall accuracies over 92% were attained for every dataset, which proved the efficiency of the created models for a broad range of activity recognition tasks.



a Many-to-one Long LSTM architecture used in [77] for activity classification.



b The proposed HAR architecture in [35].

Figure 12: Exemplifications of how to use LSTM to recognize daily activities. Taken from [77] and [35].

The first phase from the 3-part modular system proposed by Gil-Martín et al. [76] segments the acceleration signals into overlapped windows, extracting information in the frequency domain. The second module detects the performed activity at each window using CNN architectures. The third module integrates the window-level decision in longer periods of time. A LOO cross validation was used, which aims to increase generalization across subjects who are not present in the training set. When applying post-processing techniques in this work over the score vector sequence, isolated classification errors were reduced, improving the accuracy from 89.83% to 96.62%.

Table 7 presents a summary of the approaches employed by the several articles which investigated the detection of falls and daily activities using algorithms based on deep learning algorithms.

Table 7: Validation characteristics adopted by the 5 selected articles who perform DL-based algorithms for fall and activity recognition, where: Acc = accuracy, Sens = sensitivity, Spec = specificity, LOO = Leave-One-Out, CV = cross-validation, CNN = Convolutional Neural Networks, MLP = multi-layer perceptron, DAE = dense autoencoder, CAE = convolutional autoencoder and LSTM = Long Short-Term Memory Neural Networks

Work	**Model Used**	**Validation Method**	**Type of Classification**	**Results**
Altuve [75]	Bidirectional LSTM	N\A	4 ADL Recognition	Acc: 92.91%
Wang [67]	CNN, MLP, DAE and CAE	2-fold CV	FallL Detection	Acc: >99%
Chung [77]	Many-to-One LSTM	10-fold CV LOO CV	9 ADL recognition	Acc: 93%
Murad [35]	Uni, Bidirectional and cascade LSTM	Hold-out CV	ADL recognition Several Datasets	Acc: >92% For all datasets
Gíl-Martin [76]	CNN	LOO CV	12 ADL recognition	Acc: 89.83%; Acc: 96.62 (w/ post-processing)

3.4 Discussion

Several papers that use Machine and Deep Learning approaches were studied in the previous Sections for three essential phases in automatic strategies for classifying ADLs and a wealth of information was found. As in the analysis of the FRA works, it is confirmed the existence of some common points in the various studies analyzed, however, there is still a great variability of techniques that are used in each of the phases of ADL recognition. Despite the positive results, there are still a few areas that should be addressed.

In a similar way to what was examined in the literature on automatic FRA approaches (Chapter 2), accelerometers and gyroscopes are the most commonly used sensors for data collection in AI-based approaches for ADL recognition (other sensing technologies, such as pressure and EMG sensors are also used). The locations also vary between the subjects' waist, chest, leg and even wrists, although the first is the most common within the analyzed articles. It should be highlighted that in the series of studies that use deep learning algorithms to recognize activities, a larger emphasis is placed on using benchmark datasets such as *SisFall* or *UCI HAR* rather than developing their own data collection protocols.

In general, both the benchmark datasets and the data acquisition protocols are carried out in the same way: data is collected from several people as they do multiple trials of a series of activities in isolated or circuit format. The most commonly addressed activities in these works are basic ADLs related to the subjects' locomotion. Dynamic locomotion activities, such as walking, walking up and down stairs, and static activities, such as standing, sitting, or lying, have received the most attention, while transitions between activities are typically ignored in activity recognition studies.

In terms of the algorithms used in the activity recognition process, signal segmentation is generally done using sliding window approaches, with windows of 1 to 2 seconds being employed in most cases, with overlap between them. The feature extraction method in Deep Learning is either non-existent or embedded in the neural network designs employed. In general, the Machine Learning-based methods most used when it comes to classifying ADLs are KNN, SVM and DTs, although others such as Threshold methods are also used. LSTM, CNN, or hybrid architectures of CNN and LSTM are the classification models used for classifying daily activities.

Regarding the algorithms' validation, techniques like k-folds cross validation and Leave One Subject Out are commonly used to validate these procedures. In addition, the evaluation of models in benchmark datasets is commonly used for study validation and comparison [35]. However, although some works include validation techniques, others (particularly those relying on machine learning) do not refer to the way they validated their algorithms [70, 72]. The models' performance in both studied approaches is measured using metrics like accuracy, which is the most often used, as well as others like sensitivity, specificity, and the F-1 Score, which provide useful information in problems involving several classes or activities to be classified.

For the most part, the results achieved by Machine and Deep Learning-based algorithms for ADL recognition are highly satisfactory, with several works presenting accuracies over 90%. However, in most situations for both ML and DL-based approaches, only a small number of activities are detected, usually cyclical or static ADLs, like walking or sitting, leaving key activities such as transitions between body positions unexplained [71, 74, 75]. Other limitations identified in this research are related to the quality and size of the datasets used, where the amount of data used, whether through data acquisition or through public datasets, is very small and and little varied [35, 77]. This fact may lead to models that may present

high performances only on the dataset used, requiring training with a wider dataset to validate said performance [69]. As a result, there is a greater demand for large-scale data collection campaigns. Studies in the field of merging datasets and pre-processing pipelines are also required in order to properly combine and decrease disparities between data obtained from various sources [68]. According to the research conducted in this dissertation, no study has attempted to perform this method. Finally, another drawback is that the most significant information discovered by feature selection methods is not revealed in the works that use this procedure before their classification algorithms. Furthermore, when this process is not employed, all of the extracted features are directly used in the algorithms, increasing the computational cost of the developed systems.

Chapter 4

Fall Risk Assessment Tool - Architecture Conceptualization

Currently, as demonstrated by the research carried out (Chapter 2), very few studies direct their efforts towards a multifactorial FRA, with usually only one type of data being used for its analysis (either inertial data or clinical test's results data). Additionally and although some studies have already shown reasonable results in this area, continuous FRA is still a process that can be improved. Another issue that should be addressed is the usability of today's methods and hardware. These should evolve to be more readily and autonomously used by the elderly, i.e., in a free-living environment rather than under controlled conditions, as well as to reduce the quantity and size of the sensors employed, making the experience as comfortable as possible for the users. Finally, there is a need for improvements regarding the algorithms used for FRA, since many of the algorithms analyzed in Chapter 2 did not have a validation method defined or it is not well discriminated, especially in the immediate FRA and multifatorial FRA groups.

Thus, there is a need to implement a method capable of address these FRA drawbacks, providing a viable alternative to enable the daily use of FRA tools by the elderly. This tool is described in this chapter. Figure 13 demonstrates a FRA engine proposed by Danielsen et al. [22], which prompted the conceptualization of the FRA tool described in this chapter. These authors identified what kind of information is most relevant for a multifactorial FRA using wearable and other types of sensor in a daily living environment, idealizing a FRA protocol based on these findings.

All the steps to develop a **eHealth platform**-based FRA strategy in a continuous and multifactorial manner, while assessing both prospective and immediate approaches for FRA, will be discussed in the following sections. The prospective approach will be evaluated from clinical and demographic information acquired from questionnaires and medical appointments, while immediate FRA will be carried through

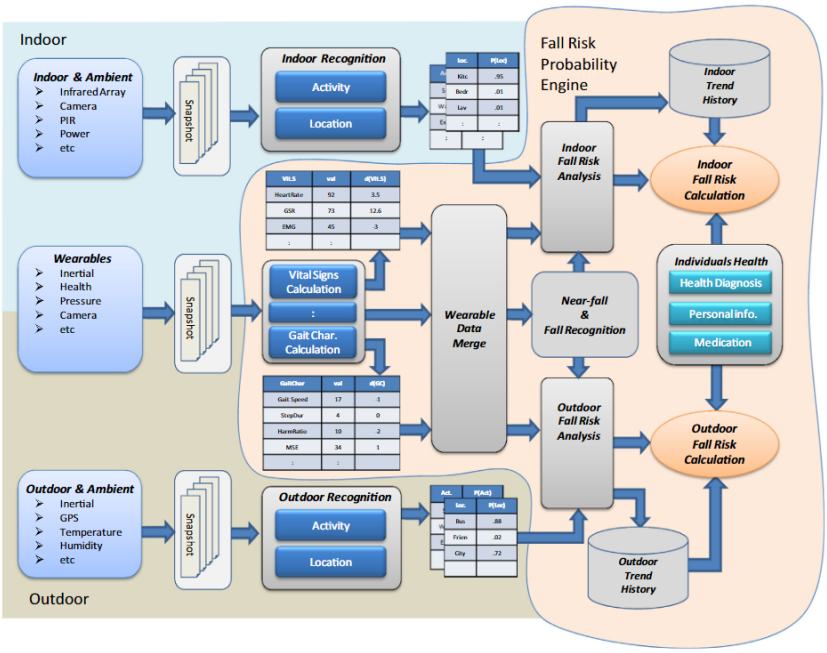


Figure 13: Calculation of fall risk using the fall risk probability engine proposed in [22].

activity recognition and detection of gait abnormalities in the elderly. These two perspectives will later be joined to give rise to an overall fall risk value. First, the main requirements of this tool will be raised (Section 4.1), followed by a data flow overview of the proposed architecture (4.2). The necessary data to carry out the FRA will be addressed in Section 4.3. Sections 4.4, 4.5, 4.6 and 4.7 approach the operation of the idealized tool's various FRA and feedback stages. Section 4.9 concludes the Chapter with a discussion regarding the idealized FRA tool.

4.1 Requirements

Based on the literature analyzed in Chapter 2, and given that the goal of this project is to develop a competitive FRA tool that solves the problems from other existing FRA tools, it was possible to gather a list of requirements that must be met. The elderly are the primary target demographic for this technology, as they are the population most at risk of falling. The tool should also be accessible to health care providers and nursing home workers so that they can provide prompt support when needed [85].

FRA is currently carried out mainly from tests and medical assessments, which only give an idea of the risk of falling relative to the moment and conditions of the referred test (Chapters 1 and 2). FRA strategies that encompass other types of analysis, usually use Machine Learning, and only one type of data – inertial – to provide a binary classification (Faller vs Non-Faller). As a result, one of the most important requirements to meet is to ensure that this tool is multifactorial, which means that it must assess risk using a variety of factors, including measurements from wearable sensors, clinical test results, physical or

psychological information from users, as well as information about the external environment. Multifactorial FRA is more advantageous since it allows for a better understanding of the various existing risk factors (discussed in Chapter 2), whether they are related to medical, psychological, and behavioral disorders in the elderly or to environmental risks.

The strategy also needs to provide a valid fall risk result, in a scale or code easy to perceive by all users, and in a continuous manner, in order to be reliable and easily available to any user (i.e., be user-friendly and subject-independent). It must also be a tool that can be used with a variety of data-gathering devices, such as instrumented waistbands, smart watches, canes, and walkers. This requires the use of tools such as eHealth platforms that enable the display of the risk level in an easily perceptible manner by the target population, as well as the establishment of communication between the aforementioned data collection and risk analysis devices and other gadgets used by caregivers.

Finally, the target audience and the time of use must be considered in this strategy. As a result, it must be constructed in such a way that it can compute a fall risk value with the fewest number of sensors possible, avoiding the use of cumbersome measuring devices and making the use of associated devices more comfortable, thereby avoiding discomfort that could lead to an increase in the risk of falling.

Thus, the strategy outlined in this dissertation has the following objectives: **i)** the collection of multifactorial data, i.e., from several different sources, such as inertial data, clinical tests and demographic data of the elderly, as well as data related to the environmental context that the elderly face; **ii)** Continuously compute a single fall risk value from the aforementioned data, taking into account both the immediate and prospective FRA perspectives and the use of AI-based methods; and **iii)** provide timely feedback in the event of a fall or near-fall situations such as slipping or other loss of balance.

4.2 Architecture Overview

Nowadays, with the electronic development of wearable sensors, and even the emergence of AI technologies capable of analyzing and learning the meaning of data coming from sensors, strategies for automatic fall risk analysis tools have come to be increasingly studied [17, 22, 28]. In general, they present an architecture divided into 3 distinct main phases of operation: **i)** the data acquisition and processing phase; **ii)** the signal analysis phase and the final decision; and **iii)** communication phase [86]. This division is also followed in the several works which have a similar objective, mentioned in Chapter 2, to provide fall risk value from multiple data sources [22, 62, 63]. The strategy that is presented in this chapter follows the type of architecture that is usually followed by the literature, and it is summarized in Figure 14.

The first block of the proposed tool (Data Acquisition and Processing) represents the different proposed data acquisition and processing techniques. In Figure 14, the second block of the presented strategy (Classification and Regression Models) is built in a modular way, with the operation of 3 classification and/or regression modules in parallel. Based on the results obtained in these 3 modules, in the third and

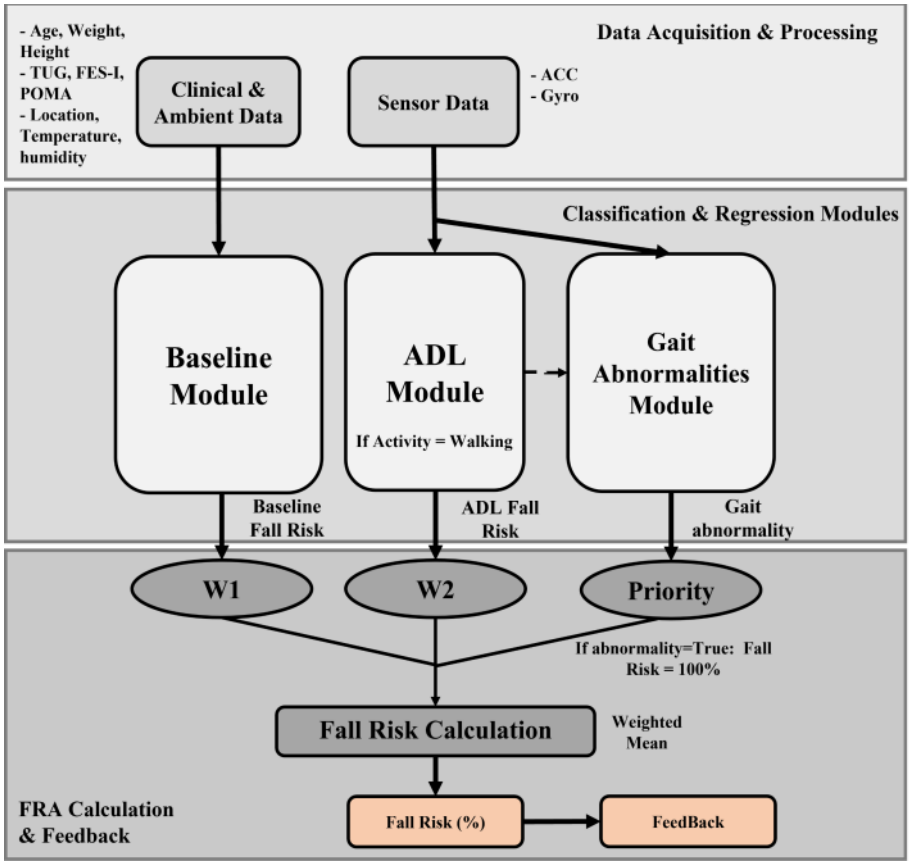


Figure 14: Proposed FRA tool architecture overview. The 3 main phases (Data acquisition & Processing, Classification & Regression, and FRA Calculation & Feedback) should be noted in the different gray areas, as well as the 3 parallel modules in the *Classification & Regression* Modules phase.

final phase (FRA Calculation and Feedback), the fall risk will be calculated and due feedback should be provided in a timely manner to the subject, family members or health professionals. An **eHealth platform** would be in charge of controlling and connecting the various blocks of the proposed strategy. This platform would also have the role of displaying in real-time the fall risk level of the users, and it would be installed on devices such as smartwatches (Figure 15), smartphones, or tablets, which can be used by both the elderly and caregivers.

In the next sections, the goals, data workflow and expected outcomes of each block will be explained, referencing the type of data to be used and the type of processing and mathematical/decision processes that will lead to a fall risk result. In Section 4.9, a discussion is made regarding the adequacy of the proposed architecture in relation to the objectives of filling the current gaps in the literature, still giving some points for improvement and future work.

Figure 15: Smartwach eHealth app example for continuous fall risk assessment.

4.3 Data Acquisition

One of the requirements of the tool presented in the previous section, is to carry out a FRA using several different fall risk sources, in order to have a better idea of all the risk factors that can influence and give rise to a fall event.

As verified in the literature review, to perform FRA automatically, wearable sensors are the most common for data acquisition, such as inertial sensors, usually located on the subjects' waist [17, 22, 27]. To analyze the kinematics of the subjects' movement, inertial data of acceleration and angular velocity will be collected from a waistband placed on the subjects' lower trunk. Thus, it will be possible to obtain information regarding the movement and behavior of the subjects throughout the day. This data will help in monitoring the movements of the elderly in real time and identify potential risk circumstances linked with their behavior, such as doing high-risk activities (e.g., climbing stairs) or situations of lack of balance, in the immediate future. In order to have access to information regarding the clinical situation of each user of this tool, the acquisitions of this type of data should be made from questionnaires, clinical tests and medical consultations, in order to be assessed as needed to perform a FRA that is more focused on the long-term future (prospective), knowing a subject's current physical and psychic skills and what this implies for their long-term fall risk. It is also suggested in this tool the use of temperature, humidity and light intensity sensors to acquire data related to the environment that surrounds the subjects, thus covering some of the risk factors associated with external parameters or ambient hazards as described in Chapter 2, and not related with the individual. These 3 acquisition modes should generate a wide range of data from different sources, thus allowing for a multifactorial fall risk analysis.

To make this FRA tool as comfortable as possible, this acquisition must be made by coupling the aforementioned sensors to devices already used in everyday life by the elderly (waistbands, canes, walkers, etc.), and other small devices used by both the elderly and health professionals (smartphones, tablets,

smartwatches, etc.), which allow for storage and embedded data processing and risk calculation, making this tool easy to use and comfortable for all users, whether they are an elderly subject or a caregiver. These devices would be connected by the aforementioned **eHealth platform**, which would be in charge of the communication between the devices used by the elderly which perform the FRA, and the ones used by the caregivers, in order to have an efficient feedback as soon as the devices used by the elderly detect a high risk of falling. Also, it would store data received from the elderly and making it available to caregivers, as well as the various results obtained from the FRA algorithms in a continuous manner.

4.4 Baseline Risk Module

As noted in Chapter 1 and Section 2.1, the major goal of the first module is to offer a baseline fall risk value based on patients' clinical (physical and psychological) and demographic data, test results related with medical fall risk scales, or even external factors such as location and weather conditions. Some of the factors intended to be used in this module can be seen in Table 8.

Table 8: List of several Baseline risk factors which can be used in the Baseline module of the FRA tool

Risk Factor Group	Factor	Value
Demographic	Age	Numeric
	Weight	
	Height	
Clinical	TUG	Numeric/Binary
	POMA	
	FES-I	
Psychological	Fall History	Binary
	Other Conditions	
	Depression	
Ambiental	Location	Numeric
	Temperature	
	Humidity	

The first phase of this block is the acquisition and storage of data, where, using simple clinical tests and questionnaires, most of the data relating to the subjects can be acquired. Data related to the environment and the location of individuals can be acquired using temperature, humidity or GPS sensors, which can be coupled to support devices used daily by the elderly, such as smartwatches, waistbands, canes or walkers. This data can be stored in databases or in electronic devices such as smartphone apps in order to be used later in the second phase of this FRA tool.

As can be seen in the second column of Table 8, there is a wide variety of indicators of the physical and psychological state of an elderly person, as well as the characteristics related to the environment that

surrounds them. Therefore, there is a need of a second stage of this module, to convert and normalize the data from the diverse types into numeric or binary values, so that it can be interpreted by mathematical models. Taking into account the aforementioned different types and magnitudes of data and its conversion, the third and final step of this module should be the calculation of the fall risk value from the converted and normalized data. Following the line of thought found in the literature, this value should be calculated using regression methods, such as Logistic Regression [63, 87, 88].

The result obtained should remain constant, constituting a baseline fall risk value appropriate to the characteristics of each individual and to the environmental conditions to which each one is subject at a given moment, until some of the parameters undergo changes, such as changes in the user's weight, new clinical tests results or even changes on the location or temperature during the day. Thus, this block serves the purpose of presenting a fall risk linked with a prospective approach of FRA, evaluating the likelihood of an aged person falling in the long term, given their skills. Figure 16 summarizes the data flow and the aforementioned steps for the correct functioning of this module.

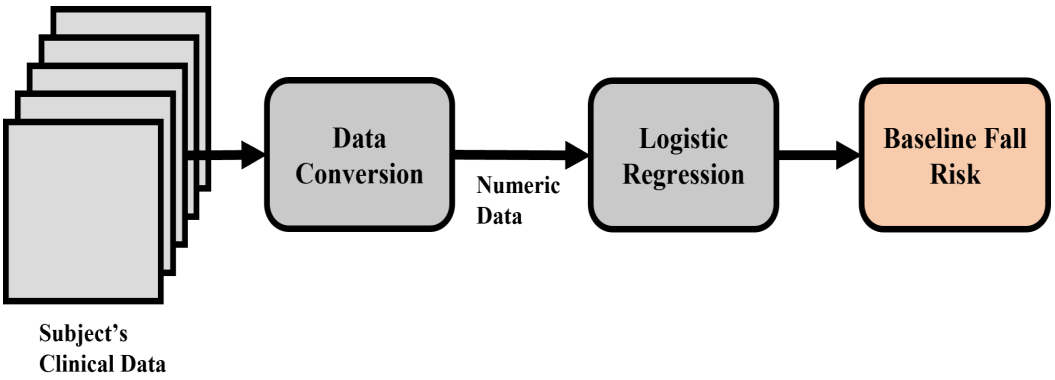


Figure 16: Baseline Fall Risk module Data flow overview.

4.5 Activity Recognition Module

As mentioned in chapter 2.1, it was shown that the increase in the probability of occurrence of a fall may be, alongside with other factors, associated with certain activities. Also, [32] reports that the effective number of falls reported per location and activity also differs, showing that the issue of daily activities recognition may be a central point of study in order to improve FRA tools.

With that scope in mind, an ADL classification module will be in constant operation, which will allow the decoding of what activities the subjects is practicing throughout the day. This activity recognition module, which will from this point on be referred as *ADL Module*, is inserted in the second phase of the proposed

FRA tool and it will be the central focus point of this dissertation, given that it is the least worked point in the literature regarding the integration of the recognition of daily activities with the risk of falling, and it is one of the first points that need to be addressed.

As mentioned in Chapters 1, 2 and 3, for the automatic assessment of the risk of falling in the elderly, as well as for the recognition of daily activities, the most used type of sensing in recent years it has been wearable sensors. The most commonly used data types for activity recognition are inertial, acceleration (from accelerometers) and angular velocity data (from gyroscopes). Thus, in a first phase of this block, it is intended that a waistband equipped with these inertial sensors makes a continuous collection of acceleration and angular velocity data from the lower trunk of the subjects as they freely carry out their daily activities. The second phase should process the inertial data obtained from wearable sensors in the first phase, computing many features that would then be employed with AI-based models to identify several postural ADLs. Fall events, as well as their direction (forward, backward, or sideways), must be identified in addition to the daily activities that this module must recognize.

After recognizing which activity is being carried out at a given moment, this will be used in a next step, where, alongside with some of the demographic data from the *Baseline module*, a fall risk level will be assigned to each activity, based on the results found in [32], regarding the probability of falling according to the activity and place of the fall, as well as according to gender and age. A summary of the data flow of these modules is presented in Figure 17.

The *ADL recognition* module of the FRA strategy proves to be the centerpiece of the tool and what will bring advances to research on this topic, since, to the best of our knowledge, no works were found whose techniques make use of the recognition of ADLs in the fall risk analysis. Despite the potential displayed by this module, it is necessary to validate the data collection methods, and which are the best features and classification models to use, so that the recognition of ADLs can be made with the shortest possible delay between the start of the activity and the classification and with low classification error rates.

In addition to this aspect, and although there are already studies that show that the probability of a fall varies with the activity [32], it depends on the combination of several other factors, making further research on this topic necessary, with extensive prospective work and large datasets of elderly subjects, their falls and its causes, in order to make a statistical association between a daily activity and the respective risk of falling.

4.6 Gait Abnormalities Module

The second phase of this FRA tool presents a final classification module, which will be referred to as *Gait Abnormalities* Module, with the aim of detecting gait abnormalities in a timely manner. The analysis of the quality of human gait has been one of the most discussed topics in the analysis of the risk of falling [56, 57, 89], therefore, a thorough and real-time analysis of the user's gait is of great importance to determine

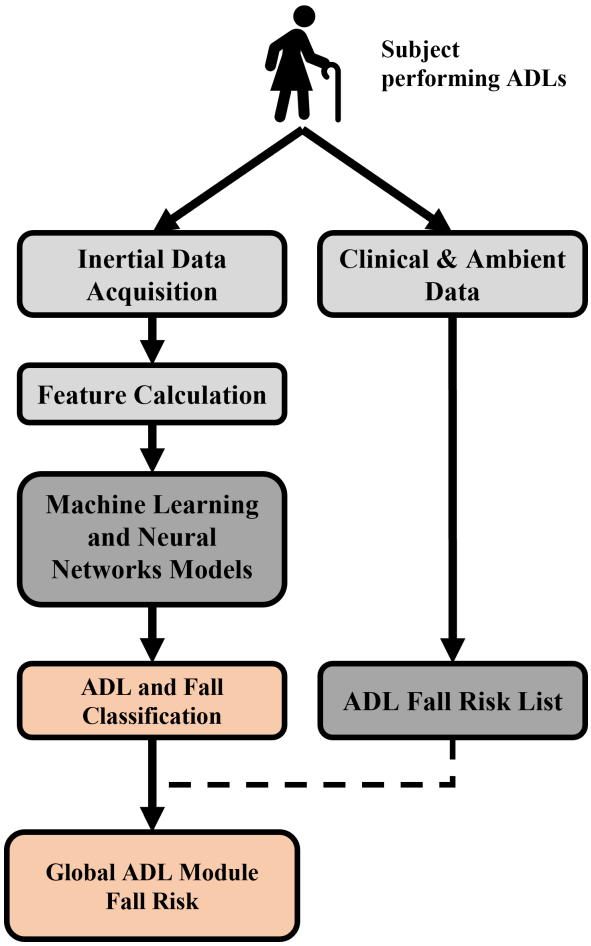


Figure 17: ADL recognition module data flow.

the fall risk of an individual. The work flow of this module will be addressed in the following paragraphs, and a quick summary of its functioning is shown on Figure 18. *Gait abnormalities* module's operation will take place according to the results attained from *ADL Module*. Thus, this module will be activated when the *ADL Module* classifies a walking activity, being only in standby while other activities are being detected. The general functioning of this *Gait abnormalities* module for the detection of gait abnormalities will be very similar to the workflow presented in the *ADL Module*.

Once again, the waistband equipped with the inertial sensors would be used to make a continuous collection of acceleration and angular velocity data from the lower trunk of the subjects. As in the previous module, AI-based methods will be applied for the analysis of the users' gait, and it is expected that these models can make a binary classification of the quality of the gait. It should be classified as normal, in case the subject is walking within a certain range of normality, or abnormal, in case the user has a walking pattern related to some disability or momentary lack of balance, such as a slip or trip [56, 89, 90].

In case the gait is considered normal, this module is not activated and keeps its collection and classification of the gait until the activity to be performed by the subject changes. In this case, the risk level

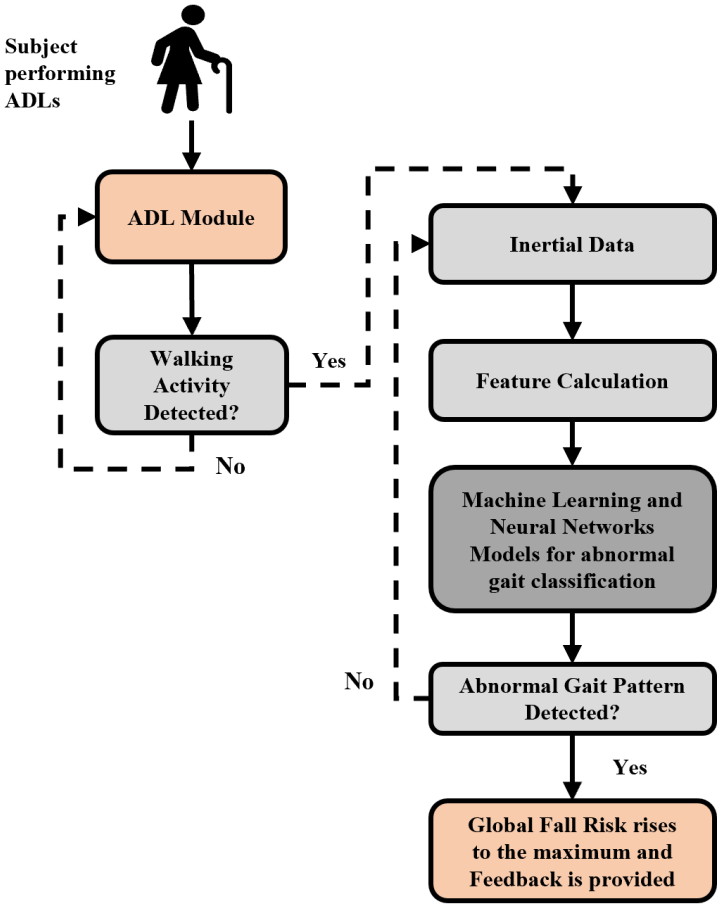


Figure 18: Gait abnormalities module data flow.

calculation is provided by the modules *Baseline* and *ADL* (See Section 4.7). On the other hand, in case the subject's gait is classified as abnormal, this block will override the aforementioned risk level calculation, raising it to a maximum value, and activating feedback processes, whether for the subject himself or for family members or health service providers. The **eHealth platform** would be responsible for the communication between the classification algorithms (and its results) and the devices used by the caregivers.

Finally, it should be noted that the presented choice of processes was due to the use of the smallest possible number of sensors and devices for FRA, allowing risk analysis using only a waistband, however, this block should be improved in the future to not only make a binary classification but a deeper analysis, for example, to the location of the center of mass (COM) and pressure (COP), perhaps with the addition of the use of instrumented shoes [57, 72, 91], or other devices and wearable sensors.

4.7 Fall Risk Assessment and Feedback

The last phase of this architecture has the main function of joining the results obtained in the 3 modules of the previous phase and calculate a fall risk level within a fall risk scale. Subsequently, and taking into account the risk value obtained, due feedback should also be provided to the users, as well as to health professionals, in case the fall risk is considered too high or a fall event is detected (Figure 14).

For the presented tool, the strategies proposed in [62] and [63] for the determination of a fall risk level served as a reference, where weights were assigned to each of the existing blocks for the analysis of fall risk. Thus, a weight of 60% (W1) was assigned to the Baseline module, and a weight of 40% (W2) to the ADL recognition module, which assesses the risk of falling according to the activity being classified. The *Gait abnormalities module* is not included in the fall risk calculations, however, as stated before, it is placed as a priority module, so that when an abnormal gait is detected, the module is prioritized above the others, causing the fall risk level to rise to its maximum.

Thus, in a continuous manner, and from the results of the *Baseline* and *ADL recognition* modules that are used to calculate the risk of falling taking into account both prospective and immediate FRA approaches, respectively, the global fall risk value will be computed as a function of the following Equation:

$$\text{Fall Risk} = (\text{Baseline Fall Risk} \times W1) + (\text{ADL Fall Risk} \times W2) \quad (1)$$

This equation represents the weighted average (*Fall Risk*) between the Baseline (*Baseline Fall Risk*) and ADL (*ADL Fall Risk*) modules, taking into account the percentage weights *W1* and *W2* assigned to these blocks, respectively. It should be remembered that the detection of any imbalance during the subjects' gait phase is not included in the calculation, however, as mentioned in Section 4.6 this block overrides this calculation in case any normality is detected, raising the risk of falling to its maximum.

If a fall is detected by the *ADL Module*, or if any gait abnormality is detected by the *Gait abnormalities* module, or even if the risk level calculated by Equation 1 exceeds a given threshold, a set of feedback processes must be triggered, which are exemplified in Figure 19. Devices such as smartphones or smartwatches or other instrumented support devices can automatically communicate with other devices. Thus, as can be seen in Figure 19, in the event of a fall or a high fall risk situation, these devices must be activated to provide feedback in a timely manner in order to prevent falls or minimize its consequences. Feedback must inform the subject of his current fall risk at all times and give warning signals in the event of a high risk situation (such as a walking in an area with low light, loss of balance, etc), for example through graphic or numerical scales, color codes or other sensory warnings, such as vibrations or warning sounds (Figure 19), light gray area. In the same situations, the referred devices must be able to come into contact with family members and health professionals so that the necessary assistance is provided as soon as possible, in order to avoid or reduce the possible consequences of falls [22]. This feedback should be

done through automatic messages, email, or even through notifications in possible apps for smartphones or other electronic devices, as shown in the dark gray area of Figure 19.

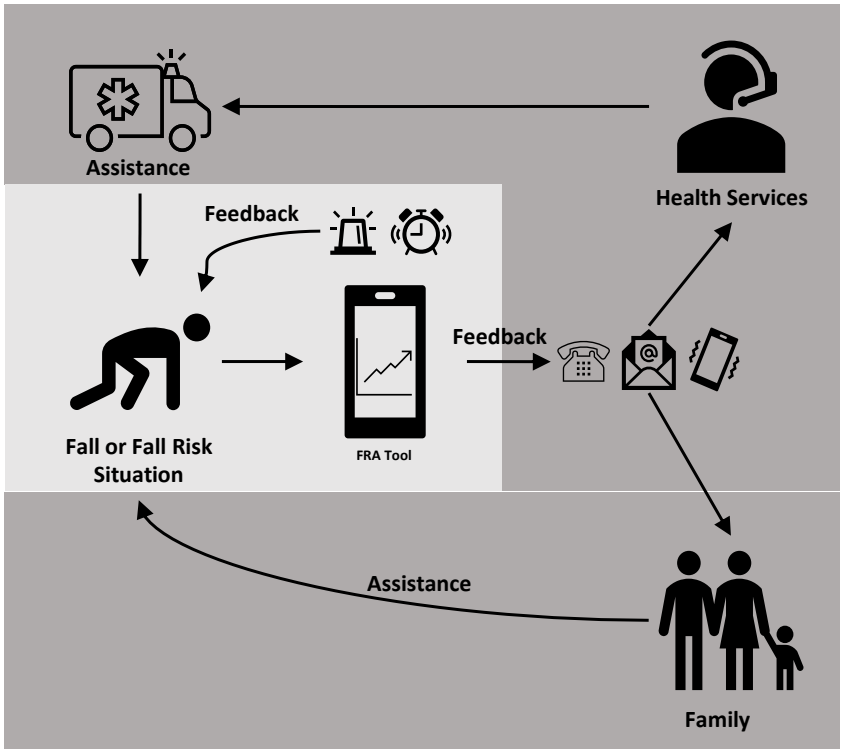


Figure 19: Warning signals and feedback processes implemented in case of a fall event or a high fall risk situation.

4.8 Data Acquisition in Nursing Homes

The first steps towards the implementation of the previously envisioned FRA strategy were taken. An initial version of the data acquisition stage (Section 4.3) was implemented offline, with the development of a waistband instrumented with an inertial sensor, as well as smartphone applications to be used in inertial data collection sessions with elderly subjects at nursing homes. This proof of concept is an essential first step for the development of the remaining blocks of the FRA tool, since this block is shared by all fall risk assessment modules.

Thus, this section will address the process of enhancement of a previous team-own instrumented waistband and smartphone applications for data acquisition, as well as an ADL inertial data acquisition protocol, which was carried out at a partner nursing home, the *Fundo Social de Braga*. The materials used to the construction of the waistband will be identified (Subsection 4.8.1). An overview of how the collection protocol took place will also be addressed (Subsection 4.8.2).

4.8.1 Waistband enhacement

The waistband used to gather data in nursing homes was based on a version previously developed by *BirdLab* members, and instrumented with several electronic components in order to collect gait data from parkinsonic subjects [90]. For the enhancement of the initial waistband version, for waist-located inertial data acquisition from ADL performed by elderly people in nursing homes, the previous sensing system (MPU-6050) was changed for an inertial sensor (**LSM6DS0X**), miniaturized (**15x20x2 mm**) and light (**10 g**), integrating a 3-axis MEMS **accelerometer** (±16 g) and **gyroscope** (±2000 °/s). An ultra-low-power high-performance three-axis magnetic sensor (**LIS3MDL**) was also added. As can be seen from Figure 20, the sensing system is located in a central region of the outer part of the waistband, connected by a cable to the remaining components.

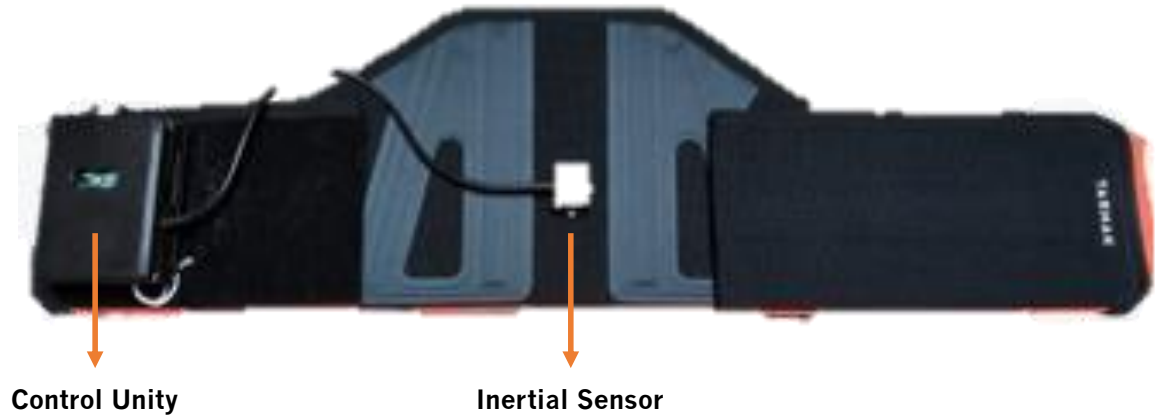


Figure 20: The instrumented waistband developed for data collection in nursing homes.

Regarding the remaining components connected to the sensing unit, a control unit, which includes the **STM32F4 Discovery** as a processor, allows for the communication between all the components, sending the necessary commands for the sensory units to start and stop the data collection, and saving the collected data in a file once the acquisition is finished. The inertial acquisition takes place at a sampling frequency of **100 Hz**, sufficient for the acquisition of human movement, without loss of information [77]. The files containing the inertial data acquired by the sensory unit are stored on a **USB pen drive** with a capacity of 32G. To power the system, the **power-bank Turbo 3000** (5 V; 2 A; 3000 mAh; 75 g) is used. The system has an autonomy of eight hours to enable daily monitoring sessions for long and continuous periods. Furthermore, a new Bluetooth module (**HC06 Bluetooth module**) was also included

to allow a better connection between the waistband and external devices, such as smartphones. The last improvement implemented in this waistband was related to its fabric, which was modified in order to be more easily adjustable to the subjects' abdominal circumference, since the previous version of the waistband, despite being elastic, was one-size-fits-all and could cause discomfort. All the referred components are depicted in the diagram presented in Figure 21.

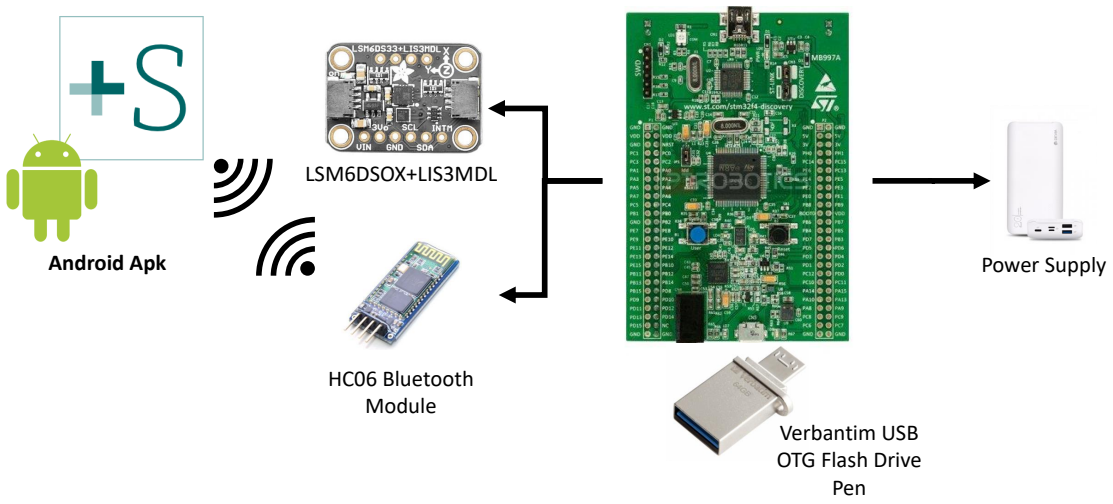
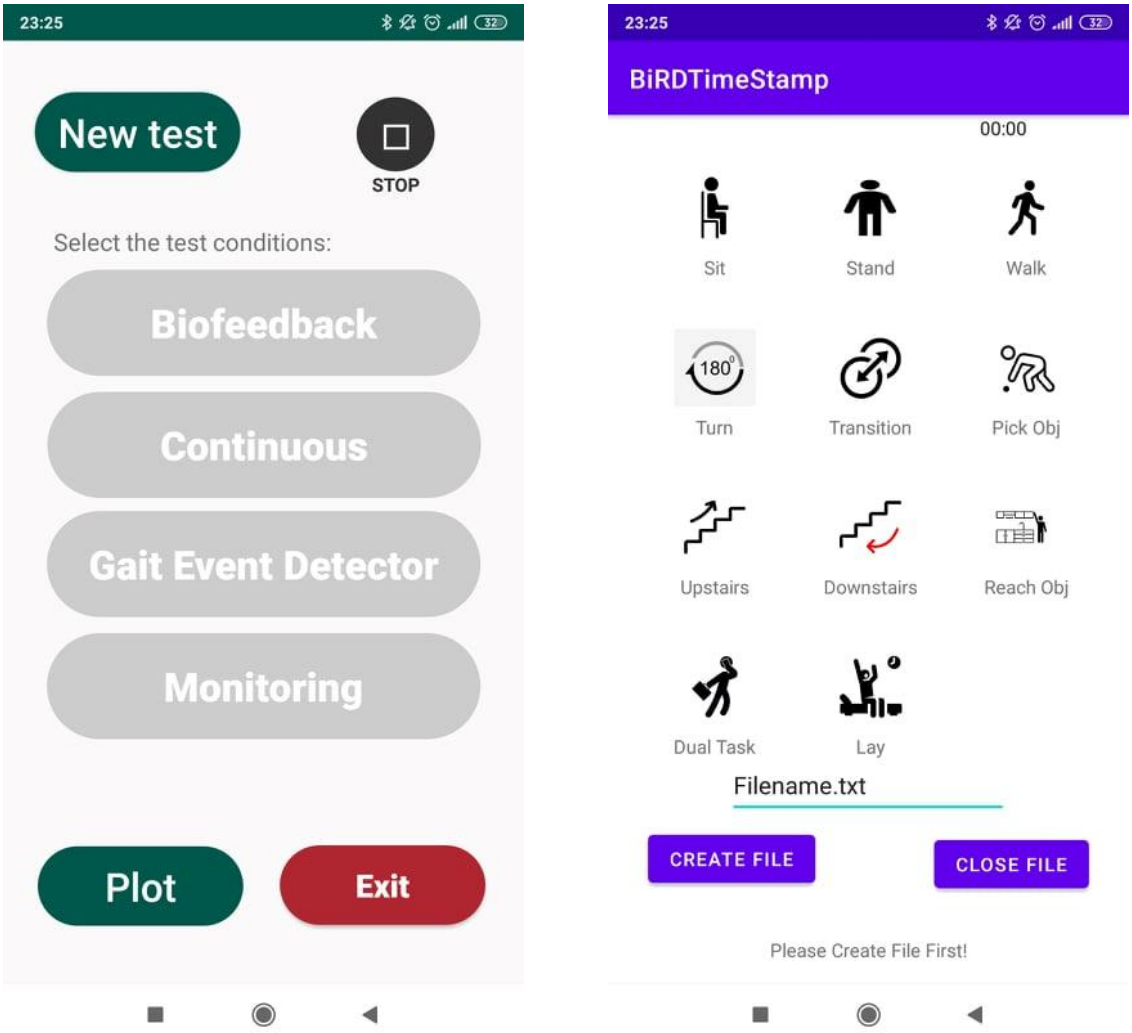


Figure 21: The waistband component's system overview, illustrating the central systems with the respective components and interfaces between them.

In order to configure a wireless data collection by the device, and to facilitate the data acquisition and subsequent processing regarding the labeling of the acquired data, **two graphic applications** for smartphones, which communicate with the control unit via Bluetooth, were created using *Kotlin* code in *Android Studio* (Figure 22). The first smartphone application (22a), developed in an Android environment firstly for the previous version of the waistband, was altered and allows the intuitive configuration of the beginning and end of the acquisition, as well as saving the file containing the inertial data from ADLs. The second application (Figure 22b), also developed in an Android environment, and collects the start and end times of the various activities performed during the acquisition, through simple clicks, saving them in a file. Moreover, it has an integrated timer so that the person responsible for data collection can time the duration of each collection

4.8.2 Data Acquisition Protocol

The data acquisition experimental protocol was elaborated in collaboration with several personnel specialized in the field of geriatrics, such as Doctor Gorjão Clara, and Marta Sousa, administrative and caregiver of the *Fundo Social de Braga* nursing home, in order to define several inclusion and exclusion



(a) Data acquisition App

(b) Data labeling App

Figure 22: Two android applications developed for the data acquisition in nursing homes. (a) Application which controls the beginning and end of the data collection, as well as saving the data acquired by the waistband in a file. (b) Application to perform the timestamp annotations of the beginning and end of each activity during the acquisition protocol, for an accurate labeling process.

criteria, the ADL to be realized by the selected subjects, and the conditions under which data acquisition must be done. Based on the exclusion criteria outlined, seniors with dementia (MMSE≤15); severe depression (EDG ≥ 11); physical disabilities (unable to walk without walking aid, FAC < 4); orthopedic, cardiac, or respiratory diseases that affect locomotion; Morbid obesity (BMI ≥ 30); and Fear of falling moderate-high (Short FES-I ≥ 14) were excluded. Table 9 shows the activities that were chosen with the target audience (elderly) and their physical constraints in mind. Finally, these activities were carried out and the inertial data was gathered under controlled conditions to ensure the individuals' safety at all levels. During the realization of these activities, each subject was accompanied by caregivers and the person responsible for the data collecting.

Table 9: Set of ADL that elderly subjects performed under controlled trials

Code	Activity
ADL001	Walking
ADL002	Standing
ADL003	Walking upstairs
ADL004	Walking downstairs
ADL005	Sit in a regular height chair, wait a moment, and up
ADL006	Sitting a moment, lying, wait a moment, and sit again
ADL007	Being on one's back change to lateral position, wait a moment, and change to one's back (Lying)
ADL008	Standing, pick something from the ground, and getting up
ADL009	Gently jump without falling (trying to reach a high object)
ADL010	Get into and out of bed without sitting

Clinicians, nurses and caregivers of nursing homes performed the various tests and clinical scales mentioned above in order to identify possible participants for data collection. After getting the clinical fall risk assessment from the several above mentioned clinical scales, a total of 8 subjects were selected to perform the experimental protocol. Thus, demographic data (age, height, body mass and gender), of the 8 selected subjects were acquired, and the set of activities present in Table 9 was executed in the form of a circuit by the elderly volunteers whose age ranged from **63 to 95** years old (**80 ± 8.31 years**), with a body mass between **42.7 and 97.5 kg** (**72.52 ± 17.79 Kg**) and a height of **151 to 171 cm** (**158 ± 6.36 cm**). All participants provided their written consent and their demographic information is resumed in Table 10.

Table 10: Demographic information regarding the eight subjects who participated in the data acquisition.

Subject	Age	Height (cm)	Weight (kg)
1	78	164	78
2	79	155	97,5
3	77	155	81,6
4	63	163	82
5	84	151	60,6
6	82	171	86,9
7	82	153	42,7
8	95	156	50,9

After the demographic data gathering, the first step for inertial data acquisition was the placement of the waistband on the subjects. This waistband must be placed so that the inertial sensor is aligned with the spine in the central region of the subjects' waist, as shown in Figure 23. Then, as previously

mentioned, each subject completed the set of activities shown in Table 9 in the form of a circuit, which is described in Figure 24. The circuit was repeated 3 times at a comfortable speed for each participant. Each activity had a duration of at least 2 minutes, with the exception of the Sit-to-Stand, Stand-to-Sit and Lay down transitions. The circuit was the following: Sit – Sit to stand (not with a specific time) – Stand – walk – pick objects from the ground – Walk – climb/descend stairs – walk – reach high objects – walk (dual task: answering the telephone/talk with a person) - Stand – lay down. The circuit was performed on the opposite side as well. While the subject performed the circuit, the responsible for the test gave voice commands to the subjects for them to change their activities, marking as well the start time of each activity in the mobile application to record the timestamps. Furthermore, if any of the subjects showed signs of fatigue, the test was paused to allow them to recover.

Figure 23: Placement of the waistband and completion of the ADL circuit by one of the subjects involved in the data collection protocol.

4.9 Discussion

In this Chapter, an architecture for a simple FRA tool was idealized, considering some gaps found in the literature, namely reducing the subjectivity presented in tests and clinical scales, allowing FRA in a continuous manner, incorporating both prospective and immediate FRA approaches, and using of a wide

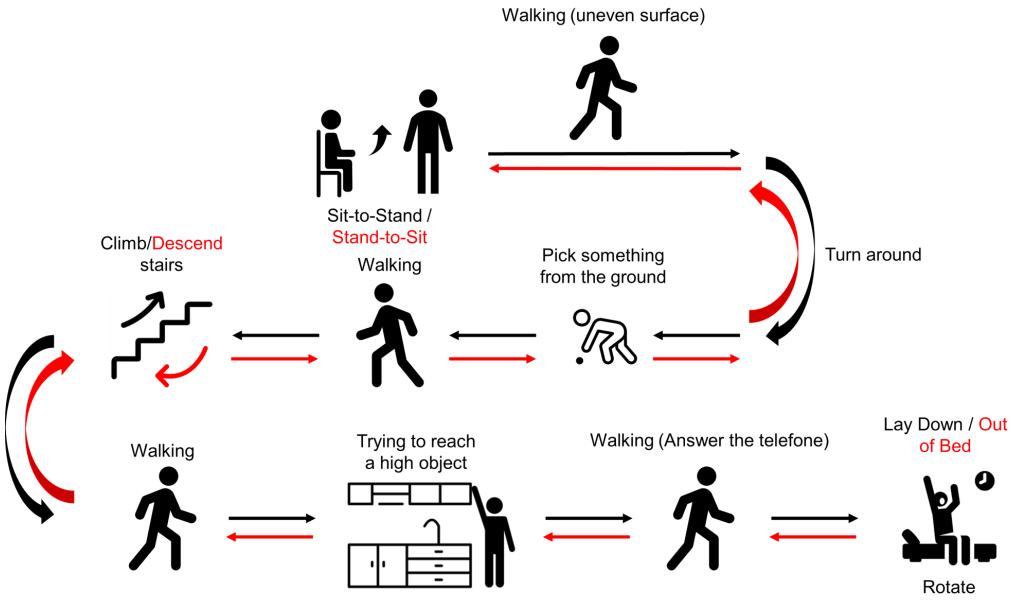


Figure 24: Circuits for inertial data collection of ADLs.

range of risk factors to carry out the assessment, while still being easy to use and comfortable for the target audience.

As initially mentioned in Chapter 1, this tool is based on recent developments in AI techniques and in data acquisition methods from wearable sensors. In this way, all modules can work using inertial data acquired by waistbands or smartwatches, thus making it an easy-to-use and comfortable tool for everyday use by the elderly. With the Logistic Regression model referred in Section 4.4, it should be possible to integrate and evaluate different types of risk factors, whether associated with the subjects, such as age, weight, fear of falling or results of clinical tests, or risk factors associated with the surrounding environment, such as location or humidity level. This module will thus make it possible to address the question regarding the need for a FRA tool that assesses the risk of falling in a multifactorial way. Another positive aspect that differentiates this FRA tool from all others presented in Chapter 2 is its ability to classify in real-time the ADL that is being carried out, and assign a risk level to from that classification.

Furthermore, the first step of this strategy, the acquisition of inertial data through devices, has been successfully initiated in institutions such as nursing homes. A team-own waistband previously equipped with an inertial sensor and a set of electronic components capable of acquiring inertial data was enhanced in order to be used for the acquisition of inertial data to feed the various classification modules of the FRA tool. An inertial data collection protocol was carried out in a nursing home using the developed waistband for collection, which allowed the acquisition of inertial data from 8 elderly people, while they practiced several ADLs in the format of an activity circuit. This data acquisition process proved to be quite positive, given the good general functioning of all components, without restricting the movements of the users of

the enhanced waistband, when performing the various tasks.

Despite the positive aspects mentioned above and the possibility of suppressing some disadvantages found in the literature, this tool is still require a validation of each module described above. As mentioned in Chapter 3, the validation of methods that make use of AI models is a complex process, which requires the use of a large set of data and several processes, in order to find the best method for a given task [69]. Taking these aspects into account, in the remaining Chapters of this dissertation, the first steps taken to validate the main module of this tool will be shown - the module that recognizes the daily activities performed by the subjects.

Chapter 5

Activities of Daily Living Recognition - Methods

From the analysis of the current literature regarding FRA (Chapter 2), it is clear that the use of ADL recognition in real-time FRA tools has been neglected. Therefore, to address this gap, a block was introduced in the proposed FRA strategy (Chapter 4) which included the automatic recognition of ADLs, as well as fall events, since the main objective is a complete FRA, preventing falls, but also giving the necessary feedback in a timely manner if any fall is detected. Furthermore, the review carried out in Chapter 3 revealed certain limitations in the various wearable sensor-based ADL recognition systems explored, such as the lack of large datasets and the lack of studies regarding which are the most relevant features for ADL recognition.

As a result, and taking into consideration the drawbacks observed in the literature, the methods used to validate the central block of ADL recognition indicated in the proposed FRA tool (Chapter 4) will be discussed in this Chapter: **i)** requirements gathering and data flow overview (Section 5.1); **ii)** dataset acquisition, with public and team-owned datasets fusion and normalization (Section 5.2); **iii)** comparative analysis with data segmentation and feature extraction and selection processes (Section 5.3); **iv)** discussion concerning the robustness of the implemented process (Section 5.4).

5.1 Requirements and Data Flow Overview

The first requirement of this module concerns the type of activities that must be classified. According to the studies mentioned in Chapter 2, the risk of falling shows to be related to the postural activity that the subjects are performing (walking, standing), or transitions between static postures (such as sitting or standing up). Thus, it is necessary for this block to recognize several cyclic locomotion activities, as well as static postures and the different transitions between them, in addition to the aforementioned falls [78].

Table 11 resumes the group of ADLs, postural transitions and fall events commonly found in the literature, as well as the label assigned to each one of the events mentioned (Chapter 3).

Table 11: Static postures and locomotion ADLs, postural transitions and Fall events selected to be recognised by the Machine and Deep Learning models.

Periodic Activities and Static Postures (Class Number)	**Transitions (Class Number)**	**Falls (Class Number)**
Walking (1)	Lying to Stand (13)	Forwards (9)
Standing (2)	Stand to Sit (14)	Backwards (10)
Sitting (3)	Sit to Stand (15)	Lateral (11)
Lying (4)	Stand to Pick to Stand (16)	Syncope (12)
Upstairs (5)	Stand to Lying (17)	
Downstairs (6)	Change Position (Lying) (18)	
Jumping (7)	Turning (19)	
Jogging (8)	Bending (20)	

According to the requirements of the FRA tool specified in Chapter 4, it must be user-friendly for the elderly, and activity recognition must be accomplished with as few sensors attached to the users as possible [22]. Thus, the ADL recognition in this Chapter was accomplished using only data from 1 inertial sensor (accelerometer and gyroscope), located at the waist, according to what is most frequently found in the literature (See Chapters 2 and 3).

ADL recognition should be as accurate as possible, with the literature showing performances of 96.70%, 96.83%, 96.83%, for accuracy, precision, and sensitivity, respectively, in 6 class classification problems using deep learning algorithms, and onde waist-located accelerometer and gyroscope [35]. Thus, taking into account the high number of activities to be classified (Table 11) and the same number of sensors used (just one inertial sensor at the waist), performances over 95% for accuracy will be considered an excellent result.

Another key requirement is related to a strong algorithm validation, which shows to be a potentially lengthy process and rarely done in the most correctly way in the analyzed works in Chapter 3. One of the ways to obtain well-validated results in ML systems is through training the models with a vast dataset, and with various validation techniques, such as hold-out data split, k-folds cross-validation or leave-one-out validation, among others [68, 69]. To fulfill this requirement, this module aims to build a vast dataset from the fusion and normalization of several different public and team-owned datasets (Section 5.2). From the research carried out in Chapter 3, this process has not been carried out before in other works of ADL recognition, despite its importance being already mentioned in several papers [68, 69]. Furthermore, given

the higher amount and variability of data, both at the subject and sensor levels, a big and diverse dataset enables for the development of more robust models.

Finally, the recognition of activities should be done in with the shortest possible delay between the start of the activity and its recognition [64, 65]. During this dissertation, a comparative analysis will be performed offline to study which AI-based method is preferable, as well as to find the most relevant features in ADL and fall events recognition. Despite this analysis will be performed offline, the models will be trained and tested from a future perspective of real-time practical use. In addition to seeking accurate recognition, it is important that the selected model classifies in a short time considering the window size and overlap used.

5.1.1 Data Flow Overview

The implemented ADL recognition strategy in this dissertation tries to uncover which classification algorithm is best suited to recognise 20 classes, including ADL and fall events. Furthermore, the determination of which information is most suited to this classification problem is also carried out in the proposed algorithm. To achieve these goals, a comparative analysis was carried out with different classification models and different Feature Selection Methods (FSM).

Figure 25 presents a data flow overview of the ADL recognition strategy, since raw data acquired from public and team-owned datasets, to the performance evaluation of several classification models. The acquisition of the data used in this strategy is explained in Section 5.2. All the steps taken to perform the comparative analysis: feature extraction, feature selection, window Labeling and AI-based models validation and evaluation process (Grey areas in Figure 25), are covered in Section 5.3.

All the mentioned steps related with the development of Machine Learning and Neural Networks models, validation and evaluation of these ADL recognition algorithms were implemented offline using the Matlab 2021b version, on a Lenovo Legion Y540: processor - intel core i5, 9th Gen; graphics card - NVIDIA® GeForce® GTX 1650; memory - 8 GB DDR4 at 2666 MHz and SSD PCIe of 512 GB.

5.2 Dataset Acquisition and Fusion Process

A research was carried out on public datasets of ADL and fall events. The selected datasets were further combined and normalized into a single dataset for a generalised classification of these events. The use of different and large datasets within the same topic makes validation and test of classification models more robust [68, 69] In addition to the aforementioned advantage, fusion and normalization of public datasets for activity recognition is a process that, according to the knowledge acquired during the research carried out in this dissertation, has not been completed by anyone else yet.

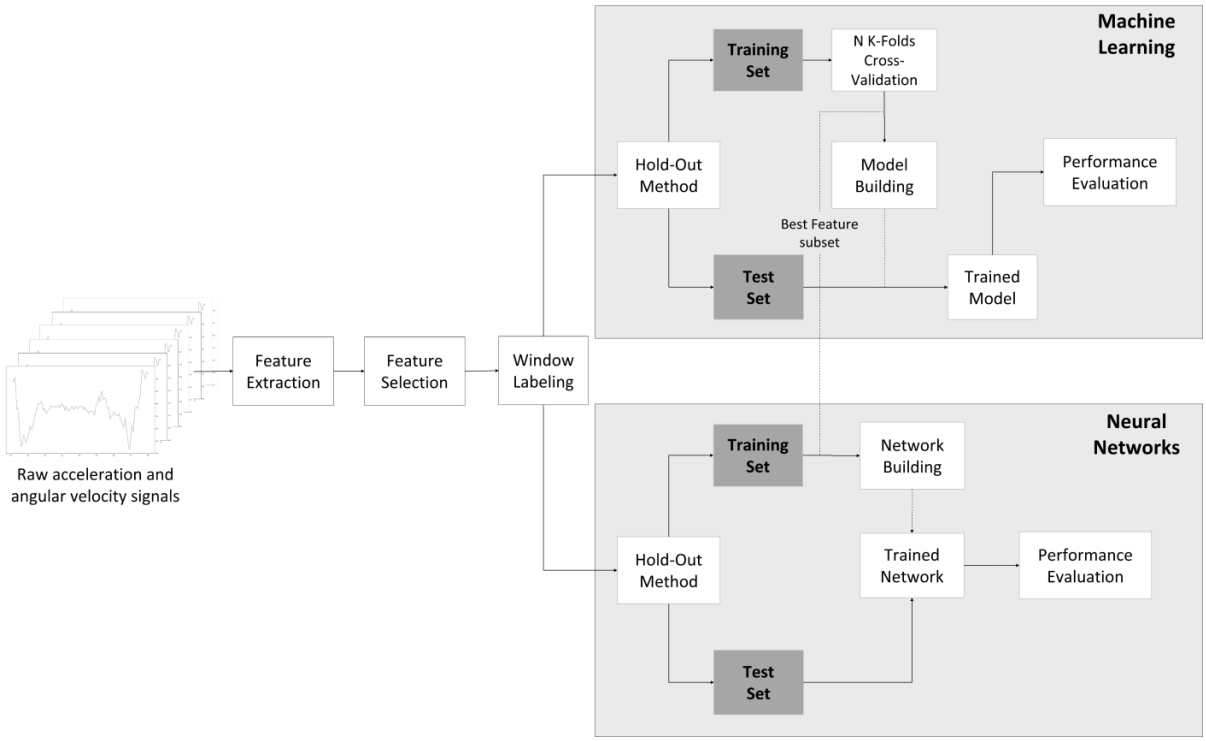


Figure 25: Dataflow diagram for the complete evaluation process of different Machine Learning and Deep Learning models for ADL recognition, alongside the best feature set selected by diverse FSM, since raw data acquired to the performance evaluation.

5.2.1 Acquisition and Description

Considering the most common information collected for activity recognition with wearable sensors, public datasets were selected if they met the following requirements: **i)** Be publicly available online for download; **ii)** Contain inertial data collected from the lower trunk (back); and **iii)** Contain postural daily activities and/or fall events. Therefore, the following keywords were used in several search engines: "Dataset", "Public", "ADL", "Falls", "IMU", "Lower Trunk", and datasets which did not met to the stipulated eligibility criteria were not considered for this dissertation. The research carried resulted in the gathering of six public datasets and three team-own datasets. The gathered public datasets were as follows.

1. Sisfall [78]: Data acquired with 23 healthy young adults (19-30 years, 149-183 cm, 42-81 kg) and 15 healthy elderly participants (60-75 years, 150-171 cm, 50-102 kg) with a device composed of two types of accelerometer and one gyroscope fixed to the waist of the participants, who performed 19 ADL and 15 fall types.

2. FallAllD [92]: Data acquired from 15 healthy subjects (21-53 years, 158-187 cm, 48-85 kg) who used 3 devices equipped with an accelerometer, a gyroscope, a magnetometer and a barometer. A total of 44 classes of ADLs and 35 classes of falls were performed.

3. FARSEEING [93]: Large-scale collaborative database to collect and share sensor signals from real-world falls. Real fall's data is acquired from either 2 locations: waist or thigh, and the acquisition devices are equipped with up to 3 sensors, namely accelerometer, gyroscope and magnetometer.

4. UCI HAR [80]: Dataset recorded from 30 healthy subjects (19-48 years) by using a waist-mounted smartphone with an embedded 3-axis accelerometer, gyroscope, and magnetometer. This dataset contains six classes of ADLs: walking, ascending stairs, descending stairs, sitting, standing, and laying.

5. Cotechini et al. [94]: Dataset acquired from 8 healthy subjects (22-29 years old, 173-187 cm, 60-94 kg) using a wearable device containing a 3-axis accelerometer and gyroscope, tied to the subject's waist, that recorded subject's acceleration and orientation. Subjects simulated 13 typologies of falls and 5 types of ADL.

6. UMAFall [95]: A dataset acquired from a total of 17 healthy subjects (18-55 years, 50-93 kg, 155-195 cm). Accelerometer, gyroscope and magnetometer data were colected from five wearable sensing devices, located on the subject's chest, waist, wrist, ankle and pocket. The participants performed 8 different ADL and 3 different typologies of falls (except by those older than 50 years, who did not perform falls).

The used team-owned datasets were as follows.

1. +Sense [90]: Dataset with data acquired from 10 healthy subjects (44.02 ± 16.42 years, 67.5 ± 16.06 kg, 172 ± 7.93 cm) and 40 subjects with Parkinson's disease (64.00 ± 10.60 years, 69.93 ± 11.41 kg, 165.93 ± 8.65 cm). A waist-mounted waistband, equipped with an accelerometer, a gyroscope and a magnetometer recorded subject's data in walking activity protocols.

2. SafeWalk [96]: Dataset acquired with 12 healthy subjects (25.33±6.33 years old, 66.92±10.07 kg, 1.74±0.11 m). Five IMUs were attached to the lower back, both back thighs, and to both feet of the subjects, who performed walking trials and front fall events.

3. InertialLab [97]: Dataset which includes data from 11 able-bodied subjects (24.53 ± 2.09 years old, 171 ± 10 cm, 65.29 ± 9.02 kg). Gyroscopes and accelerometers were attached to six lower limbs and trunk segments. Walking in varying speed and terrain (flat, ramp, and stairs) and including turns were the activities carried out by the subjects.

Table 12 resumes the public datasets found and their most relevant characteristics. The three team-own datasets added with gait and fall events data collected by the BiRDLab laboratory, are described in Table 13. Thus, after this survey, a total of 9 datasets were acquired, which make up a total dataset with inertial data of ADLs and falls from 180 subjects. All the acquired datasets were divided into several folders,

which represented the different subjects and activities. Each file in these folders represented a trial for a given activity or group of ADLs.

Table 12: Researched public datasets regarding sensing methods and location, sample frequency, participants and activities recorded. In this table: Acc = Accelerometer, GYR = Gyroscope, MAG = Magnetometer and Baro = Barometer

DataSet	Sensors	Location	Sample Frequency	Participants	ADLs Falls
SisFall [78]*	ACC & GYR	Waist	200 Hz	23 subjects<30 years 15 subjects>60 years	19 ADL 15 Falls
FALLALLD [92]*	ACC,GYR, MAG & Baro	Chest,Waist, Wrist	238 Hz	15 subjects [21–53 years]	44 ADL 35 falls
FARSEEING [93]*	ACC,GYR, MAG	Waist,Thigh	20 Hz 100 Hz	20 subjects**	Real falls
UCI HAR [80]	ACC,GYR	Waist	50 Hz	30 subjects [19-48 years]	12 ADL
Cotechini [94]*	ACC,GYR	Waist	33,33 Hz	8 subjects	5 ADL 13 Falls
UMAFall [95]	ACC,GYR, MAG	Waist,Chest, Wrists,Ankle	20 Hz	17 subjects [18-55 years]	8 ADL 3 Falls

* Several activities in these datasets were grouped into one single class of basic activities.

** Only data from 3 subjects were suitable to use.

Table 13: BiRDLab datasets description regarding sensing methods and location, sample frequency, participants and activities recorded. In this table: Acc = Accelerometer, GYR = Gyroscope and MAG = Magnetometer

DataSet	Sensors	Location	Sample Frequency	Participants	ADLs Falls
+Sense [90]	ACC, GYR, MAG	Waist	100 Hz	10 Healthy 40 Pathological	1 ADL
SafeWalk [96]		Waist,Thighs, Feet	30 Hz	12 subjects 25.33±6.33 years	1 ADL 1 Fall
InertialLab [97]		Waist	100 Hz	7 subjects [23-26 years]	5 ADL

After the acquisition, a generalized statistical study was carried out on the subjects of all datasets, where demographic information was collected from each subject, such as gender, age, weight and height, and the mean values and respective standard deviations of each of these parameters were determined (Table 14). It should be noted that this study does not cover all subjects, since not every dataset presented this demographic information or it was incomplete. *UCI HAR* dataset was excluded based on this principle.

Due to this fact, only demographic data from 150 out of the 180 subjects of the global dataset were used in this study.

Table 14: Demographic information regarding 150 of the 180 subjects of the total dataset, with maximum, minimum, mean and standard deviation computation for each factor

	Age (Years)	Weight (kg)	Height (cm)	Gender (%)
Maximum	85	102	195	M = 54% F = 46%
Minimum	18	50	149	
Mean	33.6	69.98	168.99	
Standard Deviation	±16.84	±10.99	±9.42	

As can be seen in Table 14, the global dataset presents a balanced distribution regarding to the subject's gender, containing also data from both young adults and elderly people. However, despite the presence of elderly data in the dataset, the average and standard deviation values show that the global dataset is still made up mostly of young adults, with the percentage of people over 65 years old being just over 20% of the total dataset (37/180 subjects). In short, a large dataset of 6702 files covering the numerous trials conducted by the 180 subjects was generated after the datasets' fusion process. This vast amount of data is crucial since it is far larger than any other dataset used in the research reviewed in Chapter 3.

5.2.2 Dataset Normalization Process

Due to the great variability found between datasets, regarding both the location and orientation of the sensors, and the frequencies at which the signals were acquired (Tables 12 and 13), in order to be merged, it was necessary to pass the data from all datasets by a normalization process. The different steps of this process, which can be seen in Figure 26, will be explained below.

First, from each dataset, only the data corresponding to the acceleration (accelerometer) and angular velocities (gyroscope) of the sensors located in the subjects' lumbar region were selected to be used. Data referring to other sensory information in addition to these were discarded. Then, the data were reorganized so that the orientation of the sensors corresponded to the desired one, as shown in Figure 27. Finally, all datasets underwent a resampling process so that the sampling frequency was set to 50 Hz, which is one of the most common sample frequency within activity recognition literature (See Chapter 3). Thus, after this normalization process, tables of inertial acceleration and angular velocity data similar to the one sampled in Figure 28 were obtained.

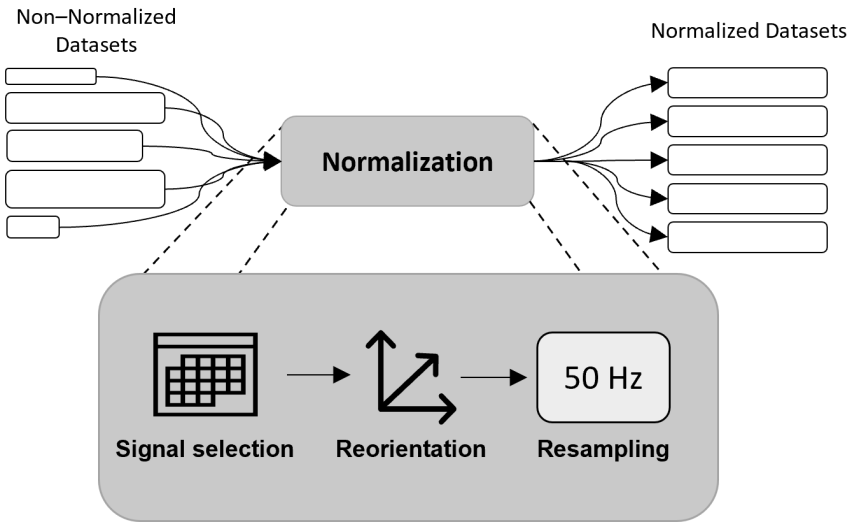


Figure 26: Normalization process steps implemented in order to build a vast dataset for ADLs recognition.

5.2.3 Labeling

The step that followed the acquisition and standardization process was the labeling, i.e., the process where to each acquired sample is assigned a "label" that identifies which ADL or type of fall event was happening at that moment. Due to the lack of information regarding the labeling of the data, this process was carried out manually or semi-automatically, depending on the information provided by the respective public dataset selected. Video and article information, descriptions of the tests carried out and the graphics of the signals themselves were used so the labeling process was as rigorous as possible. Thus, for this manual procedure, Python code was written to plot the signals present in each file, allowing easy clicks to select the start and end samples of each activity identified, labeling all samples in-between with the associated ADL value (see table 11).

As previously mentioned in Table 12, not all datasets contained the same ADLs, and some were considered two different activities despite representing the same static positions or similar activities (e.g., the standing in a room and standing in an elevator ADLs found in the *SisFall* dataset were considered as standing in this dissertation [78]). Thus, the labeling of all datasets was performed in order to find the most basic locomotion activities, static postures, transitions between those, and falls, as presented in the Table 11. After the labeling process, a study was carried out on how the labels were distributed by the global dataset, in order to assess the the balancing of the different classes (activities) to be classified. Figure 29 shows the percentage amounts of each activity present in the created dataset. As can be seen, the dataset has a greater tendency towards cyclical activities, such as walking or lying (1 - 29.73% and 4 - 18.52%, respectively), with only a small percentage of transitions between activities and fall events, such

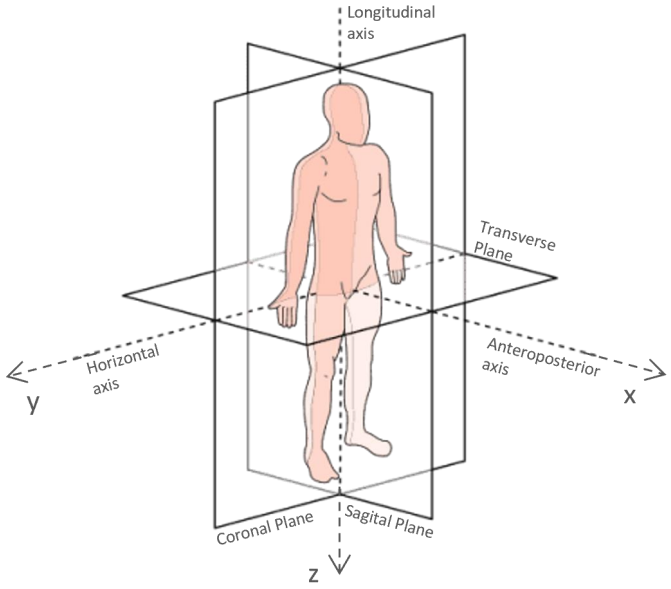


Figure 27: Desired sensor's orientation. The inertial sensors should be located on the lower trunk of the users. The arrows and letters x, y and z point in the positive direction of the Anteroposterior, Mediolateral and Longitudinal axes, respectively.

Time stamp	Acc X	Acc Y	Acc Z	Gyr X	Gyr Y	Gyr Z	Activity Label
Increasing Time			Sample 1				
			Sample n				
			...				
			Sample n				

Figure 28: Normalized inertial data table of each dataset after applying the normalization process of Figure 26.

as the fall by syncope , which is the activity with the least amount of data in the global dataset (0.27%).

5.3 AI-based classifiers: Comparative Analysis

We performed a comparative analysis, whose strategy is illustrated in Figure 25, to determine the most suitable AI-based classification model for ADL and fall events recognition, and the subset of features which attains the best performance. It should be noted that all operations were performed on the global dataset without the use of any noise filters or other sorts of processing, i.e., the raw inertial data was directly

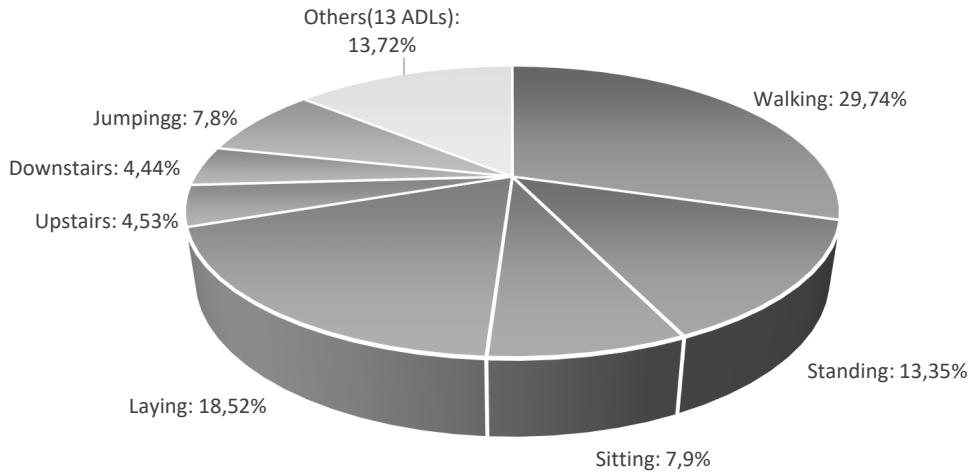


Figure 29: Percentage of each activity present in the total dataset. The activities are named according to Table 11.

implemented in the referred procedures. The following subsections will describe all the processes taken to conduct this comparative analysis, as well as theoretical definitions of the main concepts employed.

5.3.1 Feature Extraction

The first step of this block concerned the extraction of features from the raw inertial data in the global dataset (Section 5.2). These metrics were selected and calculated according to the research carried out in Chapter 3. The extraction was achieved through the sliding window method, the most used method for data segmentation and feature extraction, where a signal is segmented into several windows (subgroups) of equal size, on which different features can be calculated [64, 65].

The resolution of activity recognition can be understood as the size of the sliding window. A stream of activity data can be processed as one to observe long-term activity trends by leveraging a large window size, but if numerous actions are contained in one window, it fails to distinguish each activity. A smaller window size, may be more fitting when addressing ADLs that occur in a very short period [77].

According to the researched literature in Chapter 3, the most used sliding windows generally have a size corresponding to approximately 1 second for this type of activity classification, and the overlap between consecutive windows can vary from 50% to 87% [64, 65]. Within the scope of this project, an initial one-second window was selected to perform the comparative analysis, which corresponds to a window that contains 50 samples, with an overlap between two consecutive windows of 40 samples (80%), corresponding to an “advance” of the window of 200 ms (Figure 30).

In addition to the initial window size, for the best classifier found in the comparative analysis, various

sizes were explored to investigate the impact of the window size used for feature acquisition on classification performance. Thus, windows with **4 different sizes** were tested: **0.5 seconds**; **1 second**; **1.5 seconds**; and **2 seconds**. These were limited to 2s in order to maintain the possibility of recognizing ADLs in real-time [98, 99], in addition to allowing a good classification of transitional activities and falls, which have shorter execution times [100]. It should be noted that, due to time constrains relative to the time required to run the models for the different size and window overlaps combinations, in this dissertation the percentage of **overlap** between consecutive windows was kept at **80%** for all tests performed, despite literature suggesting that overlap can also have high impact on the classification performance and computational cost for real-time applications [98, 99]. In Chapter 6, the results of a comparative analysis between these four window sizes will be presented.

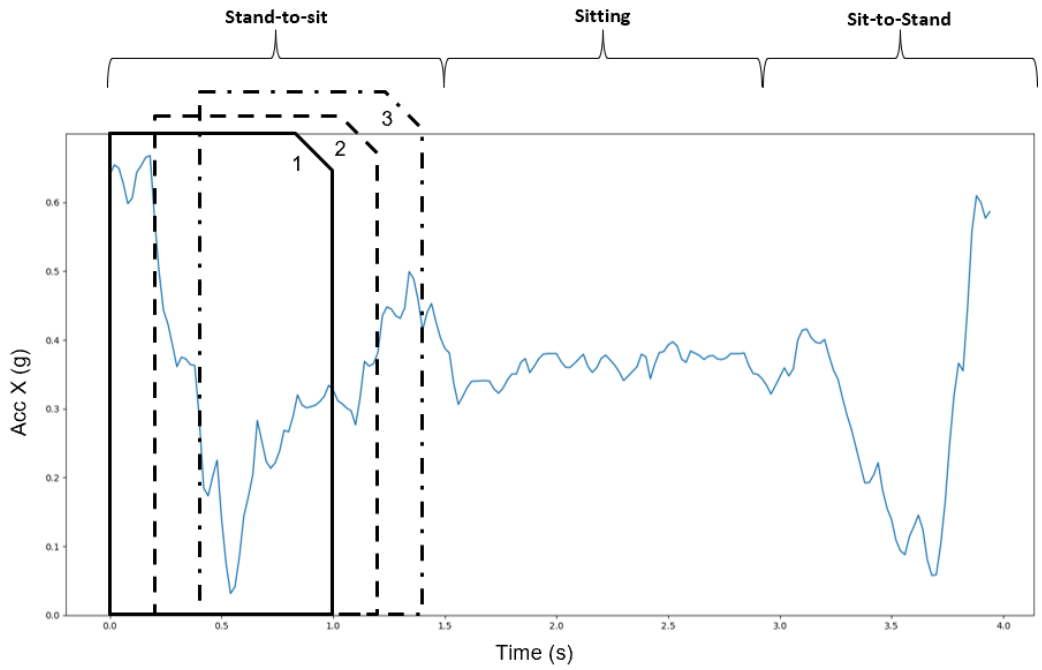


Figure 30: Signal segmentation with the sliding window technique example for feature extraction. In the image, 1, 2 and 3 represent 3 consecutive 1-sencond windows, spaced by 200ms. At the top of the graph an example is given of how samples were labeled.

The segmentation in **1s** window with **80% overlap**, and feature extraction resulted in a total set of **666660** windows. Similarly, for **0.5s, 1.5s and 2s** window sizes, **1366289, 435111 and 318516** windows were obtained, respectively. Thus, for each one, several feature values were extracted, such as the averages, maximums, minimums and standard deviations of each of the acceleration and angular velocity axis, among other metrics, making a total of **199 features** calculated. A summary of the selected features can be seen in Table 15.

Table 15: List of all extracted features from each window created. AP, V and ML refer to the anteroposterior, vertical and mediolateral axis, respectively.

Feature Number	Feature Description
[1-6]	Acceleration and angular velocity – (AP, V, ML)
[7-8]	SumVM of Acceleration and Angular Velocity
[9-24]	Skewness and Kurtosis of Acceleration, angular velocity (AP, V, ML) and SumVM signals
[25-64]	Min, max, mean, variance and Std deviation of Acceleration, angular velocity (AP, V, ML) and SumVM signals
[65-70]	Correlation between V-ML, V-AP and ML-AP axis of Acceleration and angular velocity
[71-77]	Slope, Total angular change, Resultant angular acceleration, ASMA, SMA, Absolute vertical acceleration, Cumulative horizontal displacement
[78-102]	Peak-to-Peak, Root Mean Square and Ratio Index of Acceleration, angular velocity (AP, V, ML) and SumVM signals
[103-115]	Resultant angle change, Flutuation Frequency, Resultant of Average Acceleration and Resultant of Standard Deviation (AP, V, ML)
[116-117]	Resultant of Delta changes of acceleration and angular velocity
[118-133]	Gravity component, Displacement, Displacement range, Cumulative sway length and Mean sway velocity (AP, V, ML)
[133-189]	Slope changes, Zero Crossings, Waveform Length of Acceleration, angular velocity (AP, V, ML) and SumVM signals Energy, Mean Frequency, Peak Frequency and Magnitude of Acceleration, angular velocity (AP, V, ML) and SumVM signals
[190-195]	SumVM of resultant angular velocity, average acceleration and Standard deviation, Maximum resultant angular velocity and acceleration in the horizontal plane
[196-199]	Acceleration exponential moving average, rotational angle of acceleration SumVM, Z-Score, Magnitude of Angular Displacement

A min-max scaling of the features, to values between 0 and 1, followed the feature extraction process to ensure a low computational cost when building the ADL classification models [101]. Since each of the datasets used was built from the collection of data with different devices and sensors with various ranges and sensitivities, this feature scaling process was carried out for each of the initial public datasets separately, in order to reduce possible bias derived from the different types, ranges and sensitivities of the sensors used in each dataset.

Since the labeling of dataset's signals was performed at a sample level (Section 5.2.3), the feature

windows created by the sliding window method also needed to be labeled. The window labeling was automatically carried out according to two different methods: **i)** Last Sample Labeling Method (LSLM) , where the window label corresponds to the label of last sample of that given window; and **ii)** Mode Labeling Method (MLM), where the label of a given window would be the mode of all labels present in that window [98]. In Chapter 6, the results of a comparative analysis between these two methods will be presented.

5.3.2 Feature Selection

As demonstrated in the previous step, in practical terms a limitless number of features can be computed over a signal window. However, a large feature sets can be problematic since real-time systems cannot afford the time to compute them. At the same time, in many cases it is difficult to know which features are relevant to a given task, and which ones are compromising that task. As a result, the ability to select features from a huge feature set is critical [74, 102].

Feature Selection (FS) aims at improving the performance of a prediction system, allowing faster and more cost-effective models, while providing a better understanding of the inherent regularities in data [103]. Therefore, a feature ranking and FS process was carried out with several types of FSM: Filtering, Wrapper and Embedded [104]. Since in this project one of the objectives is to understand what kind of improvements a feature selection process can introduce, both in terms of performance and computational costs reduction, 9 FSM were selected and applied to the previously calculated features [103]. The selected methods and their respective type is indicated in table 16.

Table 16: Other Feature Selection Methods tested in this dissertation for ADL recognition

Feature Selection Method [Reference]	**FSM Type**
Infnite Latent Feature Selection (**ILFS**) [103]	Filtering
Unsupervised Feature Selection with Ordinal Locality (**UFSOL**) [105]	Wrapper
Feature Selection with Adaptive Structure Learning (**FSASL**) [106]	Wrapper
Minimum-Redundancy Maximum-Relevancy (**MRMR**) [107]	Filtering
Relief-F [108]	Filtering
Mutual Information Feature Selection (**MutInfFS**) [109]	Filtering
Feature Selection Via Concave Minimization (**FSV**) [110]	Embedded
Correlation-based Feature Selection (**CFS**) [111]	Filtering
Least Absolute Shrinkage and Selection Operator (**LASSO**) [108]	Embedded
Principal Component Analysis (**PCA**) [112]	Filtering

Principal Component Analysis (PCA) is a process whose main aim is reducing the dimensionality of a dataset, while retaining as much as possible of the variation present in it, transforming the feature set into Principal Components (PCs) in order to lower the number of features used in the classification [112]. In this dissertation, the PCA method was used in two different ways: **i)** adapted as a tenth FSM; and **ii)** as a limiter on the number of features to be used during the comparative analysis.

Firstly, PCA was adapted to provide a feature ranking, as presented in Figure 31. The PCs which correspond to a cumulative percent explained of 70% were selected and a resultant and proportional PC is obtained and used to rank the extracted features [112]. The second role of PCA was to limit the number of features searched in the comparative analysis, reducing its computational cost. For this, the number of PCs corresponding to a data cumulative percent of 70% is investigated. From the PCs found, a resulting PC is formed. Thus, instead of using all 199 extracted features in the comparative analysis, only the number of features with a resulting PC value greater than $\frac{1}{199}$ multiplied by 2 are used (Figure 31). Therefore, a higher number of features than those selected from PCA is selected for comparative analysis, still with great contributions to the variability of the data while reducing the computational costs of the comparative analysis. The success of this procedure will be determined by achieving the best classification performance with a lower number of features than the number of extracted features in the previous step (199).

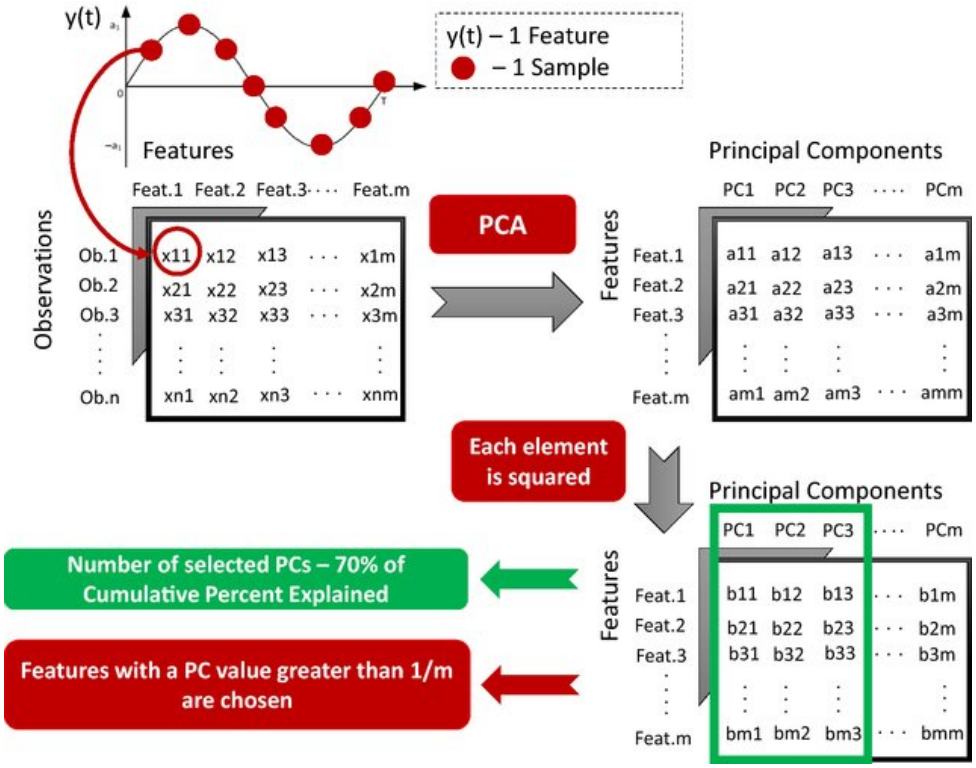


Figure 31: PCA-based procedure to rank and obtain the most crucial features to limit the computational cost of comparative analysis. Taken from [113]

5.3.3 Classification Models

In the recent past, researchers have investigated the ADL classification problem with a wide variety of AI-based techniques, since a proper classification system has to be capable of learn and tolerate noise from data [114]. Supervised learning has been a very important field in this matter, bringing about a great number of algorithms for human activity recognition (HAR) [115].

Machine Learning models such as Support Vector Machines (SVMs), K-Nearest Neighbors (K-NN), Decision Trees (DTs) or even Ensemble Learning classifiers have been the most commonly used in the past years [64, 67, 76, 115]. They have achieved promising results, despite the referred throwbacks regarding feature analysis, validation process, number of activities recognised and the size of the used dataset (See Chapter 3). In Table 17 the Machine Learning models used for the comparative analysis carried out in this dissertation are exposed. As it can be seen, some of the most prevalent classifiers in recent literature will be evaluated in this dissertation.

Table 17: Machine Learning classification models studied in this dissertation

Machine Learning Models	**Reference**
K-Nearest Neighbors (**K-NN**)	[64, 76, 115]
Ensemble Learning	[115, 116]
Decision Trees **(DT)**	[64, 76, 115]
Linear Discriminant Analysis **(LDA)**	[117]
Quadratic Discriminant Analysis **(QDA)**	[117]

5.3.3.1 K-Nearest Neighbors (K-NN)

KNN algorithm compares each new instance with all dataset available and the instance closest by distance metrics is used to perform classification When all samples of the dataset have the same relevance, which is rarely the case, computing the distance between two examples is simple, and selecting which attributes are most significant varies by application. Because every sample of the dataset must be checked for every instance that needs to be classified, the time and complexity of the methods rises according to the dataset size [118].

5.3.3.2 Ensemble Learning

The Ensemble Learning (EL) method creates multiple instances of traditional ML methods and combines them to evolve a single optimal solution to a problem. This approach is capable of producing better predictive models compared to the traditional approach. The top reasons to employ the EL method include situations where there are uncertainties in data representation, solution objectives, modeling techniques, or the existence of random initial seeds in a model. The instances or candidate methods are called base learners. Each base learner works independently as a traditional ML method, and the eventual results are combined to produce a single robust output. The combination could be done using any of the averaging (simple or weighted) methods and voting (majority or weighted) for classification methods [116].

5.3.3.3 Decision Trees

In DT, the goal is to generate a model that predicts the value of a target variable based on numerous input variables. A decision tree is constituted by a condition/internal node, based on which the tree splits into branches. The end of the branch that does not split any longer is the decision/leaf. An attribute/node must first be selected to place at the root and make one branch for its every possible value, which separates the example into subsets. This process continues iteratively in every branch until every instance of it have the same classification. The decision of which attribute to split is made based on its measure of purity, measured in bits. To classify an unknown instance, its directed through the tree accordingly to the values of its attributes in the nodes, when a leaf is reached, the instance is classified accordingly to the class that the leaf is assigned to [119].

5.3.3.4 Discriminant Analysis

Discriminant analysis (DA) is a method used in statistics, pattern recognition, and other fields, to find combinations of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a classifier, or, more commonly, for dimensionality reduction before classification. It works similarly to PCA, however, DA searches for the most variance between classes. This means that a DA system will contain information that maximizes the difference between classes. Therefore, DA requires class labelling and is a supervised machine learning method. Moreover, DA is more suitable to classification problems in which classes are known during the training phase. Hence, DA is often used in problems such as face recognition and medical diagnosis suggestion, applications in which there is abundant previously known data to train an DA system to classify new cases [120].

5.3.3.5 Neural Networks

Recently, a fast development and advancement of deep learning techniques has made possible to achieve promising results [64, 65]. These deep learning strategies do not need previous feature extraction step since they can learn features from sequential signals. Hence, this type of strategies can be used for both feature learning and activity classification [76].

Using the literature analysis carried out in Chapter 3, four distinct neural network architectures were adapted to the problem of this dissertation. Thus, architectures of Convolutional Neural Networks (CNNs) [65, 76, 114], Long Short-Term Memory Neural Networks (LSTMs) [35, 65], Bidirectional LSTMs (BiDirLSTMs) [65, 121] , and Convolutional LSTMs (ConvLSTMs) [65] were built. These are shown in Figures 32a), 32b), 32c) and 32d), respectively.

Although networks work properly using raw data, some studies refer the option of using previously acquired features as inputs [59, 121]. Thus, in order to have a comparative study on how both neural networks and Machine Learning models behave with the same type of data, in the training and testing

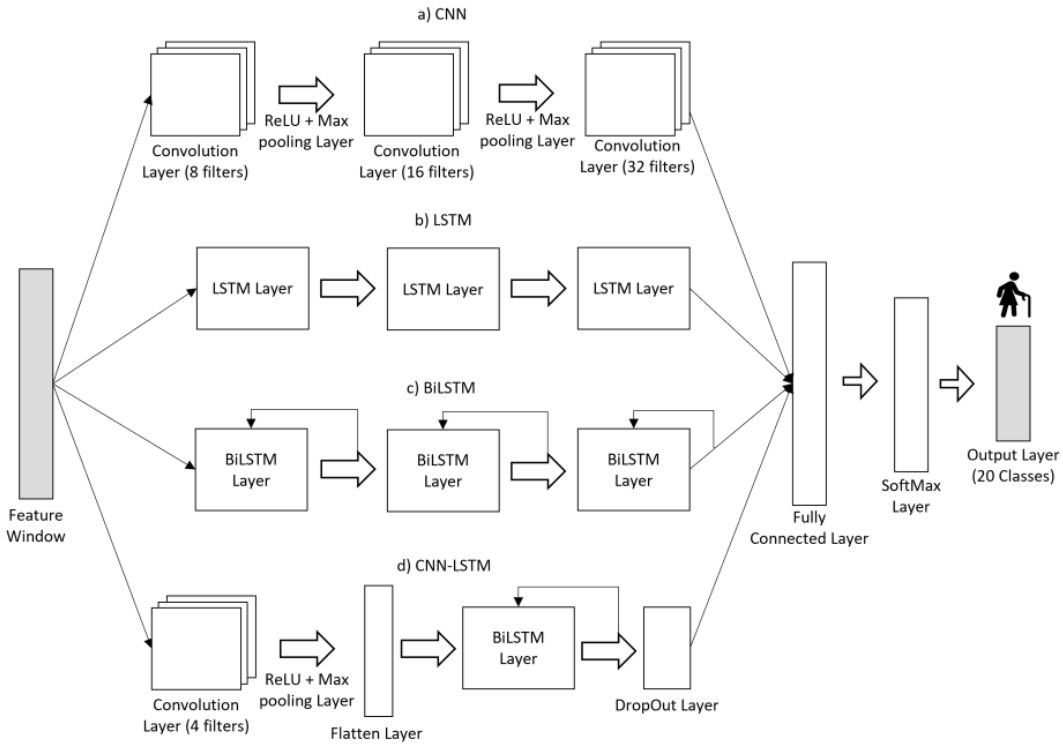


Figure 32: Neural Networks architectures used in this manuscript. The input shown represents a single feature window. a) The CNN identifies correlations from the various features provided. b) The LSTM network detects crucial temporal features. c) The BiLSTM network has a similar operating mode as LSTM, but with bidirectional LSTM layers. d) The hybrid CNN-LSTM extracts temporal patterns using convolutional features from the CNN convolutional layer. Adapted from [122].

of the created architectures, the two best feature subsets found in the first CV stage, carried out with Machine Learning models, were used as inputs. During all of the operations, test specifications such as the loss function used, number of Epochs, the optimizer employed, number of hidden layers, batch size and the Learning Rate were kept constant for all the architectures. Table 18 provides a summary of all these characteristics and respective values.

Table 18: Specifications for the use of the Deep Learning models depicted in Figure 32

Specification	**Value**
Epoch Number	100
Hidden Layers	150
Batch Size	64
Optimizer	Adam [123]
Learning Rate	0.001 (Constant)
Loss Function	Cross Entropy Loss

The employment of neural networks in this study was achieved through the implementation of the base architecture without resorting to internal layer optimizations. As a result, it is expected that the outcomes

obtained may not be optimal, hence efforts to improve their design should be included in future works, in order to identify how deep learning methods work best in the context of recognizing ADLs and fall events.

5.3.4 Validation Methods Description

5.3.4.1 Hold-Out Method

The hold-out method is one of the simplest methods to evaluate a classifier, and is widely used in classification problems, where the total dataset is divided into two distinct groups: one group to train the models, and another to test them and evaluate their performance (Figure 33) [124]. When a particular model is trained and tested with the same dataset, it ends up overfitting, i.e., it becomes overly familiar to a certain type of data, losing its ability to generalize and make a good classification when receiving new data to recognize. Thus, this step is extremely important to avoid overfitting problems in relation to the training data, since it causes the created model to be trained with a certain set of data, and to be tested with data that it has never observed before [69].

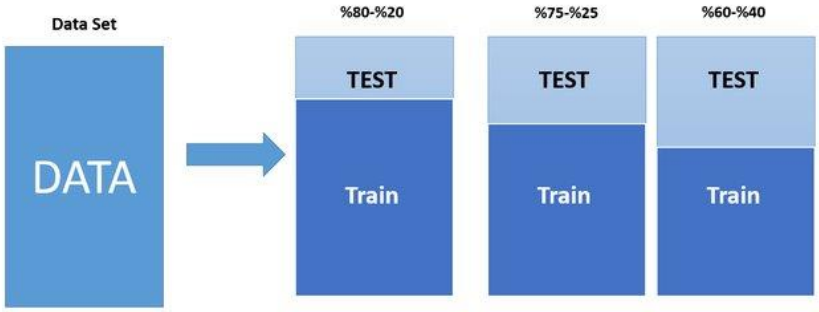


Figure 33: Test and training set separation with the hold out method. Taken from [125]

.

5.3.4.2 K-Folds Cross-Validation

K-Folds CV is a statistical method used to estimate the generalization power of machine learning models. It is important since observing the average error among k folds, a good projection of the expected error for a model built on the full dataset is made. It is commonly used to compare and select models for a given predictive model because it is easy to understand, easy to implement, and results in estimates that generally have a lower bias than other methods [124, 126]. The Figure 34 shows the cross-validation process, from the data split to the performance evaluation metrics.

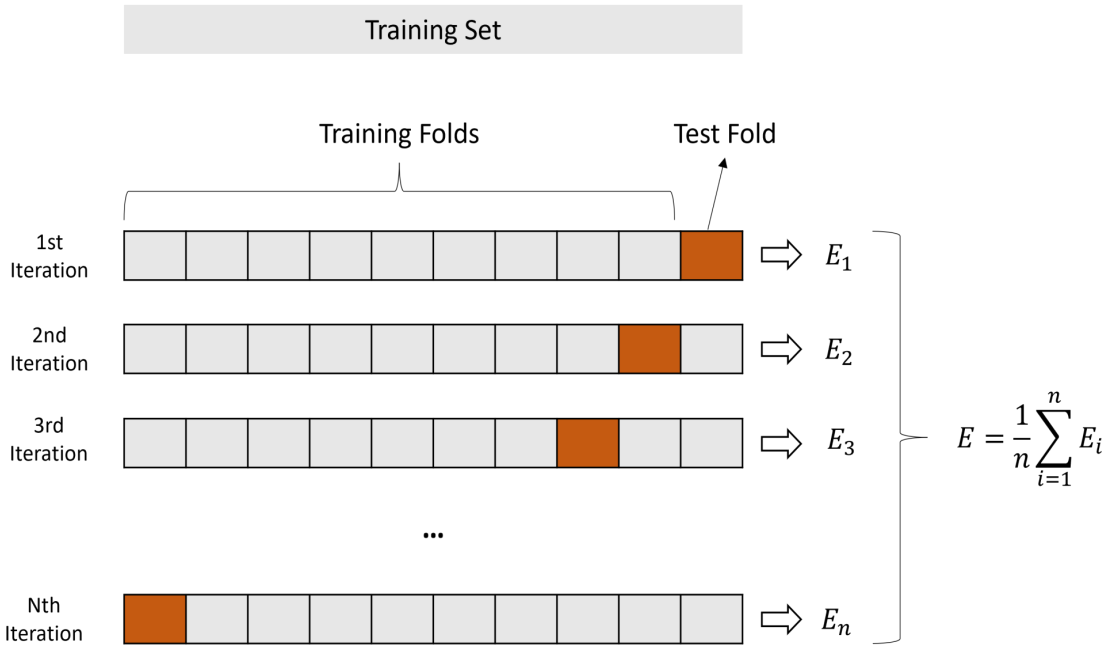


Figure 34: A general example of a k-Folds cross-validation process. Adapted from [127]

.

5.3.5 Classification Models Building and Evaluation

The first step taken for the model's building and evaluation in the comparative analysis was an initial data split. For this division, HO method was used, where the created feature **windows were divided into 70% for model training data and 30% for test data**. Due to the existence of an unbalanced dataset, this data split was done **randomly**, in order to maintain the same proportion of labels, i.e., 70% of the windows of each activity were assigned to a training set, and 30% were chosen for a test set, in order to test the trained models. With this splitting methodology, it is expected to avoid that all windows of certain activities are placed entirely either in the training data group or in the test data group, causing no activity to be only trained or only tested.

To analyze which would be the **most suitable Machine Learning model** (Table 17) and **best subset of features**v(Table 16) for the classification of ADLs and fall events, **two steps of k-folds cross-validation** were performed to the training data from the HO method. In the first step, the training data was divided into 5 folds and **one repetition** of this method was performed. In addition to the performance analysis of the aforementioned combinations, the subset of features which presented best results was also investigated, in order to allow the creation of faster models with a lower computational weight, while still obtaining the ideal performance [103, 112, 113]. The **two classifiers which presented the best performance were selected** for the second CV stage. In the second stage, with the classifiers which presented better results, the training data were divided again in 5 folds, and **10 repetitions** of the CV

process were performed. In this step, no progressive analysis was performed, using the ideal feature subset obtained from the first phase in the CV process. From the results of this 5-10-Fold CV, the two best combinations of classifier and feature selection method were chosen according to their performance results.

Finally, for the two best classifier obtained after the cross-validation phase, the classification models were trained with all training data and tested with all test data from the HO method. Using a **gridsearch** technique [128], the hyperparameters of each classification model were optimized and the models were tested before and after the **hyperparameter optimization** process, in order to ascertain the influence that this process had on improving the performance results of ADL recognition. Furthermore, at this stage the two window labeling methods presented in Section 5.3.1 were compared. Thus, the optimized classifiers were trained and tested with windows labeled by the two processes and it was verified which one presented the best performance.

In order to understand if deep-learning methods can be a viable alternative to Machine Learning methods, the **classification** of ADLs and fall events was also carried out using the four **Neural Networks** architectures presented in Figure 32. As a result, the **two best feature subsets** acquired from the validation steps were used as **input** to train and evaluate these deep learning models. The performance of neural networks with the two best set of features was analyzed and from the performance results obtained, a **choice between the machine learning-based and deep learning-based models** was made, for the **most suitable classification model for ADL recognition**.

5.3.6 Window Studies and Classification Time

As is widely verified in the literature, the impact which the window size used for feature extraction has on the performance of classification models was analyzed. Starting from the initial window of 1s (50 samples), the features of Table 15 were recalculated in windows whose size was described in Section 5.3.1. Thus, the models which performed better in Section 5.3.5 were tested for the different windows created. Finally, the **training and classification times** for the classifiers were computed and assessed as well during this study. Based on this analysis, the **classifier with the best potential to recognize ADL and fall events in real-time** was chosen, taking into account the **performance outcomes** of the various models as well as the corresponding **training and classification times**.

5.3.7 Model Evaluation Metrics

The evaluation process of AI-based algorithms is an essential part of any project, since it is the way in which the quality of the developed model can be quantified [129]. There are several ways to analyze the performance of Machine and Deep Learning classification models, and depending on number of classes, the type and balancing of the data used to train and test the models, these may or may not be the most

appropriate [130]. The evaluation of classifiers is critical in the learning process because it allows access to the performance of multiple algorithms that cannot be compared in any other way. During the comparative analysis, the following list of performance evaluation metrics was used to assess the classification models performance (Table 19) [67, 75, 76].

Table 19: Performance metrics used to evaluate the different AI-based models

Performance Evaluation Metric	**Reference**
Accuracy **(ACC)**	[67, 76, 114]
Sensitivity **(Sens)**	[35]
Specificity **(Spec)**	[35]
Precision **(Prec)**	[35]
F1 Score **(F1S)**	[76]
Mathews Correlation Coefficient **(MCC)**	[131]

Since the acquired dataset is unbalanced, i.e. certain activities appear in much greater quantity than others (For example, Walking vs Fall Forward), a bigger importance was given to the last 5 performance metrics, although accuracy was also calculated. Usually, the results are stored in a confusion matrix, as seen in Figure 35, which allows visualising the performance of the classifier.

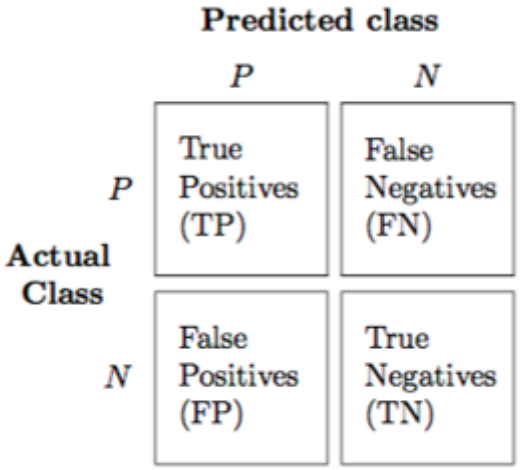


Figure 35: Confusion Matrix example.

The proper and incorrect classifications in a Confusion Matrix are typically referred to by a variety of names. True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) are the four types of classification found. The parameters TP, TN, FN, and PF are used to determine the several aforementioned performance evaluation metrics, such as precision (Prec), sensitivity (Sens), specificity (Spec), and accuracy (ACC).

5.3.7.1 Accuracy

Equation 2 shows that accuracy (ACC) is an evaluation metric that reflects the percentage of correct outcomes obtained by the classifier. The main disadvantages of ACC are that it ignores the differences between types of errors and their reliance on the data class distribution, despite the fact that the distinction between them is generally relevant in actual applications.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Although this metric is one of the first to examine when evaluating a classifier, it cannot accurately characterize the efficacy of the classifier when the number of class windows is not balanced. As a result, alternative measures that reflect the more particular features of the evaluation must be calculated.

5.3.7.2 Sensitivity

Sensitivity, presented in equation 3, measures the proportion of current positives that are correctly identified.

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

5.3.7.3 Specificity

Specificity displays the proportion of negatives that are correctly identified, and this metric is presented in equation 4.

$$Specificity = \frac{TN}{N + FP} \tag{4}$$

5.3.7.4 Precision

Precision, is the metric that indicates the percentage of correct positive results of all positive results obtained by the classifier, given by equation 5.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

5.3.7.5 F1 Score

Precision and Sensitivity are combined to form the F1 Score (F1S). It can be seen from equation 6 that the TN number is not taken into account in the calculation method, thus we can have the same value of this metric whether the TN value in the categorization results is high or low.

$$F1Score = \frac{2 \times (Sens + Prec)}{Sens + Prec} \tag{6}$$

5.3.7.6 Mathews Correlation Coefficient

The Matthews Correlation Coefficient (MCC) is a metric used in Machine Learning for evaluating the quality of binary classifications, i.e., when there are only two classes. If there are more than two classes, the classification is done by joining several classes with respect to one another, this process being iterative until there are no more possible combinations - Equation 7.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

5.4 Discussion

In this Chapter, an AI-based algorithm for the recognition of 16 ADLs and four fall events was provided with the goal of filling several gaps in the literature, especially the lack of large and varied datasets to increase the robustness of the algorithms. A survey of the most relevant features for ADL recognition was also conducted, and approaches to reduce the computational cost of this recognition were offered through a comparative analysis.

The process of merging several public datasets containing ADLs and falls should mitigate the insufficiency of data for validation of ADL recognition models, since for a segmentation in windows of 1s and with an overlap of 80%, 666660 windows were obtained to be used in the construction and evaluation of the classifiers. With the application of two validation methods, the hold-out method as well as two phases of cross-validation, in addition to the large amount of data used, the classification models compared in this algorithm will be trained and tested with different data sets, allowing its robust validation.

The use of PCA as a feature number limiter in comparative analysis should allow for a reduction in computing cost. Furthermore, it is expected that the applied feature selection procedures would be able to locate the most relevant feature combinations for the classification of activities, providing good results without the need to employ the entire set of extracted features. Furthermore, studies on the size of the window and classification time should provide insights into the potential use of the better performing classifiers in real-time applications, such as FRA. All of these strategies were combined to create a robust procedure for choosing and validating the best algorithm for ADL and fall events recognition.

Chapter 6

Activities of Daily Living Recognition - Results and Discussion

In the following subsections, the main results concerning the PCA application (Section 6.1), **comparative analysis** (Section 6.2) carried out for the choice of the two best combinations of **AI-based classification Model** and **subset of features** ranked by different FSM, considering the recognition of 16 ADLs and 4 fall events, will be shown and analyzed. Furthermore, the main results obtained in the **evaluation** process of the different **Neural Networks** architectures explained in Chapter 5 are also exposed in this Chapter (Section 6.2.3). Results regarding the other mentioned studies (**window size study and classification time analysis**) carried out on the said best combinations will also be exhibited (Section 6.3). Finally, Section 6.5 presents a discussion regarding the results obtained, highlighting the positive aspects and recognizing possible limitations. Other intermediate results are shown from Appendix A to E.

6.1 PCA Outcomes

The application of PCA as a limiter on the number of features used in comparative analysis gave rise to the graph illustrated in Figure 36. As can be seen, more than 70% of the variation present in the dataset is covered by **11 PCs**, which correspond to **55 of the 199 extracted features**. Thus, the **limit** applied in the comparative analysis was **110 features**, according to Ribeiro et al. [113].

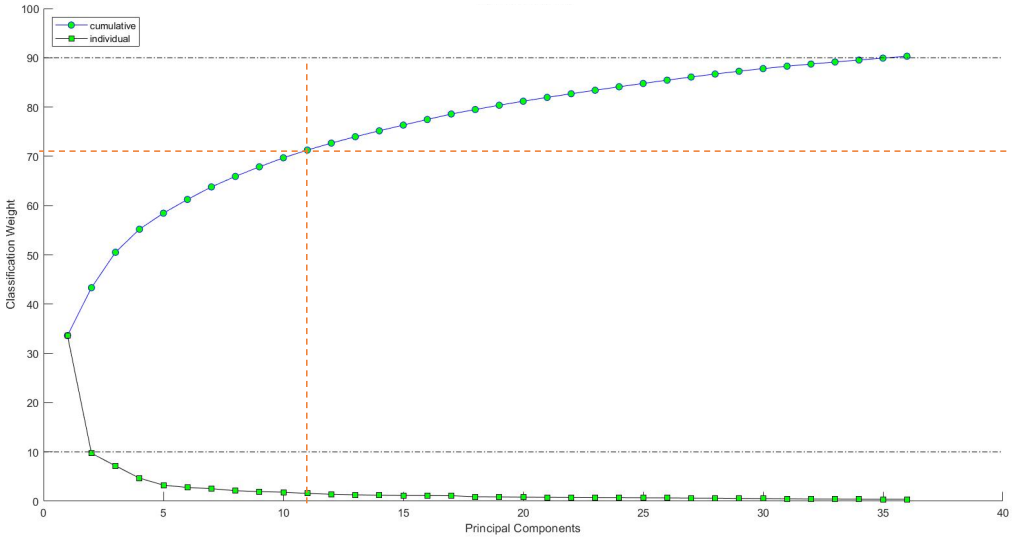


Figure 36: Scree plot for the first application of PCA to the dataset. As it can be observed, 11 Principal Components corresponded to more than 70% of the variation present in the dataset, which correspond to 55 of the 199 extracted features.

6.2 Comparative Analysis Results

6.2.1 Cross-Validation Outcomes

First, when evaluating the best ML algorithm and the best feature set, through the first step of the five-fold CV (one repetition), it was possible to identify that the **best performance results**, regardless of the feature selection model used, were obtained for the **K-NN classifier, with a squared inverse distance weighting function** (Accuracy: 93.63%, Sensitivity: 84.17%, Specificity: 99.64%, Precision: 86.80%, F1-Score: 85.43%, MCC: 85.10%) and **Ensemble Learning classification model** (Accuracy: 94.59%, Sensitivity: 82.22%, Specificity: 99.68%, Precision: 90.54%, F1-Score: 85.80%, MCC: 85.78%). In the opposite direction, the **lowest performance** was achieved when using the **Discriminant Analysis Classification with quadratic function** with the first **109 best features** ranked by **UFSOL** (Accuracy: 30.14%, Sensitivity: 33.85%, Specificity: 96.35%, Precision: 32.24%, F1-Score: 25.04%, MCC: 25.72%). The complete results comparing the different feature selection methods and classifiers can be encountered from Tables 27 to 31 in Appendix A. It is imperative to disclose that the performance evaluation metrics presented in this first stage are the mean values between the twenty different classes.

The four best combinations were chosen based on two criteria: the best overall performance amongst all computed evaluation metrics and the number of features used to train the classifiers. From the performance results obtained in the first stage of CV (Table 20), the following four classifiers and respective feature subset were chosen for the next CV step:

- **K-NN (squared inverse distance weighting function) – 85 features ranked by Relief-F**;
- **K-NN (squared inverse distance weighting function) – 85 features ranked by PCA**;
- **K-NN (squared inverse distance weighting function) – 70 features ranked by FSASL**;
- **Ensemble Learning – 65 features ranked by PCA**;

As can be seen from the Tables presented in appendix A, for the best classifiers (K-NN and Ensemble Learning), it is noted that for certain feature selection methods, the best performance results are obtained for a subset of features smaller than 110 features, which was the limit established when performing the initial PCA process. With the four best performing models, the number of iterations of the five-fold CV process was increased to ten in the second iteration of this method. The results obtained in this last validation step are also described in Table 20. Comparing the results obtained, from the first to the second iteration of CV there was a decrease of 0.01% in terms of ACC and of 0.05% of MCC for the K-NN classifier with the first 85 features ranked by Relief-F. In the case of the Ensemble Learning classification model with the 65 best features ranked by the PCA, no changes were observed in the ACC, with an increase of 0.01% in the MCC from the first to the second iteration.

Table 20: Comparison of the best classification results (ACC, Sens, Spec, Prec, F1S, MCC), attained after the second stage 5-10 k-fold cross-validation step)

CV Method	ML Model	FSM	Nº of Features	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
Five-Fold (1 rep.)	**K-NN**	**Relief-F**	85	93.63	84.17	99.64	86.80	85.43	85.10
		PCA	85	92.99	84.08	99.60	86.01	85.01	84.63
		FSASL	70	91.49	81.39	99.51	83.66	82.48	82.02
	Ensemble	**PCA**	65	94.59	82.22	99.68	90.54	85.80	85.78
Five-Fold (10 rep.)	**K-NN**	**Relief-F**	85	93.62	84,12	99,64	86,75	85,38	85,05
		PCA	85	92.95	83.91	99.60	85.88	84.86	84.48
		FSASL	70	91.48	81.40	99.51	83.59	82.45	81.99
	Ensemble	**PCA**	65	94.59	82.18	99.68	90.64	85.79	85.79

The **Ensemble Learning model with a subset of 65 features selected by PCA** (Accuracy: 94.59%, Sensitivity: 82.18%, Specificity: 99.68%, Precision: 90.64%, F1-Score: 85.79%, MCC: 85.79%) presented the best performance results. In addition, it was the case that used the smallest number of features. The model with the best performance is then selected and moves on to the next phase where it will be intensively tested to improve its performance. Despite getting worse results in 5 of the 6 performance evaluation metrics, when compared to the best performance result, the **K-NN with the first 85 features ranked by Relief-F** was also selected for further tests. This particular classification model presented a

higher sensitivity. The remaining ranked feature subsets presented in the table, were not taken into account for the next phases.

6.2.2 Hold-Out Outcomes

The two best models of the CV process were selected to be trained with all the training data obtained by the HO method and tested with the test set created by the same method. The results obtained for the performance of these models, before and after the hyperparameter optimization process, are presented in Table 21. The results obtained in this phase show an inversion of what happened in the CV process, since for the test of these two classifiers with unseen data, the performance of K-NN (Accuracy: 95.17% and MCC: 89.50%) was superior to that of Ensemble Learning (Accuracy: 68.93% and MCC: NaN%).

After the optimization process described in Chapter 5, the **KNN hyperparameters** were: (i) distance — minkowski; (ii) distance weight - squared inverse; (iii) exponent — 0.5; and (iv) number of neighbours — 1, and **Ensemble learning hyperparameters** were: (i) Method - Bag; and ii) number of learning cycles - 37. The **optimization process led to improvements in the performance of both classifiers**, as evidenced by Table 21. Regarding K-NN, this process led to an increase in Accuracy values from 95.17% to 97.27%, as well as increases in other performance evaluation metrics. As for Ensemble Learning, the optimization of its hyperparameters also resulted in significant improvements in all metrics, with Accuracy rising from 68.93% to 95.44%, reaching performance results closer to the ones observed for K-NN classifier.

Table 21: Final test results for the K-NN Model with 85 features ranked by the Relief-f and the Ensemble Learning classifier with the first 65 features ranked by PCA. The tests were made on the test data from the hold-out phase, before and after the models' hyperparameters optimization

ML Model	FSM	Nº of Features	Optimization	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
K-NN	Relief-F	85	**No**	95.17	89.28	99.72	90.28	89.76	89.50
			Yes	97.27	92.90	99.84	93.79	93.34	93.19
Ensemble Learning	PCA	65	**No**	68.93	28.72	98.04	NaN	NaN	NaN
			Yes	95.44	85.97	99.73	91.67	88.43	88.36

The results obtained when testing the optimized version of the two selected classifiers with the different window labeling methods show that labeling method can impact the performance of the models (Figure 37). By analyzing the graph, it is clear that the use of the **Mode Labeling Method** originates **better performance** results for both classifiers, when compared to the **Last Sample Labeling Method**.

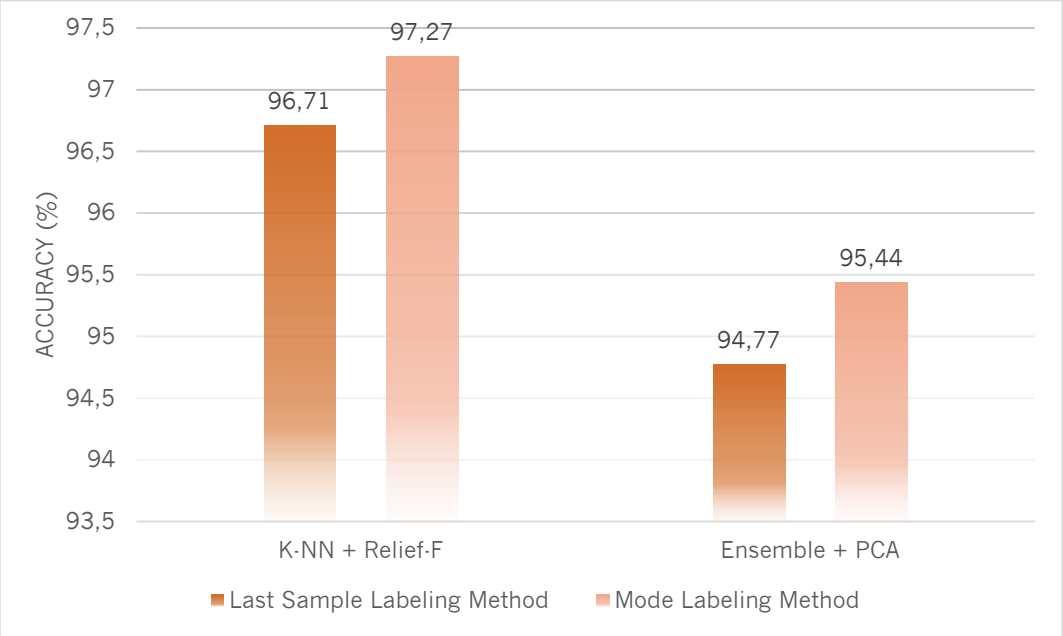


Figure 37: Labeling method study. This graphic shows an improvement of roughly 1% in accuracy values in both models. This study was realized with the optimized versions of each model.

6.2.3 Deep Learning Outcomes

As mentioned in Section 5.3.5, several architectures of Neural Networks were applied to classify the same 20 classes. In order to make a comparative analysis between the performance of Neural Networks and the best Machine Learning classifiers from Sections 6.2.1 and 6.2.2, the input for the training and testing of these architectures were the first 85 and 65 features ranked by the Relief-F and PCA feature selection methods, respectively (Appendix D, Tables 35 and 36, respectively).

The final results, obtained for each of the tests performed, can be observed in Table 22. The remaining results regarding the confusion matrices and computation time are shown in Appendix E, from Figure 42 to 49, and Table 37. Through direct observation of the table, **the best performance results were obtained for the BiLSTM when used with the 85 features raked by the Relief-F** feature selection method as the input (Accuracy: 92.55%, Sensitivity: 81.14%, Specificity: 99.57% , Precision: 85.56%, F1-Score: 83.14%, MCC: 82.83%). In the opposite direction, **the CNN applied with the subset of features ranked by PCA attained the worst performance results** (Accuracy: 57.01%, Sensitivity: 37.06%, Specificity: 97.22%, Precision: 54.67%, F1-Score: 35.47%, MCC: 37.87%). In addition to the above, from Table 37 in Appendix E it is visible that the computation times for the LSTM and bidirectional LSTM

architectures with both Feature subsets are lower than CNN and CNN-LSTM architectures.

From the tests carried out with the neural networks, it can be seen that, as was previously observed for the Machine Learning classifiers, **the best results regarding performance evaluation metrics are obtained using Relief-F** as a feature selection method. However, **the differences** between the results, for the same architecture, **are generally small** (92.55% accuracy for Relief-F vs 91.48% accuracy for PCA).

Table 22: Results for the test of the 4 Deep Learning architectures with the 85 first features ranked by Relief-f and the 65 first features ranked by PCA.

FSM	Feature Number	Architecture	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
Relief-F	85	**CNN**	57.01	37.06	97.22	54.67	35.47	37.87
		LSTM	92.06	79.58	99.55	84.25	81.02	81.01
		CNN-LSTM	88.84	74.48	99.36	75.24	74.53	74.06
		BiLSTM	92.55	81.14	99.57	85.56	83.14	82.83
PCA	65	**CNN**	42.67	26.46	96.15	54.49	22.27	24.90
		LSTM	91.46	77.81	99.51	84.38	80.61	80.38
		CNN-LSTM	88.55	74.33	99.35	75.09	74.36	73.88
		BiLSTM	91.48	79.33	99.52	83.32	80.67	80.52

6.3 Window Size Study and Classification Time

The last study evaluates which window size provides better performances result, as explained in Section 5.3.5. Due to the great diversity in the classes to be classified (**static postures, cyclical activities, transitions between postures and falls**) and their duration, this study became imperative, in order to find the size that best fits the activities to be classified.

The results attained in this study, for the optimized K-NN and Ensemble Learning classifiers, are depicted in Table 23. It should be noted that during the feature extraction step for each of the sizes of the windows, the labeling was performed according to the Mode Labeling Method, because it was the one that presented the best results in terms of performance metrics.

Both tables show a **decreasing** trend in the **performance** of the two classifiers as the **window size is increased from 0.5s to 2s**.

In addition to the results obtained regarding performance metrics, the **time required to perform the training and testing** of the K-NN and Ensemble Learning classifiers for each of the window size used in this last study was also computed. This exercise has the objective of studying the **possibility of using one of the combinations which presented better performance results in real-time**

Table 23: Window size comparative study results for the two best optimized models and respective feature subset

ML Model + FSM	Window size (s)	Window Overlap (%)	ACC (%)	Sens (%)	Spec (%)	Precn (%)	F1S (%)	MCC (%)
K-NN + Relief-f	0.5	80	98.22	95.20	99.90	96.04	95.62	95.52
	1		97.27	92.90	99.84	93.79	93.34	93.19
	1.5		96.30	91.73	99.79	91.15	91.41	91.22
	2		95.33	90.53	99.74	88.51	89.44	89.22
Ensemble + PCA	0.5		96.53	88.94	99.79	94.09	91.29	91.21
	1		95.44	85.97	99.73	91.67	88.43	88.36
	1.5		95.01	85.60	99.71	90.76	87.64	87.62
	2		94.51	85.21	99.68	89.37	86.92	86.79

situations. Table 24 depicts the results obtained for training and testing time of each classifier and window size combination.

Table 24: Classification time for the training and testing of the two best combinations of Machine Learning (ML) model and Feature Selection Method (FSM), for each of the selected windows for the window size study.

ML Model + FSM	Window size (s)	Window Overlap (%)	Test Windows	Train Time (s)	Test Time (s)	Test Time per Window (s)
K-NN + Relief-f	0.5	80	409740	4.36	213588.88	0.521
	1		199997	4.18	66782.58	0.334
	1.5		130421	4.70	12633.08	0.097
	2		95482	4.09	6752.47	0.071
Ensemble + PCA	0.5		409740	829.55	15.99	3.90×10^{-5}
	1		199997	279.03	8.54	4.27×10^{-5}
	1.5		130421	145.21	5.68	4.35×10^{-5}
	2		95482	100.23	3.94	4.13×10^{-5}

Through direct observation of the Table, the **K-NN classifier** has a **training time** of around four seconds, regardless of the size of the windows, **much lower when compared to the Ensemble classifier**. Moreover, the training time of this last model shows an increasing trend as the window size decreases. On the other hand, the **time required to test only one window** (Last column of Table 24) **is much lower for Ensemble Learning classifier**, when compared to the time required for testing with the K-NN classifier. Moreover, while test time per window for Ensemble Learning is independent from window size, test time per window for knn shows an increasing trend as the window size decreases.

Thus, ensemble learning with the first 65 features ranked by the PCA was chosen as the classification model with the highest potential for real-time ADL and fall event recognition. In Apendix C, Figure 41 and Table 34 include the confusion matrix and the performance metrics obtained for each of the evaluated classes, respectively, and their analysis allowed us to extract several lessons about the process described in Chapter 5 and classifier behavior.

6.4 Benchmark

The classifiers with best performance results obtained for the tested Machine and Deep Learning models in Sections 6.2 and 6.3 were compared to those obtained in some of the various studies analyzed in Chapter 3 with similar objectives and classification models. The comparison of performance metrics, the number of classified classes, and the number of sensors used can be seen in Tables 25 and 26, for Machine Learning classifiers and Neural Networks models, respectively.

As can be seen from Table 25, the accuracy results obtained in this dissertation for the combination of the K-NN Classifier with first 85 feature ranked by Relief-F, are **superior to the presented literature works**, with a classification problem of **20 different classes, more than any other analyzed work**. Moreover, **only data from one inertial sensor** was used to extract features and perform the classification. From Table 26 it is noted that the best Deep Learning results achieved in this dissertation **are similar, although slightly inferior, in relation to the results of other works found in the literature**, for other architectures of the same neural network type. However, once again the results in this dissertation were obtained for a classification of 20 different classes,more than in any other analyzed work.

Table 25: Machine Learning literature approaches for ADL recognition and benchmark

Work	Classifier	Number of Classes	Number of Sensors	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)
[74]	SVM	4	1 ACC 1 GYR 1 MAG	88.3	-	-	-	79.8
[132]	SVM	7	1 ACC	93.90	-	-	-	88.11
[70]	SVM	4	1 ACC	95.58	-	-	-	-
[34]	SVM	14	1 ACC 1 GYR	81	-	-	-	-
This dissertation	K-NN	20	1 ACC 1 GYR	98.22	95.20	99.90	96.04	95.62

Table 26: Deep Learning literature approaches for ADL recognition and benchmark

Work	Neural Network	Number of Classes	Inputs	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)
[67]	CNN & MLP	2 Fall Detection	3 Raw Data	99.94	98.71	99.96	-	-
[76]	CNN	12	9 Raw Data	93.24	-	-	-	92.73
[35]	DRNN (LSTM)	6	6 Raw Data	96.70	96.83	-	96.83	-
[77]	LSTM	9	36 Raw Data	93.00	-	-	-	86.5
[75]	LSTM	6	6 Raw Data	92.91	92.93	-	93.58	-
This dissertation	**LSTM**	**20**	**85 features**	**92.06**	**79.58**	**99.55**	**84.25**	**81.02**

6.5 Discussion

In this master's thesis, an algorithm to recognize sixteen ADLs and four different types of fall (twenty classes in total) from several AI-based classification models and feature selection models was built in order to find the combination that presents better performance in this type of classification. The performance of the different combinations was evaluated from the results presented in Sections 6.2 and 6.3, using the following parameters: **i)** Performance evaluation metrics, **ii)** Subset of features used, and **iii)** Training and classification times of the models.

The Machine Learning-based classification was established with a combination of different classifiers and feature selection methods, and proved to be an accurate strategy. This process allowed the most relevant features to be found for the classification of 20 distinct activity classes (Apendix D), which should aid in a time-effective and low computation strategy for real-time ADL recognition. The PCA-based approach for reducing the computational burden of the comparison analysis may therefore be judged effective since for most of the tested models and FSM, the CV best performance was achieved with fewer features than those initially extracted.

The success achieved regarding the performance of the classification models, through different validation methods, namely hold-out and cross-validation, shows the robustness of the applied processes. The creation of a vast dataset also positively influenced the performance results obtained. According to the results presented in the Tables 23 and 24, the **Ensemble Learning** classifier with a subset of the **65 first features** ranked by the feature selection method **PCA**, with a window of 0.5s and an overlap of 80%, was the model which showed the most potential for the classification of the 20 ADL classes in real-time.

Despite **not being the best performer in terms of evaluation metrics**, it had a **lower classification time, and lower computational cost**, due to the use of a smaller feature subset, allowing in theory, its deployment in real-time situations.

Regarding the analysis made to the window sizes, according to the literature, windows with a larger size have the tendency to perform worst in the recognition of shorter activities or transitions, and better in the classification of cyclical or static activities, which are maintained for longer periods of time. Furthermore, as the window size increases, the model's capability to recognize ADLs in real time decreases. [98, 99]. As stated in Table 23, for smaller windows, the performance of the classifiers analyzed in this dissertation increases, thus, corroborating what is described in the literature.

Through direct observation of the Table 34 in Appendix C it is noticed that **shorter activities** (e.g. Falls and Turning) **are generally recognized with lower accuracy** and, as the **duration of the activities increases**, e.g., cyclic activities or static postures, **the accuracy** of this ADL **increases**. **The activities with the best results are walking and lying, while the worst are Syncope Fall and Fall Backwards**. The amount of data available for each of the activities may have influenced this event, since, as mentioned in Section 5.2, the **constructed dataset is unbalanced**, containing more samples related to cyclic activities and static postures (Walking, Standing, Sitting, Lying, Upstairs, Downstairs, Running and Jumping) than for different fall events and transitional activities. Furthermore, after examining the comparison between the ground truth and the classification produced for the chosen model, some windows right after or before the transition between two activities were misclassified, as seen in Figures 38 and 39 (blue boxes), in addition to, on other occasions, also presenting misclassifications during the execution of certain cyclic activities (orange boxes in Figures 38 and 39).

These misclassifications when transitioning between two activities can be **justified by patterns of movement different from the usual at the beginning or end of certain activities**. As can be seen in Figure 39, when transitioning from class 2 (Standing) to class 5 or 6 (UpStairs and DownStairs, respectively), there is regularly a misclassification with class 1 (Walking). The **similarities between the movement patterns of walking and going up or down stairs** can also justify the misclassifications highlighted by the orange boxes during these activities.

However, from the confusion matrix of Figure 41 from Appendix C, and from the classification graphs of the previous figures, it is noticed that **only one ADL class is not confused with the walking class** (class 18, which is changing position while lying). In some of these cases it may be normal due to **similar movement patterns** (e.g. UpStairs, DownStairs, Running), as previously demonstrated. Another strong reason could be the existing imbalance between the classes of the constructed dataset, with great **bias for cyclic activities, especially walking**. This fact can also justify the misclassifications found even with activities that have completely different acceleration patterns (e.g. Standing). This fact reveals the **need to still make improvements to the created dataset**, either through the acquisition of more data from the activities which are less represented, or through **data augmentation** methods, to assess in more

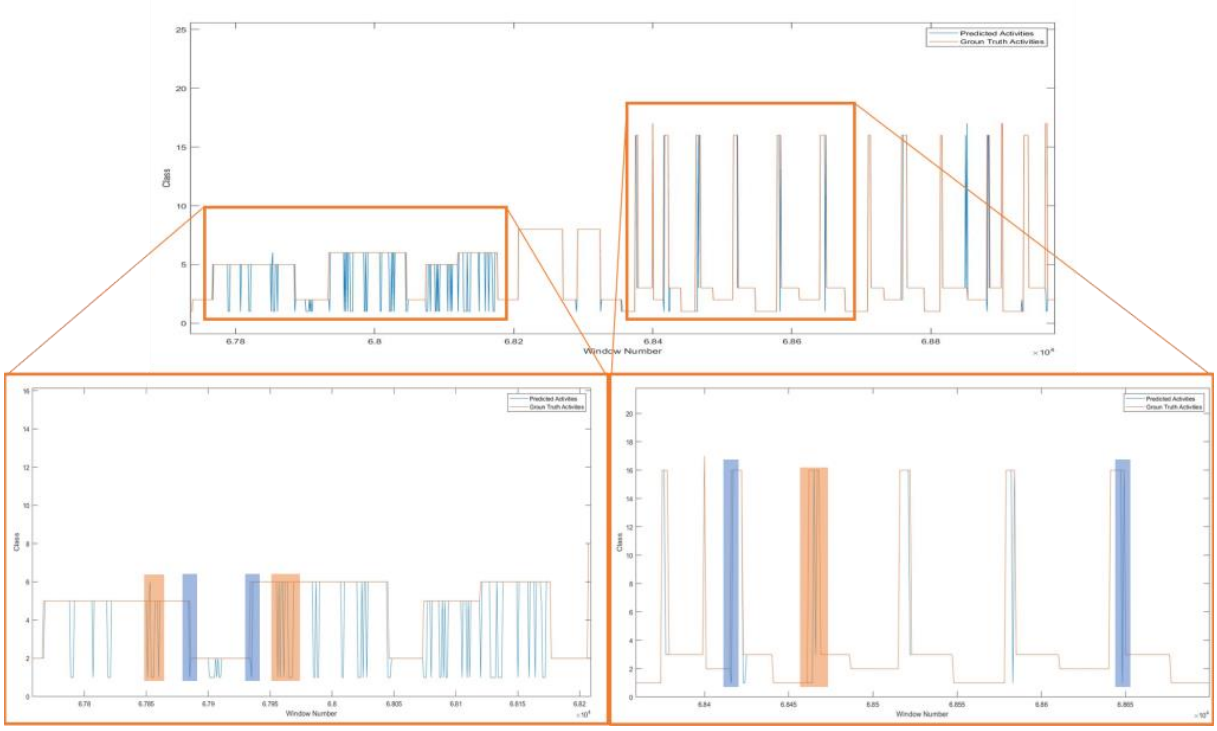


Figure 38: Comparison between the ground truth (orange line) with the classification output of the chosen chosen model with unseen data (blue line).

depth the influence that a balanced dataset may have in mitigating these misclassifications [77]. **Post-classification post-processing techniques such as filters or majority voting** can help eliminate other sporadic misclassifications that can arise due to unexpected changes in the movement patterns of a given activity. Moreover, **activity transition diagrams** can help in **reducing the number of classes for a given classification based in the last classification** that was made [100] (Figure 40).

In short, the process carried out for **the classification of sixteen ADL and four types of fall events proved to be successful**, taking into account the general results of the performance evaluation metrics achieved for the K-NN and Ensemble Learning classifiers. Another positive aspect was the fact that **the window classification time (0.039ms) for Ensemble Learning is much lower than the selected window advance** (100ms advance for a 500ms window), allowing in theory the deployment of this model in real-time. In addition, for the ensemble learning classifier, the comparative analysis of the set of features extracted by the different feature selection models provided a **reduction in the number of features by more than 65%** from the 199 initial features, to a subset of 65 features, also supporting the possibility of applying this type of models in real-time systems. Furthermore, it enables the identification of a group of features which produce superior results in the recognition of various ADL and fall occurrences, even when compared with results of similar works found in the literature (Section 6.4, Table 25).

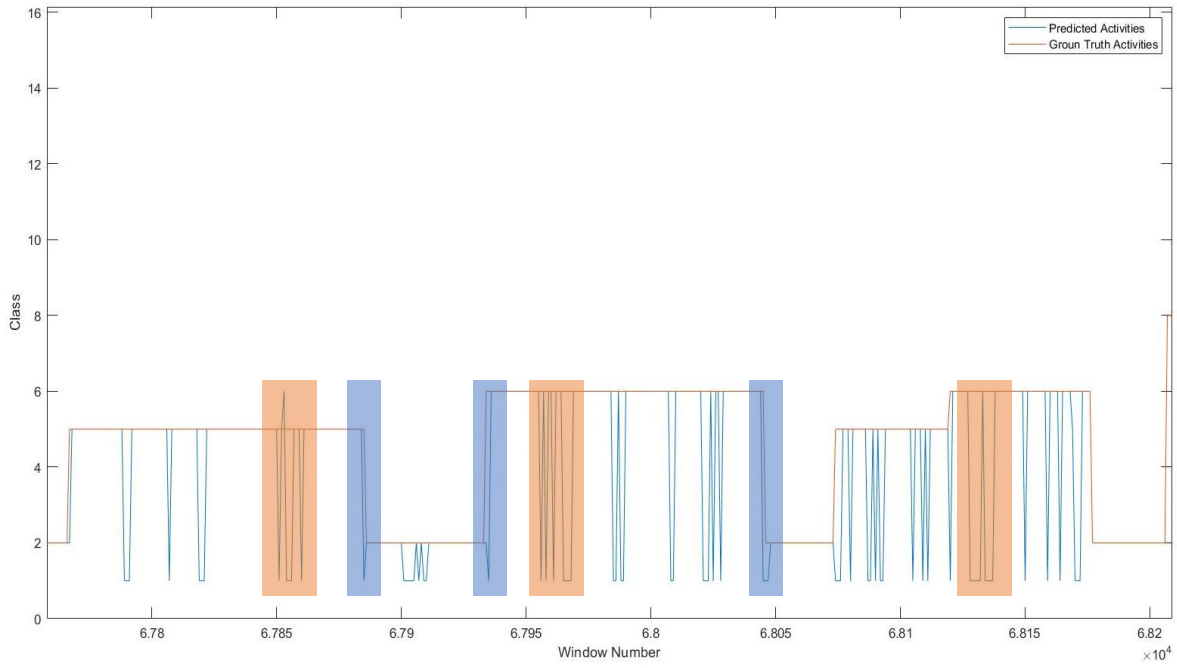


Figure 39: Comparison between the ground truth (orange line) and the classification output of the chosen best Machine Learning model with unseen data (blue line). Blue boxes represent model misclassifications when transitioning between two activities. Orange boxes represent model misclassifications during the execution of certain cyclic activities.

A direct comparison between the two best results obtained in tests performed with neural networks (Table 22) and with the process developed for Machine Learning Classifiers (Table 23) shows that the use of different architectures with the same features ranked by the best feature selection methods found in Section 6.2 as inputs, **did not produce as good results as when using the K-NN and Ensemble Learning classifiers**.

The tested **LSTM and BiLSTM achieved the best results** in terms of performance metrics, as well as in relation to the computation time of training and testing them. As mentioned in the literature, **LSTM have a more suitable architecture for classifying sequential data** [77], and this better behavior is expected compared to other architectures, such as CNN [35]. Moreover, as investigated in the literature, **the results achieved by BiLSTM were slightly superior to the basic LSTM** [65, 121]. The feature selection method had little influence on the results obtained for the performance of neural networks, however **the 85 features ranked by Relief-f returned results regarding performance metrics slightly higher than the 65 features ranked by the PCA**, as what happened in the studies carried out on Machine Learning models.

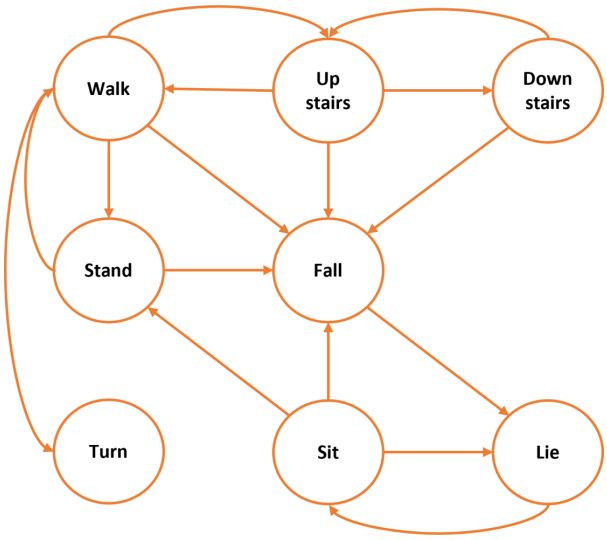


Figure 40: An example of an Activity transition diagram. Adapted from [100]

CNNs are generally more used in problems involving image inputs and in this process **the implemented CNN architecture received as input 1D arrays of features extracted from sequential data**. This method of use may justify the **poor results obtained with the CNNs in terms of performance metrics**, given that the **convolution processes applied to extract features from the arrays used were not carried out in the most effective way**. Despite these poor performance results, other architectures involving this type of networks should continue to be tested, with different type of inputs, such as the raw inertial data, since several studies claim to obtain positive results in the recognition of activities with the use of this type of neural networks [64, 67, 76].

Thus, this initial study of the behavior of several neural networks showed that, despite positive results in some of the cases tested, it **did not reach the desired potential**, when compared to the results obtained with the machine learning classifiers and other deep learning-based studies found in literature (Section 6.4, Table 26). This fact raises the **need to carry out several studies and some changes in the future**, in order to obtain performance results higher than those found in the literature. The architectures used must be improved in the search for better results in order to be comparable with the studies currently carried out in the literature. Furthermore, ablation studies should be carried out in order to better understand the influence that different parameters have on the performance of these networks. Finally, and as mentioned in most of the works analyzed in the literature, **new tests should be carried out** in the future with windows segmented from **raw inertial data as inputs**, rather than features, given the ability of neural networks to carry out a process extraction of features [76].

Despite the generalized positive results achieved by the different processes applied with machine learning and deep learning-based classifiers, questions regarding the **data split method must still be investigated**. The data division was done randomly and restricted by the data imbalance and the fact

that not all subjects realized the same 20 activity classes, so that all activities were trained and tested in the same proportion (70% of the data for training and 30% for the classifier test). This fact suggests the possibility of the **existence of very similar data in the two sets**, which may lead to overfitting of the models when testing with new subjects' data features, even though they were tested with windows that were not included in their training. Thus, new data split methods should be tested, such as **dividing the data by the subjects or by the trials** that make up the dataset [133], still trying to guarantee the balance of the labels in the two sets.

Chapter 7

Conclusions

This dissertation confirms the state-of-the art concern that falls among the elderly are a serious risk that can result in deadly or other non-fatal consequences, as well as with high societal and economic impact to families and society. As a result, efficient approaches to combat the stated problem are critical, and any endeavor to minimize or prevent a fall can save numerous lives. As a result, real-time fall risk analysis techniques are critical in combating the problem of geriatric falls.

Despite the fact that several steps have already been taken in this direction, one of the conclusions reached in this dissertation is that these tools still require improvements to make them easier and more comfortable to use for stakeholders, and to allow them to freely execute their movements while being clearly presented with a level of risk of falling.

The design of a Fall risk assessment strategy, which resulted in a three-part modular architecture through an eHealth platform that allows real-time FRA using only a waistband equipped with one inertial sensor and electronic gadgets readily available these days, was discussed in Chapter 4. The inclusion of few wearable sensors enables for a more comfortable use of this equipment. Furthermore, the proposed modular strategy should allow for a multifactorial risk assessment, which encompasses numerous risk elements, making it more thorough and dependable. It also enables the the combination of both prospective and immediate FRA approaches to provide a single fall risk level, using AI-based technologies.

Concerning the ADL recognition discussed in Chapter 5, this topic was chosen since it is critical for assessing fall risk, despite being rarely used in the FRA methodologies studied in this dissertation. According to studies, it is statistically necessary to consider the activities that a person undertakes, since the number of falls varies based on the activity that was being carried out at the time of the fall [32].

Machine and Deep Learning-based methodologies were analyzed in a 20 class (between ADLs and fall events) classification problem. A procedure of fusion and normalization of public datasets was carried out

in order to generate a big enough dataset, larger than any existent in the researched state-of-the-art, to validate the activity classification models. Further, other relevant datasets of the team, including data from parkinsonic and healthy subjects, with different ages, weights and heights were included, which allows a high variability that brings robustness to the developed methods. These aspects were intended to address major concerns reported in the literature related to the lack of useful data for the validation of activity recognition algorithms, specially considering the targeted population.

Regarding the machine learning results attained in this dissertation for the ADL recognition problem, a comparative analysis was performed by using different machine learning models and feature selection methods, which allowed the author to find the most relevant features to classify 20 distinct activity classes (Appendix D). The selected classification algorithm for building the final machine learning model was the Ensemble Learning classifier with the first 65 features ranked the PCA feature selection method. The performance obtained for this algorithm was: Accuracy- 96.53%; Sensitivity- 88.94%; Specificity- 99.79%; Precision- 94.09%; F1-Score- 91.29%; and MCC- 91.21%. Furthermore, the classification time per window for this combination of classifier and subset of features was lower than 4.5e-5s for every window size tested, which is also lower than the time between 2 consecutive classification windows for every window size tested at a sampling frequency of 50 Hz for the machine used in these tests, which shows the potential of using this procedure in real-time ADL recognition and FRA applications. Although the deep learning outcomes for the same classification problem were not as good as in the prior procedure, their potential was demonstrated, indicating that they could be a good option in the future with appropriate modifications and new testing.

In short, the work herein presented enables to answer the RQs outlined in Chapter 1.

- **RQ 1**: **What is the best strategy to perform FRA in real-time and in a multi-factorial way?** This RQ is addressed in Chapter 4.

  The best FRA strategy that can be implemented to perform in real-time and in a multi-factorial way is still a much discussed topic and without a definitive answer. However, through the research work carried out, it is concluded that a modular FRA tool will be the most suitable, where each module will work for the fall risk assessment of a certain set of risk factors (multifactorial), whether related to the movement and balance of users (real-time FRA), whether related to other demographic or environmental characteristics (prospective FRA). With the development of sensor technology and AI, it is concluded that the application of these in each of the modules will be the ideal way to carry out a better Fall Risk Assessment, providing relevant information to clinical caregivers and even the elderly through an eHealth platform.

  A module based on prospective approaches to FRA is essential in order to provide a general characterization of an elderly person's abilities in the face of possible fall situations in the long term future, based on information such as clinical scale results, demographic information, physical and

psychological aspects of the elderly, as well as information regarding their surroundings. On this type of data, regression algorithms can be used to calculate a baseline fall risk value.

Immediate FRA approaches should also be evaluated in another module in order to provide insights into the elderly person's current moment, ie, information regarding the activities that the elderly practices in real time, as well as the manner in which they practice these ADLs, enabling the analysis of the current moment and detecting situations of imbalance or near fall, such as slips or trips, in addition to the activity to be carried out. The analysis of waist-mounted intertial data using AI-based classification algorithms is the best solution for this module. Furthermore, activity recognition has extra weight since the gathered inertial signals can be used to evaluate the evolution of the efficiency of performance of various activities in the medium/long term, allowing diagnoses regarding the evolution of the elderly's physical capacity.

These blocks can be combined to assign a single fall risk value, simplifying analysis by caretakers or the elderly themselves, who, as the tool's target demographic, must clearly understand when they are in situations where they are at high risk of falling, in order to avoid them. eHealth systems embedded in devices such as smartphones are the best approach for calculating and displaying the risk level, enabling communication between the elderly and carers easier in the event of a high fall risk or even a fall event.

- **RQ 2**: **What is the best ML and DL-based strategy to implement for real-time ADL and fall events recognition?** This RQ is addressed in Chapters 5 and 6.

  From the work results presented in Chapter 6, the best Machine Learning algorithm for ADL and fall events recognition is the Ensemble Learning classifier, which achieved an ACC, Sens, Spec, Prec, F1S and MCC of 96.53%, 88.94%, 99.79%, 94.09%, 91.29% and 91.21%, respectively, for a 0.5s window and the use of the first 65 features ranked by PCA. Moreover, this classification model combination presented the smallest test time per window (< 4.5e-5s), a value much lower than the time between two consecutive classifications for every window size tested, thus being the method which presented a higher potential for real-time applications when compared with the K-NN classifier. This classifier, despite having achieved better performance results, has a classification time for each window size tested greater than the time between two consecutive classifications, making it impossible to use it in real-time with the equipment used: a Lenovo Legion Y540: processor - intel core i5, 9th Gen; graphics card - NVIDIA® GeForce® GTX 1650; memory - 8 GB DDR4 at 2666 MHz and SSD PCIe of 512 GB.

  From the work results presented in Chapter 6, the best Deep Learning algorithm for ADL and fall events recognition was the bidirectional LSTM architecture described in Figure 32, Chapter 5, combined with the 85 most important features ranked by Relief-F. When testing the trained model

with unseen data it was achieved an ACC, SENS, SPEC, PREC, F1S and MCC of 92.06%, 79.58%, 99.55%, 84.25%, 81.02% and 81.01%, respectively.

Finally, it should be noted that these two results were obtained using data that already included a portion of the desired target audience, the elderly, in order to ensure a good performance when using these models in the future for the purpose of this dissertation, the FRA in a continuous and multifactorial manner in the elderly population.

- **RQ 3**: **What are the most suitable features for recognizing ADL and fall events from Machine Learning models?** This RQ is addressed in Chapters 5 and 6.

  To unveil which were the most suitable features for recognizing ADL and fall events, a progressive analysis was made to a set of 199 features in order to discover which number and which subset of features gave the best result in terms of performance evaluation metrics. The 65 most important features ranked by PCA combined with the Ensemble Machine Learning classifier achieved the highest overall performance amongst all combinations tested - Appendix D, Table 36. With a 10-5-Fold Cross-Validation during the progressive analysis, it was achieved 99.08%, 99.39%, 97.72%, 99.47%, 99.43%, and 97.00% for ACC, SENS, SPEC, PREC, F1S and MCC, respectively, for a 1s window. When testing the trained model with unseen data it was achieved an ACC, SENS, SPEC, PREC, F1S and MCC of 99.65%, 100%, 84.44%,99.64%, 99.82% and 91.73%, sequentially. The 85 most significant features resulting from the Relief-F combined with the K-NN machine learning classifier achieved the second highest overall performance amongst all combinations tested - Appendix D, Table 35. With a 10-5-Fold Cross-Validation during the progressive analysis, it was achieved 93.62%, 84.12%, 99.64%, 86.75%, 85.38%, and 85.05% for ACC, Sens, Spec, Prec, F1S and MCC, respectively, and a 1s window. When testing the trained model with unseen data, for the same window size, it was achieved an ACC, Sens, Spec, Prec, F1S and MCC of 97.27%, 92.90%, 99.84%, 93.79%, 93.34% and 93.19%, respectively, after the model optimization.

- **RQ 4**: **Which inclusion and exclusion criteria, ADL and circuits should be included in a protocol to collect relevant, variable data for training Machine Learning models to recognize ADL and fall events from elderly in real-life environments?** This RQ is addressed in Chapter 4.

  In order to collect relevant, variable data for training Machine Learning models to recognize ADL and fall events from elderly in real-life environments, inclusion and exclusion criteria must first be applied, in order to decide on which subjects the collection protocols should be applied. From the elaboration of a collection protocol together with doctors and caregivers from nursing homes, it was concluded that seniors with dementia (MMSE≤15); severe depression (EDG ≥ 11); physical disabilities (unable to walk without walking aid, FAC < 4); orthopedic, cardiac, or respiratory diseases

that affect locomotion; Morbid obesity (BMI ≥ 30); and Fear of falling moderate-high (Short FES-I ≥ 14) should be excluded from ADL data acquisition protocols. The ADL addressed in acquisition protocols must take into account the target audience, which in this dissertation are the elderly, and be as close as possible to those that they can perform in a free-living context. In this way, the basic movements of locomotion (such as walking, running or walking up or downstairs), static postures (such as standing or siting) and the transitory activities (e.g., sit-to-stand and stand-to-sit transitions) presented in Tables 11 and 9 are essential in any protocol for collecting inertial data. Finally, these activities must be carried out in the form of a circuit, integrating numerous ADL in a sequential manner, in order to ensure a more natural and continuous execution by the subjects, best emulating the free-living contexts of the elderly.

7.1 Future Work

The Fall Risk Assessment tool and eHealth platform idealized in Chapter 4, comprising all its individual modules and their integration, must be validated as future work. To begin, the systems that make up each of its modules must be built and tested in accordance with the guidelines outlined in Chapters 2 and 3, using a large amount of data, ideally from the elderly or high-risk users, to match the tool to its target users as closely as possible. After validation, other methods for assessing other risk variables, as well as various types of sensors for obtaining more information, particularly in connection to the surrounding environment, should be included. Moreover, despite the fact that a waistband has been enhanced to begin collecting inertial data in nursing homes, and that this collection has already commenced, future work must be carried out in order to be able to miniaturize all components of the waistband, making it even more comfortable to use.

Although the constructed dataset contains a large amount of successfully normalized data for the validation of activity recognition algorithms, it still requires the application of several strategies for its improvement, such as the addition of more data from elderly people, and the use of data balancing techniques, such as class weights, in order to balance the quantity of samples of activities represented on a smaller scale. Furthermore, data augmentation processes can be used in order to obtain an even more diversified dataset. Key stakeholders such as older individuals and healthcare experts should be involved in any future work on building these datasets [85]. These concerns were already considered in this work (Section 4.8) and will continue to be addressed.

Despite the excellent results of the proposed ADL and fall event detection algorithms using Machine Learning, new methods must be verified during the data splitting process for validation in the future, to guarantee that the great results attained in this dissertation were not due to the usage of extremely comparable data during the train and test stages of the model. Adaptive windows can be a good alternative in order to avoid the lower performances of the classifier for fall events and shorter activities. Another critical

aspect in the validation of classification models that should be addressed in more depth in the near future is the collection of inertial data from a large number of older subjects, since they are the target audience of this dissertation, with a wide range of ADL performed. The creation of a dataset of this kind will be a valuable asset to validate future ADL recognition models, in order to be implemented in the design FRA tool. Furthermore, future validations should be performed in real time from data gathering techniques used in nursing homes.

As mentioned in this dissertation, despite comparable results with the literature, continued development of the tested neural networks and the usage of raw inertial data as input to the networks should be processes to be implemented. Ablation studies should be conducted to determine the impact of the various stages of each architecture on ADL and fall events recognition.

Bibliography

[1] W. H. Organization. *Ageing and health*. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed: 2021-12-05.

[2] W. H. Organization. *Falls*. https://www.who.int/news-room/fact-sheets/detail/falls. Accessed: 2021-09-30.

[3] N. M. Peel. "Epidemiology of Falls in Older Age." In: *Canadian Journal on Aging / La Revue canadienne du vieillissement* 30.1 (2011), pp. 7–19. doi: 10.1017/S071498081000070X.

[4] L. Z. Rubenstein. "Falls in older people: epidemiology, risk factors and strategies for prevention." In: *Age and Ageing* 35.suppl$_2$ (Sept. 2006), pp. ii37–ii41. issn: 0002-0729. doi: 10.1093/ageing/afl084.

[5] A. M. Tromp, S. M. F. Pluijm, J. H. Smit, D. J. H. Deeg, L. M. Bouter, and P. Lips. "Fall-risk screening test: A prospective study on predictors for falls in community-dwelling elderly." English. In: *Journal of Clinical Epidemiology* 54.8 (Aug. 2001), pp. 837–844. issn: 0895-4356, 1878-5921. doi: 10.1016/S0895-4356(01)00349-3.

[6] M. Tinetti, M. Speechley, and S. Ginter. "Risk Factors for Falls among Elderly Persons Living in the Community." In: *New England Journal of Medicine* 319.26 (1988). cited By 4686, pp. 1701–1707. doi: 10.1056/NEJM198812293192604.

[7] B. Vellas, F. Cayla, H. Bocquet, F. de Pemille, and J. L. Albarede. "Prospective Study of Restriction of Activity in Old People After Falls." In: *Age and Ageing* 16.3 (May 1987), pp. 189–193. issn: 0002-0729. doi: 10.1093/ageing/16.3.189.

[8] *WHO European health information at your fingertips.* en. url: https://gateway.euro.who.int/en/indicators/hfamdb_17-deaths-accidental-falls/visualizations/#id=29059&tab=graph (visited on 05/20/2021).

[9] J. Parkkari. "Fall-induced injuries among elderly people." undefined. In: *Lancet* 350 (1997), p. 1174.

[10] M. Hemmatpour, R. Ferrero, B. Montrucchio, and M. Rebaudengo. "A Review on Fall Prediction and Prevention System for Personal Devices: Evaluation and Experimental Results." In: *Advances in Human-Computer Interaction* 2019 (2019). issn: 1687-5893. doi: 10.1155/2019/9610567.

[11] W. H. Organization, W. H. O. Ageing, and L. C. Unit. *WHO global report on falls prevention in older age*. World Health Organization, 2008.

[12] D. Hendrie, S. Hall, G. Arena, and M. Legge. "Health System Costs of Falls of Older Adults in Western Australia." In: *Australian health review : a publication of the Australian Hospital Association* 28 (Jan. 2005), pp. 363–73. doi: 10.1071/AH040363.

[13] I. Nurmi-Lüthje and P. Lüthje. "Incidence and costs of falls and fall injuries among elderly in institutional care." In: *Scandinavian journal of primary health care* 20 (June 2002), pp. 118–22. doi: 10.1080/pri.20.2.118.122.

[14] B. S. Roudsari, B. E. Ebel, P. S. Corso, N.-A. M. Molinari, and T. D. Koepsell. "The acute medical care costs of fall-related injuries among the U.S. older adults." In: *Injury* 36.11 (2005), pp. 1316–1322. issn: 0020-1383. doi: https://doi.org/10.1016/j.injury.2005.05.024.

[15] D Carey and M Laffoy. "Hospitalisations due to falls in older persons." In: *Irish medical journal* 98.6 (2005), pp. 179—181. issn: 0332-3102.

[16] P. Kannus, J. Parkkari, H. Sievänen, A. Heinonen, I. Vuori, and M. Järvinen. "Epidemiology of hip fractures." In: *Bone* 18.1, Supplement 1 (1996). Proceedings of the International Symposium on Physical Loading, Exercise, and Bone, S57–S63. issn: 8756-3282. doi: https://doi.org/10.1016/8756-3282(95)00381-9.

[17] L. Montesinos, R. Castaldo, and L. Pecchia. "Wearable Inertial Sensors for Fall Risk Assessment and Prediction in Older Adults: A Systematic Review and Meta-Analysis." In: *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society* 26.3 (2018), pp. 573–582. issn: 1558-0210. doi: 10.1109/TNSRE.2017.2771383.

[18] S. Mathias, U. S. L. Nayak, and B. Isaacs. "Balance in elderly patients: the "get-up and go"test." In: *Archives of physical medicine and rehabilitation* 67 6 (1986), pp. 387–9.

[19] S. W. Muir, K. Berg, B. Chesworth, and M. Speechley. "Use of the Berg Balance Scale for Predicting Multiple Falls in Community-Dwelling Elderly People: A Prospective Study." In: *Physical Therapy* 88.4 (Apr. 2008), pp. 449–459. issn: 0031-9023. doi: 10.2522/ptj.20070251.

[20] M. E. Tinetti, T. F. Williams, and R. Mayewski. "Fall risk index for elderly patients based on number of chronic disabilities." eng. In: *The American Journal of Medicine* 80.3 (Mar. 1986), pp. 429–434. issn: 0002-9343. doi: 10.1016/0002-9343(86)90717-5.

[21] B. Najafi, K. Aminian, F. Loew, Y. Blanc, and P. Robert. "Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly." In: *IEEE Transactions on Biomedical Engineering* 49.8 (2002), pp. 843–851. doi: 10.1109/TBME.2002.800763.

[22] A. Danielsen, H. Olofsen, and B. A. Bremdal. “Increasing fall risk awareness using wearables: A fall risk awareness protocol.” In: *Journal of Biomedical Informatics* 63 (2016), pp. 184–194. issn: 1532-0464. doi: 10.1016/j.jbi.2016.08.016.

[23] K. C. Fleming, J. M. Evans, D. C. Weber, and D. S. Chutka. “Practical Functional Assessment of Elderly Persons: A Primary-Care Approach.” In: *Mayo Clinic Proceedings* 70.9 (1995), pp. 890–910. issn: 0025-6196. doi: https://doi.org/10.4065/70.9.890.

[24] G. P. Wolf-Klein, F. A. Silverstone, N. Basavaraju, C. J. Foley, A. Pascaru, and P. H. Ma. “Prevention of falls in the elderly population.” In: *Archives of physical medicine and rehabilitation* 69.9 (1988), pp. 689–91.

[25] L. Z. Rubenstein, K. R. Josephson, and D. Osterweil. “Falls and Fall Prevention in the Nursing Home.” In: *Clinics in Geriatric Medicine* 12.4 (1996). Gait and Balance Disorders, pp. 881–902. issn: 0749-0690. doi: https://doi.org/10.1016/S0749-0690(18)30206-4.

[26] J. Howcroft, J. Kofman, and E. D. Lemaire. “Review of fall risk assessment in geriatric populations using inertial sensors.” In: *Journal of NeuroEngineering and Rehabilitation* 10.1 (2013). issn: 1743-0003. doi: 10.1186/1743-0003-10-91.

[27] Y. S. Delahoz and M. A. Labrador. “Survey on Fall Detection and Fall Prevention Using Wearable and External Sensors.” In: *Sensors* 14.10 (2014), pp. 19806–19842. doi: 10.3390/s141019806.

[28] R. Rucco, A. Sorriso, M. Liparoti, G. Ferraioli, P. Sorrentino, M. Ambrosanio, and F. Baselice. “Type and Location of Wearable Sensors for Monitoring Falls during Static and Dynamic Tasks in Healthy Elderly: A Review.” In: *Sensors (Basel, Switzerland)* 18.5 (2018). issn: 1424-8220. doi: 10.3390/s18051613.

[29] W. Saadeh, S. A. Butt, and M. A. B. Altaf. “A Patient-Specific Single Sensor IoT-Based Wearable Fall Prediction and Detection System.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.5 (2019). Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering, pp. 995–1003. issn: 1558-0210. doi: 10.1109/TNSRE.2019.2911602.

[30] A. Leone, G. Rescio, L. Giampetruzzi, and P. Siciliano. “Smart EMG-based Socks for Leg Muscles Contraction Assessment.” In: *2019 IEEE International Symposium on Measurements Networking (M N)*. 2019, pp. 1–6. doi: 10.1109/IWMN.2019.8804991.

[31] S. Shan and T. Yuan. “A wearable pre-impact fall detector using feature selection and Support Vector Machine.” In: *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS* (2010), pp. 1686–1689.

[32] S. Lord and C. Sherrington. “Falls in Older People: Risk Factors and Strategies for Prevention.” In: *Inj. Prev.* 9 (Jan. 2001).

[33] S. Lord. "Predictors of nursing home placement and mortality of residents in intermediate care." In: *Age and ageing* 23.6 (1994), pp. 499—504. issn: 0002-0729. doi: 10.1093/ageing/23.6.452.

[34] W. Gomaa, R. Elbasiony, and S. Ashry. "ADL Classification Based on Autocorrelation Function of Inertial Signals." In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). 2017, pp. 833–837. doi: 10.1109/ICMLA.2017.00-53.

[35] A. Murad and J.-Y. Pyun. "Deep Recurrent Neural Networks for Human Activity Recognition." In: *Sensors* 17.11 (2017). issn: 1424-8220. doi: 10.3390/s17112556.

[36] "Falls by elderly people at home: prevalence and associated factors." In: 17 (Nov. 1988), pp. 365–372. issn: 0002-0729. doi: 10.1093/ageing/17.6.365.

[37] A. J. Campbell, J. Reinken, B. C. Allan, and G. S. Martinez. "Falls in old age: a study of frequency and related clinical factors." eng. In: *Age and Ageing* 10.4 (Nov. 1981), pp. 264–270. issn: 0002-0729. doi: 10.1093/ageing/10.4.264.

[38] "Home hazards and falls in the elderly: the role of health and functional status." In: 85 (Apr. 1995), pp. 509–515. issn: 0090-0036. doi: 10.2105/ajph.85.4.509.

[39] J. Mahoney, M. Sager, N. C. Dunham, and J. Johnson. "Risk of falls after hospital discharge." eng. In: *Journal of the American Geriatrics Society* 42.3 (Mar. 1994), pp. 269–274. issn: 0002-8614. doi: 10.1111/j.1532-5415.1994.tb01750.x.

[40] M. Speechley and M. Tinetti. "Falls and injuries in frail and vigorous community elderly persons." eng. In: *Journal of the American Geriatrics Society* 39.1 (Jan. 1991), pp. 46–52. issn: 0002-8614. doi: 10.1111/j.1532-5415.1991.tb05905.x.

[41] S. R. Lord, J. A. Ward, P. Williams, and K. J. Anstey. "Physiological Factors Associated with Falls in Older Community-Dwelling Women." In: *Journal of the American Geriatrics Society* 42.10 (1994), pp. 1110–1117. doi: https://doi.org/10.1111/j.1532-5415.1994.tb06218.x.

[42] S. R. Lord, J. A. Ward, P. Williams, and K. J. Anstey. "An epidemiological study of falls in older community-dwelling women: the Randwick falls and fractures study." In: *Australian Journal of Public Health* 17.3 (1993), pp. 240–245. doi: https://doi.org/10.1111/j.1753-6405.1993.tb00143.x.

[43] L. Seematter-Bagnoud, V. Wietlisbach, B. Yersin, and C. J. Büla. "Healthcare Utilization of Elderly Persons Hospitalized After a Noninjurious Fall in a Swiss Academic Medical Center." en. In: *Journal of the American Geriatrics Society* 54.6 (2006), pp. 891–897. issn: 1532-5415. doi: 10.1111/j.1532-5415.2006.00743.x.

[44] S. C. C. Fabrício, R. A. P. Rodrigues, and M. L. d. Costa Junior. "Falls among older adults seen at a São Paulo State public hospital: causes and consequences." en. In: *Revista de Saúde Pública* 38 (Feb. 2004), pp. 93–99. issn: 0034-8910, 1518-8787. doi: 10.1590/S0034-89102004000100013.

[45] R. N. Ferreira, N. F. Ribeiro, and C. P. Santos. "Fall Risk Assessment Using Wearable Sensors: A Narrative Review." In: *Sensors* 22.3 (2022). issn: 1424-8220. doi: 10.3390/s22030984. url: https://www.mdpi.com/1424-8220/22/3/984.

[46] M. W. Rivolta, M. Aktaruzzaman, G. Rizzo, C. L. Lafortuna, M. Ferrarin, G. Bovi, D. R. Bonardi, and R. Sassi. "Automatic vs. clinical assessment of fall risk in older individuals: A proof of concept." In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). ISSN: 1558-4615. 2015, pp. 6935–6938. doi: 10.1109/EMBC.2015.7319987.

[47] M. W. Rivolta, M. Aktaruzzaman, G. Rizzo, C. L. Lafortuna, M. Ferrarin, G. Bovi, D. R. Bonardi, A. Caspani, and R. Sassi. "Evaluation of the Tinetti score and fall risk assessment via accelerometry-based movement analysis." In: *Artificial Intelligence in Medicine* 95 (2019), pp. 38–47. issn: 0933-3657. doi: 10.1016/j.artmed.2018.08.005.

[48] W. Tang, G. Fulk, S. Zeigler, T. Zhang, and E. Sazonov. "Estimating Berg Balance Scale and Mini Balance Evaluation System Test Scores by Using Wearable Shoe Sensors." In: *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. 2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI). ISSN: 2641-3604. 2019, pp. 1–4. doi: 10.1109/BHI.2019.8834631.

[49] A. Shahzad, S. Ko, S. Lee, J.-A. Lee, and K. Kim. "Quantitative Assessment of Balance Impairment for Fall-Risk Estimation Using Wearable Triaxial Accelerometer." In: *IEEE Sensors Journal* 17.20 (2017). Conference Name: IEEE Sensors Journal, pp. 6743–6751. issn: 1558-1748. doi: 10.1109/JSEN.2017.2749446.

[50] Z. Yang, C. Song, F. Lin, J. Langan, and W. Xu. "A Smart Environment-Adapting Timed-Up-and-Go System Powered by Sensor-Embedded Insoles." In: *IEEE Internet of Things Journal* 6.2 (2019). Conference Name: IEEE Internet of Things Journal, pp. 1298–1305. issn: 2327-4662. doi: 10.1109/JIOT.2018.2844837.

[51] S. Saporito, M. A. Brodie, K. Delbaere, J. Hoogland, H. Nijboer, S. M. Rispens, G. Spina, M. Stevens, and J. Annegarn. "Remote timed up and go evaluation from activities of daily living reveals changing mobility after surgery." In: *Physiological Measurement* 40.3 (2019). Publisher: IOP Publishing, p. 035004. issn: 0967-3334. doi: 10.1088/1361-6579/ab0d3e.

[52] F. Buisseret, L. Catinus, R. Grenard, L. Jojczyk, D. Fievez, V. Barvaux, and F. Dierick. "Timed Up and Go and Six-Minute Walking Tests with Wearable Inertial Sensor: One Step Further for the Prediction of the Risk of Fall in Elderly Nursing Home People." In: *Sensors* 20.11 (2020). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, p. 3207. doi: 10.3390/s20113207.

[53] J. Silva and I. Sousa. "Instrumented timed up and go: Fall risk assessment based on inertial wearable sensors." In: *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA). 2016, pp. 1–6. doi: 10.1109/MeMeA.2016.7533778.

[54] G. Rescio, A. Leone, A. Caroppo, and P. Siciliano. "A preliminary study on fall risk evaluation through electromiography systems." In: *2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL)*. 2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL). 2015, pp. 219–221. doi: 10.1109/IMCTL.2015.7359590.

[55] S. Parvaneh, J. Mohler, N. Toosizadeh, G. S. Grewal, and B. Najafi. "Postural Transitions during Activities of Daily Living Could Identify Frailty Status: Application of Wearable Technology to Identify Frailty during Unsupervised Condition." In: *Gerontology* 63.5 (2017), pp. 479–487. issn: 0304-324X, 1423-0003. doi: 10.1159/000460292.

[56] G. Cola, M. Avvenuti, A. Vecchio, G.-Z. Yang, and B. Lo. "An On-Node Processing Approach for Anomaly Detection in Gait." In: *IEEE Sensors Journal* 15.11 (2015). Conference Name: IEEE Sensors Journal. issn: 1558-1748. doi: 10.1109/JSEN.2015.2464774.

[57] M. Di Rosa, J. M. Hausdorff, V. Stara, L. Rossi, L. Glynn, M. Casey, S. Burkard, and A. Cherubini. "Concurrent validation of an index to estimate fall risk in community dwelling seniors through a wireless sensor insole system: A pilot study." In: *Gait & Posture* 55 (2017), pp. 6–11. issn: 0966-6362. doi: 10.1016/j.gaitpost.2017.03.037.

[58] W. Sansrimahachai, M. Toahchoodee, R. Piakaew, T. Vijitphu, and S. Jeenboonmee. "Real-time fall risk assessment system based on acceleration data." In: *2017 International Conference on Orange Technologies (ICOT)*. 2017 International Conference on Orange Technologies (ICOT). 2017, pp. 33–36. doi: 10.1109/ICOT.2017.8336083.

[59] D. Gagnon, B.-A. J. Menelas, and M. J.-D. Otis. "Qualitative Risk of Falling Assessment Based on Gait Abnormalities." In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. 2013 IEEE International Conference on Systems, Man, and Cybernetics. 2013, pp. 3966–3971. doi: 10.1109/SMC.2013.677.

[60] A. Mehmood, A. Nadeem, M. Ashraf, M. S. Siddiqui, K. Rizwan, and K. Ahsan. "A Fall Risk Assessment Mechanism for Elderly People Through Muscle Fatigue Analysis on Data From Body Area Sensor Network." In: *IEEE Sensors Journal* 21.5 (2021), pp. 6679–6690. issn: 1558-1748. doi: 10.1109/JSEN.2020.3043285.

[61] H. Chandra. *An intuitive approach to DTW — Dynamic Time Warping*. en. June 2020. url: https://towardsdatascience.com/an-intuitive-approach-to-dtw-dynamic-time-warping-f660ccb77ff4 (visited on 03/13/2022).

[62] V. Annese and D. De Venuto. "FPGA based architecture for fall-risk assessment during gait monitoring by synchronous EEG/EMG." In: *2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*. 2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI). 2015, pp. 116–121. doi: 10.1109/IWASI.2015.7184953.

[63] B. R. Greene, K. McManus, S. J. Redmond, B. Caulfield, and C. C. Quinn. "Digital assessment of falls risk, frailty, and mobility impairment using wearable sensors." en. In: *npj Digital Medicine* 2.1 (Dec. 2019), pp. 1–7. issn: 2398-6352. doi: 10.1038/s41746-019-0204-z.

[64] R. Baghezza, K. Bouchard, A. Bouzouane, and C. Gouin-Vallerand. "From Offline to Real-Time Distributed Activity Recognition in Wireless Sensor Networks for Healthcare: A Review." In: *Sensors* 21.8 (2021). issn: 1424-8220. doi: 10.3390/s21082786.

[65] R. Elangovan, T. Perumal, and s Padmavathi. "Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review." In: *IEEE Sensors Journal* PP (Mar. 2021), pp. 1–1. doi: 10.1109/JSEN.2021.3069927.

[66] Q. Wang, H. Luo, J. Wang, L. Sun, Z. Ma, C. Zhang, M. Fu, and F. Zhao. "Recent Advances in Pedestrian Navigation Activity Recognition: A Review." In: *IEEE Sensors Journal* (Jan. 2022), pp. 1–1. doi: 10.1109/JSEN.2022.3153610.

[67] G. Wang, Q. Li, L. Wang, Y. Zhang, and Z. Liu. "Elderly Fall Detection with an Accelerometer Using Lightweight Neural Networks." en. In: *Electronics* 8.11 (Nov. 2019), p. 1354. issn: 2079-9292. doi: 10.3390/electronics8111354.

[68] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano. "Trends in human activity recognition using smartphones." In: *Journal of Reliable Intelligent Environments* 7.3 (Sept. 2021), pp. 189–213. issn: 2199-4676. doi: 10.1007/s40860-021-00147-0.

[69] T. Shany, K. Wang, Y. Liu, N. H. Lovell, and S. J. Redmond. "Review: Are we stumbling in our quest to find the best predictor? Over-optimism in sensor-based models for predicting falls in older adults." In: *Healthcare Technology Letters* 2.4 (2015), pp. 79–88. issn: 2053-3713. doi: https://doi.org/10.1049/htl.2015.0019.

[70] S. S. Kambhampati, V. Singh, M. S. Manikandan, and B. Ramkumar. "Unified framework for triaxial accelerometer-based fall event detection and classification using cumulants and hierarchical decision tree classifier." In: *Healthcare Technology Letters* 2.4 (2015), pp. 101–107. doi: https://doi.org/10.1049/htl.2015.0018.

[71] P. Gupta and T. Dallas. "Feature selection and activity recognition system using a single triaxial accelerometer." In: *IEEE transactions on bio-medical engineering* 61.6 (2014), pp. 1780–1786. issn: 1558-2531. doi: 10.1109/TBME.2014.2307069.

[72] C. Moufawad el Achkar, C. Lenoble-Hoskovec, A. Paraschiv-Ionescu, K. Major, C. Büla, and K. Aminian. "Instrumented shoes for activity classification in the elderly." In: *Gait & Posture* 44 (2016), pp. 12–17. issn: 0966-6362. doi: 10.1016/j.gaitpost.2015.10.016.

[73] H. Nguyen, K. Lebel, S. Bogard, E. Goubault, P. Boissy, and C. Duval. "Using Inertial Sensors to Automatically Detect and Segment Activities of Daily Living in People With Parkinson's Disease." In: *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society* 26.1 (2018), pp. 197–204. issn: 1558-0210. doi: 10.1109/TNSRE.2017.2745418.

[74] M. Awais, L. Chiari, E. A. F. Ihlen, J. L. Helbostad, and L. Palmerini. "Physical Activity Classification for Elderly People in Free-Living Conditions." In: *IEEE Journal of Biomedical and Health Informatics* 23.1 (2019). Conference Name: IEEE Journal of Biomedical and Health Informatics, pp. 197–207. issn: 2168-2208. doi: 10.1109/JBHI.2018.2820179.

[75] M. Altuve, P. Lizarazo, and J. Villamizar. "Human activity recognition using improved complete ensemble EMD with adaptive noise and long short-term memory neural networks." In: *Biocybernetics and Biomedical Engineering* 40.3 (2020), pp. 901–909. issn: 0208-5216. doi: https://doi.org/10.1016/j.bbe.2020.04.007.

[76] M. Gil-Martín, R. San-Segundo, F. Fernández-Martínez, and J. Ferreiros-López. "Improving physical activity recognition using a new deep learning architecture and post-processing techniques." In: *Engineering Applications of Artificial Intelligence* 92 (2020), p. 103679. issn: 0952-1976. doi: https://doi.org/10.1016/j.engappai.2020.103679.

[77] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong. "Sensor Data Acquisition and Multimodal Sensor Fusion for Human Activity Recognition Using Deep Learning." In: *Sensors* 19.7 (2019). issn: 1424-8220. doi: 10.3390/s19071716.

[78] A. Sucerquia, J. D. López, and J. F. Vargas-Bonilla. "SisFall: A Fall and Movement Dataset." In: *Sensors* 17.1 (2017). issn: 1424-8220. doi: 10.3390/s17010198.

[79] *UCI Machine Learning Repository: PAMAP2 Physical Activity Monitoring Data Set*. url: https://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring (visited on 03/16/2022).

[80] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. "A Public Domain Dataset for Human Activity Recognition Using Smartphones." In: *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (2013).

[81] M. Zhang and A. A. Sawchuk. "USC-HAD: A Daily Activity Dataset for Ubiquitous Activity Recognition Using Wearable Sensors." In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. UbiComp '12. Association for Computing Machinery, 2012, pp. 1036–1043. isbn: 9781450312240. doi: 10.1145/2370216.2370438.

[82] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. del R. Millán, and D. Roggen. "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition." In: *Pattern Recognition Letters* 34.15 (2013). Smart Approaches for Human Action Recognition, pp. 2033–2042. issn: 0167-8655. doi: https://doi.org/10.1016/j.patrec.2012.12.014.

[83] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster. "Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom." In: *IEEE Transactions on Information Technology in Biomedicine* 14.2 (2010), pp. 436–446. doi: 10.1109/TITB.2009.2036165.

[84] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster. "Activity Recognition from On-Body Sensors: Accuracy-Power Trade-Off by Dynamic Sensor Selection." In: *Wireless Sensor Networks*. Ed. by R. Verdone. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 17–33. isbn: 978-3-540-77690-1.

[85] J. Sumner, L. Chong, A. Bundele, and Y. Wei Lim. "Co-Designing Technology for Aging in Place: A Systematic Review." In: *The Gerontologist* 61.7 (June 2020), e395–e409. issn: 0016-9013. doi: 10.1093/geront/gnaa064.

[86] M. A. Habib, M. S. Mohktar, S. B. Kamaruzzaman, K. S. Lim, T. M. Pin, and F. Ibrahim. "Smartphone-Based Solutions for Fall Detection and Prevention: Challenges and Open Issues." In: *Sensors* 14.4 (2014), pp. 7181–7208. issn: 1424-8220. doi: 10.3390/s140407181.

[87] C.-L. Chan, Y.-J. Chen, K.-P. Chen, and S.-J. Chiu. "Elderly inpatient fall risk factors: A study of decision tree and logistic regression." In: *The 40th International Conference on Computers Indutrial Engineering*. 2010, pp. 1–6. doi: 10.1109/ICCIE.2010.5668424.

[88] P.-L. Chen, H.-Y. Lin, J. R. Ong, and H.-P. Ma. "Development of a fall-risk assessment profile for community-dwelling older adults by using the National Health Interview Survey in Taiwan." In: *BMC Public Health* 20.1 (2020), p. 234. issn: 1471-2458. doi: 10.1186/s12889-020-8286-8. (Visited on 02/28/2022).

[89] W. Sansrimahachai and M. Toahchoodee. "Mobile-phone based immobility tracking system for elderly care." In: *2016 IEEE Region 10 Conference (TENCON)*. 2016 IEEE Region 10 Conference (TENCON). 2016, pp. 3550–3553. doi: 10.1109/TENCON.2016.7848718.

[90] H. Gonçalves, R. Moreira, A. Rodrigues, and C. Santos. "Finding Parameters around the Abdomen for a Vibrotactile System: Healthy and Patients with Parkinson's Disease." en. In: *Journal of Medical Systems* 42.11 (Oct. 2018), p. 232. issn: 1573-689X. doi: 10.1007/s10916-018-1087-2. (Visited on 03/22/2022).

[91] H. Arndt, S. Burkard, G. Talavera, J. Garcia, D. Castells, M. Codina, J. Hausdorff, A. Mirelman, R. Harte, M. Casey, L. Glynn, M. Di Rosa, L. Rossi, V. Stara, J. Rösevall, C. Rusu, C. Carenas, F. Breuil, E. Reixach, and J. Carrabina. "Real-Time Constant Monitoring of Fall Risk Index by Means of Fully-Wireless Insoles." eng. In: *Studies in Health Technology and Informatics* 237 (2017), pp. 193–197. issn: 1879-8365.

[92] M. Saleh, M. Abbas, and R. B. Le Jeannès. "FallAllD: An Open Dataset of Human Falls and Activities of Daily Living for Classical and Deep Learning Applications." In: *IEEE Sensors Journal* 21.2 (2021), pp. 1849–1858. doi: 10.1109/JSEN.2020.3018335.

[93] J. Klenk, L. Schwickert, L. Palmerini, S. Mellone, A. Bourke, E. A. F. Ihlen, N. Kerse, K. Hauer, M. Pijnappels, M. Synofzik, K. Srulijes, W. Maetzler, J. L. Helbostad, W. Zijlstra, K. Aminian, C. Todd, L. Chiari, C. Becker, and for the FARSEEING Consortium. "The FARSEEING real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls." In: *European Review of Aging and Physical Activity* 13.1 (2016), p. 8. issn: 1861-6909. doi: 10.1186/s11556-016-0168-9.

[94] V. Cotechini, A. Belli, L. Palma, M. Morettini, L. Burattini, and P. Pierleoni. "A dataset for the development and optimization of fall detection algorithms based on wearable sensors." In: *Data in Brief* 23 (2019), p. 103839. issn: 2352-3409. doi: https://doi.org/10.1016/j.dib.2019.103839.

[95] E. Casilari, J. A. Santoyo-Ramón, and J. M. Cano-García. "UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection." In: *Procedia Computer Science* 110 (2017). 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) / 12th International Conference on Future Networks and Communications (FNC 2017) / Affiliated Workshops, pp. 32–39. issn: 1877-0509. doi: https://doi.org/10.1016/j.procs.2017.06.110.

[96] N. F. Ribeiro, J. André, L. Costa, and C. P. Santos. "Development of a Strategy to Predict and Detect Falls Using Wearable Sensors." en. In: *Journal of Medical Systems* 43.5 (Apr. 2019), p. 134. issn: 1573-689X. doi: 10.1007/s10916-019-1252-2. (Visited on 03/22/2022).

[97] J. Figueiredo, S. P. Carvalho, J. P. Vilas-Boas, L. M. Gonçalves, J. C. Moreno, and C. P. Santos. "Wearable Inertial Sensor System towards Daily Human Kinematic Gait Analysis: Benchmarking Analysis to MVN BIOMECH." In: *Sensors* 20.8 (2020). issn: 1424-8220. doi: 10.3390/s20082185.

[98] B. Fida, I. Bernabucci, D. Bibbo, S. Conforto, and M. Schmid. "Varying behavior of different window sizes on the classification of static and dynamic physical activities from a single accelerometer." In: *Medical Engineering & Physics* 37.7 (2015), pp. 705–711. issn: 1350-4533. doi: https://doi.org/10.1016/j.medengphy.2015.04.005.

[99] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas. "Window Size Impact in Human Activity Recognition." In: *Sensors* 14.4 (2014), pp. 6474–6499. issn: 1424-8220. doi: 10.3390/s140406474.

[100] M. H. M. Noor, Z. Salcic, and K. I.-K. Wang. "Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer." In: *Pervasive and Mobile Computing* 38 (2017), pp. 41–59. issn: 1574-1192. doi: https://doi.org/10.1016/j.pmcj.2016.09.009.

[101] J. Zhang, T. E. Lockhart, and R. Soangra. "Classifying Lower Extremity Muscle Fatigue During Walking Using Machine Learning and Inertial Sensors." en. In: *Annals of Biomedical Engineering* 42.3 (Mar. 2014), pp. 600–612. issn: 1573-9686. doi: 10.1007/s10439-013-0917-0. (Visited on 02/21/2022).

[102] J. Bins and B. Draper. "Feature selection from huge feature sets." In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 2. 2001, 159–165 vol.2. doi: 10.1109/ICCV.2001.937619.

[103] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli. "Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach." In: (2017), pp. 1407–1415. doi: 10.1109/ICCV.2017.156.

[104] G. Roffo. *Feature Selection Library (MATLAB Toolbox)*. 2016.

[105] J. Guo, Y. Guo, X. Kong, and R. He. "Unsupervised feature selection with ordinal locality." In: *2017 IEEE International Conference on Multimedia and Expo (ICME)* (2017), pp. 1213–1218.

[106] L. Du and Y.-D. Shen. "Unsupervised Feature Selection with Adaptive Structure Learning." In: *arXiv:1504.00736 [cs]* (Apr. 2015). arXiv: 1504.00736. url: http://arxiv.org/abs/1504.00736 (visited on 02/21/2022).

[107] H. Peng, F. Long, and C. Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1226–1238. doi: 10.1109/TPAMI.2005.159.

[108] *Computational Methods of Feature Selection*. en.

[109] M. Zaffalon and M. Hutter. "Robust Feature Selection by Mutual Information Distributions." In: *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*. UAI'02. Morgan Kaufmann Publishers Inc., 2002, pp. 577–584. isbn: 1558608974.

[110] P. Bradley and O. L. Mangasarian. "Feature Selection via Concave Minimization and Support Vector Machines." In: *Machine Learning Proceedings of the Fifteenth International Conference(ICML '98*. Morgan Kaufmann, 1998, pp. 82–90.

[111] G. Roffo and S. Melzi. "Ranking to Learn: Feature Ranking and Selection via Eigenvector Centrality." In: *arXiv:1704.05409 [cs, stat]* (Apr. 2017). (Visited on 02/21/2022).

[112] "Introduction." en. In: *Principal Component Analysis*. Ed. by I. T. Jolliffe. Springer Series in Statistics. New York, NY: Springer, 2002, pp. 1–9. isbn: 9780387224404. doi: 10.1007/0-387-22440-8_1. (Visited on 02/21/2022).

[113] N. F. Ribeiro, P. Mouta, and C. P. Santos. "Two kinematic data-based approaches for cane event detection." en. In: *Journal of Ambient Intelligence and Humanized Computing* (May 2021). issn: 1868-5145. doi: 10.1007/s12652-021-03313-7. (Visited on 02/21/2022).

[114] J. Suto, S. Oniga, C. Lung, and I. Orha. "Comparison of offline and real-time human activity recognition results using machine learning techniques." en. In: *Neural Computing and Applications* 32.20 (Oct. 2020), pp. 15673–15686. issn: 1433-3058. doi: 10.1007/s00521-018-3437-x.

[115] O. D. Lara and M. A. Labrador. "A Survey on Human Activity Recognition using Wearable Sensors." In: *IEEE Communications Surveys Tutorials* 15.3 (2013), pp. 1192–1209. doi: 10.1109/SURV.2012.110112.00192.

[116] *Ensemble Machine Learning Explained in Simple Terms*. en. July 2020. url: https://jpt.spe.org/twa/ensemble-machine-learning-explained-simple-terms (visited on 03/05/2022).

[117] R. A. Fisher. "138: The Use of Multiple Measurements in Taxonomic Problems." en. In: (1936).

[118] H. A. Abu Alfeilat, A. B. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, and V. S. Prasath. "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review." In: *Big Data* 7.4 (2019), pp. 221–248. doi: 10.1089/big.2018.0175.

[119] C. Kingsford and S. L. Salzberg. “What are decision trees?” en. In: *Nature Biotechnology* 26.9 (Sept. 2008), pp. 1011–1013. issn: 1546-1696. doi: 10.1038/nbt0908-1011. (Visited on 03/18/2022).

[120] A. P. D. Silva and A. Stam. “Discriminant analysis.” In: *Reading and understanding multivariate statistics*. Washington, DC, US: American Psychological Association, 1995, pp. 277–318. isbn: 9781557982735.

[121] S. Ashry, T. Ogawa, and W. Gomaa. “CHARM-Deep: Continuous Human Activity Recognition Model Based on Deep Neural Network Using IMU Sensors of Smartwatch.” English. In: *IEEE Sensors Journal* 20.15 (Aug. 2020), pp. 8757–8770. issn: 1530-437X. doi: 10.1109/JSEN.2020.2985374.

[122] A. E. Frank, A. Kubota, and L. D. Riek. “Wearable activity recognition for robust human-robot teaming in safety-critical environments via hybrid neural networks.” In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019, pp. 449–454. doi: 10.1109/IROS40897.2019.8968615.

[123] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. en. Ed. by F. Bach. Adaptive Computation and Machine Learning series. Cambridge, MA, USA: MIT Press, Nov. 2016. isbn: 9780262035613.

[124] E. Allibhai. *Holdout vs. Cross-validation in Machine Learning*. en. Oct. 2018. url: https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f (visited on 03/25/2022).

[125] F. Kayaalp and M. Başarslan. “Open Source Data Mining Programs: A Case Study on R.” In: *Duzce University Journal of Science and Technolgy* 6 (Apr. 2018), pp. 455–468.

[126] J. Brownlee. *A Gentle Introduction to k-fold Cross-Validation*. en-US. May 2018. url: https://machinelearningmastery.com/k-fold-cross-validation/ (visited on 02/22/2022).

[127] J. Ashfaque and A. Iqbal. “Introduction to Support Vector Machines and Kernel Methods.” In: (2019).

[128] T. Agrawal. “Hyperparameter Optimization Using Scikit-Learn.” en. In: *Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient*. Ed. by T. Agrawal. Berkeley, CA: Apress, 2021, pp. 31–51. isbn: 9781484265796. doi: 10.1007/978-1-4842-6579-6_2.

[129] A. Mishra. *Metrics to Evaluate your Machine Learning Algorithm*. en. May 2020. url: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 (visited on 02/23/2022).

[130] M. Sunasra. *Performance Metrics for Classification problems in Machine Learning*. en. Feb. 2019.

[131] D. Chicco and G. Jurman. “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.” In: *BMC Genomics* 21.1 (Jan. 2020), p. 6. issn: 1471-2164. doi: 10.1186/s12864-019-6413-7.

[132] T. Althobaiti, S. Katsigiannis, and N. Ramzan. “Triaxial Accelerometer-Based Falls and Activities of Daily Life Detection Using Machine Learning.” In: *Sensors* 20.13 (2020). issn: 1424-8220. doi: 10.3390/s20133777.

[133] A. Ignatov. “Real-time human activity recognition from accelerometer data using Convolutional Neural Networks.” In: *Applied Soft Computing* 62 (2018), pp. 915–922. issn: 1568-4946. doi: https://doi.org/10.1016/j.asoc.2017.09.027.

[134] B. Cates, T. Sim, H. M. Heo, B. Kim, H. Kim, and J. H. Mun. “A Novel Detection Model and Its Optimal Features to Classify Falls from Low- and High-Acceleration Activities of Daily Life Using an Insole Sensor System.” In: *Sensors* 18.4 (2018). issn: 1424-8220. doi: 10.3390/s18041227. url: https://www.mdpi.com/1424-8220/18/4/1227.

[135] S. Ranakoti, S. Arora, S. Chaudhary, S. Beetan, A. Sandhu, P. Khandnor, and P. Saini. “Human Fall Detection System over IMU Sensors Using Triaxial Accelerometer: ICCI-2017.” In: Jan. 2019, pp. 495–507. isbn: 978-981-13-1131-4. doi: 10.1007/978-981-13-1132-1_39.

[136] S. Zhao, W. Li, W. Niu, R. Gravina, and G. Fortino. “Recognition of human fall events based on single tri-axial gyroscope.” In: Mar. 2018, pp. 1–6. doi: 10.1109/ICNSC.2018.8361365.

[137] M. Iosa, T. Marro, S. Paolucci, and D. Morelli. “Stability and harmony of gait in children with cerebral palsy.” English. In: *Research in Developmental Disabilities* 33.1 (Jan. 2012), pp. 129–135. issn: 0891-4222. doi: 10.1016/j.ridd.2011.08.031.

[138] A. T. Özdemir and B. Barshan. “Detecting Falls with Wearable Sensors Using Machine Learning Techniques.” In: *Sensors* 14.6 (2014), pp. 10691–10708. issn: 1424-8220. doi: 10.3390/s140610691.

[139] F. Bianchi, S. J. Redmond, M. R. Narayanan, S. Cerutti, and N. H. Lovell. “Barometric Pressure and Triaxial Accelerometry-Based Falls Event Detection.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.6 (2010), pp. 619–627. doi: 10.1109/TNSRE.2010.2070807.

[140] M. Kangas, A. Konttila, I. Winblad, and T. Jämsä. “Determination of simple thresholds for accelerometry-based parameters for fall detection.” In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2007 (Feb. 2007), pp. 1367–70. doi: 10.1109/IEMBS.2007.4352552.

[141] I. P. E. S. Putra, J. Brusey, E. Gaura, and R. Vesilo. “An Event-Triggered Machine Learning Approach for Accelerometer-Based Fall Detection.” In: *Sensors* 18.1 (2018). issn: 1424-8220. doi: 10.3390/s18010020. url: https://www.mdpi.com/1424-8220/18/1/20.

Appendix A

5 fold - 1 repetition Cross-Validation Results

In this appendix, the complete results regarding the 5-1 Cross-Validation step explained in chapter 6 are presented.

Table 27: Comparison of the best classification results (ACC, Sens, Spec, Prec, F1S, MCC), selected by the highest MCC, for the K-NN machine learning classifier trained with the features ranked by the various feature selection method (fist stage 5-1 k-fold cross-validation step)

FSM	Nº of Features	ACC (%)	Sens (%)	Spec (%)	Precision (%)	F1S (%)	MCC (%)
ILFS	193	93,25	84,05	99,61	86,52	85,23	84,88
UFSOL	85	92,62	81,24	99,58	83,86	82,49	82,11
FSASL	70	91,49	81,39	99,51	83,66	82,48	82.02
MRMR	97	93,03	84,40	99,60	86,27	85,30	84,93
Relief-F	85	93,63	84,17	99,64	86,80	85,43	85,10
MutInfFS	56	88,05	74,79	99,32	78,08	76,36	75,72
FSV	55	88,28	75,00	99,33	78,33	76,58	75,97
CFS	110	91,61	80,79	99,52	83,69	82,16	81,73
LASSO	70	89,99	77,16	99,43	80,83	78,88	78.37
PCA	85	92,99	84,08	99,60	86,01	85,01	84,63

Table 28: Comparison of the best classification results (ACC, Sens, Spec, Prec, F1S, MCC), selected by the highest MCC, for the DA Linear classifier trained with the features ranked by the various feature selection method (fist stage 5-1 k-fold cross-validation step)

FSM	Nº of Features	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
ILFS	72	46,62	36,79	97,11	32,87	30,93	30,21
UFSOL	101	43,82	34,23	97,01	32,51	28,82	28,38
FSASL	60	47,50	41,51	97,24	36,20	33,39	33,32
MRMR	61	50,81	41,68	97,40	37,20	34,48	34,39
Relief-F	66	56,89	45,84	97,67	38,03	38,48	37,85
MutInfFS	62	45,43	36,89	97,04	31,92	30,69	29,73
FSV	55	42,95	36,58	96,94	31,71	29,96	29,09
CFS	104	49,74	43,79	97,30	36,77	35,70	35,34
LASSO	73	45,43	37,09	97,09	34,34	31,47	30,88
PCA	65	48,74	43,73	97,30	36,44	34,27	34,32

Table 29: Comparison of the best classification results (ACC, Sens, Spec, Prec, F1S, MCC), selected by the highest MCC, for the DA Quadratic classifier trained with the features ranked by the various feature selection method (fist stage 5-1 k-fold cross-validation step)

FSM	Nº of Features	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
ILFS	75	28,37	34,91	96,26	30,75	25,88	26,07
UFSOL	109	30,14	33,85	96,35	32,24	25,04	25,72
FSASL	81	43,49	43,09	97,06	36,65	32,44	33,03
MRMR	61	43,35	42,94	97,04	36,95	32,18	33,01
Relief-F	55	54,01	47,87	97,55	38,37	39,32	38,72
MutInfFS	97	31,58	36,09	96,42	32,37	26,49	26,99
FSV	97	31,58	36,09	96,42	32,37	26,49	26,99
CFS	109	32,50	38,93	96,46	33,33	28,03	28,69
LASSO	77	34,88	36,32	96,61	34,53	27,84	28,50
PCA	67	48,74	46,65	97,33	39,02	34,83	35,85

Table 30: Comparison of the best classification results (ACC, Sens, Spec, Prec, F1S, MCC), selected by the highest MCC, for the Decision Trees classifier trained with the features ranked by the various feature selection method (fist stage 5-1 k-fold cross-validation step)

FSM	Nº of Features	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
ILFS	109	85,61	66,68	99,18	68,76	67,66	66,88
UFSOL	109	86,59	68,23	99,24	69,91	69,04	68,30
FSASL	85	88,03	69,98	99,33	72,11	70,99	70.35
MRMR	96	87,56	69,33	99,30	71,11	70,18	69,50
Relief-F	109	87,41	69,20	99,29	71,02	70,07	69,38
MutInfFS	110	83,45	62,79	99,06	64,90	63,79	62,89
FSV	109	83,47	62,82	99,06	64,95	63,83	62,92
CFS	109	86,42	67,32	99,23	69,23	68,23	67,49
LASSO	101	86,15	67,20	99,22	69,03	68,07	67,31
PCA	74	88,22	70,37	99,34	72,29	71,29	70,65

Table 31: Comparison of the best classification results (ACC, Sens, Spec, Prec, F1S, MCC), selected by the highest MCC, for the Ensemble Learning classifier trained with the features ranked by the various feature selection method (fist stage 5-1 k-fold cross-validation step)

FSM	Nº of Features	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
ILFS	175	94,04	81,07	99,65	89,98	84,69	84,76
UFSOL	189	94,01	80,90	99,64	89,89	84,54	84,62
FSASL	81	94,21	81,65	99,66	90,09	85,25	85,23
MRMR	110	94,19	81,30	99,65	90,15	85,00	85,03
Relief-F	125	94,09	81,25	99,65	90,31	84,90	85,00
MutInfFS	195	93,93	80,79	99,64	89,95	84,48	84,57
FSV	199	93,91	80,70	99,64	89,77	84,38	84,45
CFS	199	93,93	80,79	99,64	89,84	84,45	84,53
LASSO	69	92,88	78,50	99,58	88,32	82,40	82,48
PCA	65	94,59	82,22	99,68	90,54	85,80	85,78

Appendix B

Final tests K-NN and Ensemble Learning Results

In this appendix it is represented the complete performance metrics results per class regarding chapter's 6 final test with the Ensemble Learning and K-NN classifiers.

Table 32: Final test Results per class for Ensemble Learning classifier with the first 65 features ranked by PCA for a 0,5s window

Class	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
Walking	99,17	99,17	98,63	96,82	97,98	97,13
Standing	97,34	97,34	99,22	95,23	96,27	95,68
Siting	97,99	97,99	99,78	97,46	97,73	97,53
Laying	99,11	99,11	99,59	98,24	98,67	98,37
UpStairs	92,63	92,63	99,87	97,04	94,78	94,57
DownStairs	94,31	94,31	99,89	97,53	95,89	95,72
Jumping	89,57	89,57	99,98	96,25	92,79	92,81
Jogging	98,35	98,35	99,9	98,82	98,59	98,47
Fall Fowards	81,30	81,3	99,88	88,98	84,97	84,88
Fall Backwards	81,02	81,02	99,93	90,44	85,47	85,48
Lateral Fall	82,46	82,46	99,88	85,18	83,8	83,68
Syncope Fall	61,86	61,86	99,98	90,88	73,62	74,93
Lying-to-Stand	93,23	93,23	99,91	94,92	94,07	93,97
Stand-to-Sit	84,27	84,27	99,75	84,72	84,5	84,24
Sit-to-Stand	87,93	87,93	99,9	92,45	90,13	90,02
Pick objects from ground	84,46	84,46	99,97	97,34	90,45	90,57
Stand-to-Lying	86,45	86,45	99,95	95,11	90,57	90,57
Change position (Lying)	94,40	94,4	99,95	94,81	94,6	94,55
Turning	82,31	82,31	99,96	92,26	87	87,08
Bending	90,68	90,68	99,98	97,26	93,85	93,86

Table 33: Final test Results per class for K-NN classifier with the first 85 features ranked by Relief-F for a 0,5s window

Class	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
Walking	99,11	99,11	99,50	98,82	98,96	98,52
Standing	96,56	96,55	99,45	96,43	96,49	95,95
Siting	97,35	97,35	99,75	97,05	97,20	96,95
Laying	98,29	98,29	99,54	97,97	98,13	97,71
UpStairs	98,63	98,63	99,92	98,22	98,43	98,35
DownStairs	98,39	98,39	99,93	98,47	98,43	98,36
Jumping	93,88	93,88	99,96	93,02	93,45	93,41
Jogging	99,36	99,36	99,97	99,61	99,49	99,44
Fall Fowards	85,79	85,79	99,85	87,36	86,57	86,41
Fall Backwards	87,66	87,66	99,90	88,31	87,98	87,88
Lateral Fall	85,93	85,93	99,91	88,78	87,33	87,24
Syncope Fall	82,33	82,33	99,97	86,92	84,57	84,55
Lying-to-Stand	93,40	93,40	99,88	93,18	93,29	93,17
Stand-to-Sit	88,32	88,32	99,83	89,86	89,08	88,90
Sit-to-Stand	88,57	88,57	99,88	91,33	89,93	89,80
Pick objects from ground	95,48	95,48	99,94	95,52	95,50	95,45
Stand-to-Lying	91,67	91,67	99,92	93,27	92,46	92,37
Change position (Lying)	92,76	92,76	99,93	93,61	93,18	93,11
Turning	88,54	88,54	99,96	92,41	90,43	90,40
Bending	95,98	95,98	99,96	95,72	95,85	95,81

Appendix C

Ensemble Learning Best window: Results per Class and Confusion Matrix

The results presented in this appendix refer to the performance evaluation metrics attained per class for the best combination found in Chapter X Section Y, the Ensemble Learning classifier and the 65 most important features ranked by PCA, with a 0,5s window.

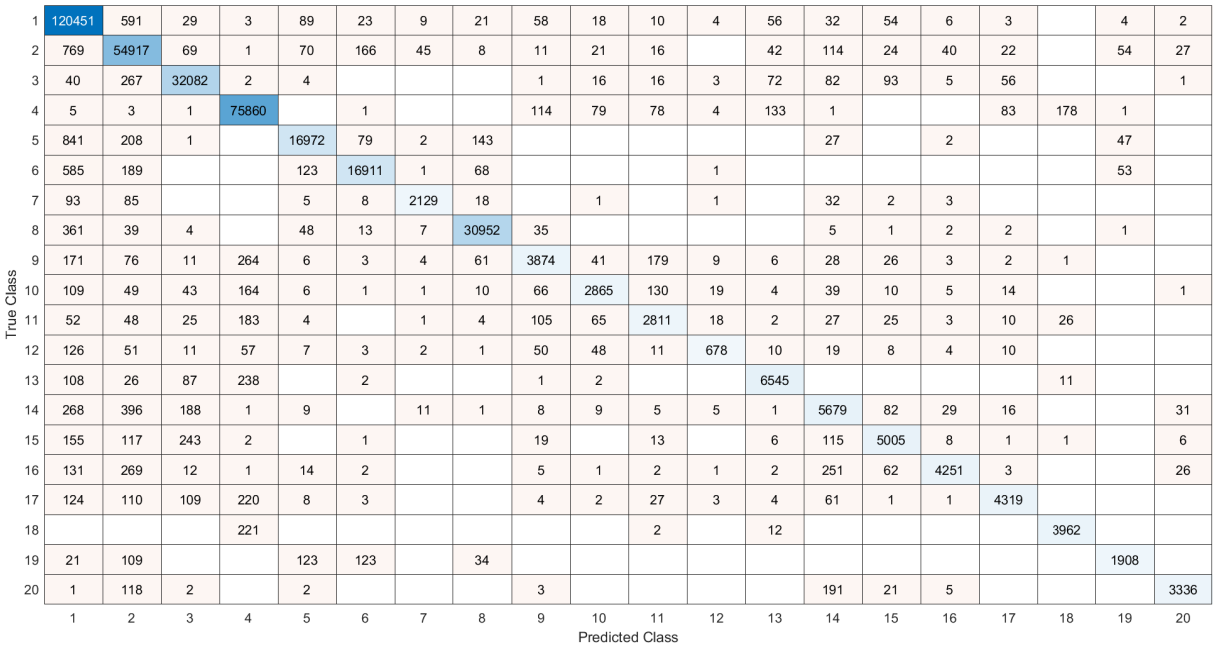


Figure 41: Confusion matrix

Table 34: Final test Results per class for Ensemble Learning classifier with the first 65 features ranked by PCA - 0,5s window

Class	ACC (%)	Sens (%)	Spec (%)	Prec (%)	F1S (%)	MCC (%)
Walking	99,17	99,17	98,63	96,82	97,98	97,13
Standing	97,34	97,34	99,22	95,23	96,27	95,68
Siting	97,99	97,99	99,78	97,46	97,73	97,53
Laying	99,11	99,11	99,59	98,24	98,67	98,37
UpStairs	92,63	92,63	99,87	97,04	94,78	94,57
DownStairs	94,31	94,31	99,89	97,53	95,89	95,72
Jumping	89,57	89,57	99,98	96,25	92,79	92,81
Jogging	98,35	98,35	99,9	98,82	98,59	98,47
Fall Fowards	81,30	81,3	99,88	88,98	84,97	84,88
Fall Backwards	81,02	81,02	99,93	90,44	85,47	85,48
Lateral Fall	82,46	82,46	99,88	85,18	83,8	83,68
Syncope Fall	61,86	61,86	99,98	90,88	73,62	74,93
Lying-to-Stand	93,23	93,23	99,91	94,92	94,07	93,97
Stand-to-Sit	84,27	84,27	99,75	84,72	84,5	84,24
Sit-to-Stand	87,93	87,93	99,9	92,45	90,13	90,02
Pick objects from ground	84,46	84,46	99,97	97,34	90,45	90,57
Stand-to-Lying	86,45	86,45	99,95	95,11	90,57	90,57
Change position (Lying)	94,40	94,4	99,95	94,81	94,6	94,55
Turning	82,31	82,31	99,96	92,26	87	87,08
Bending	90,68	90,68	99,98	97,26	93,85	93,86

Appendix D

Best Feature Subsets

In this appendix the complete list of the selected feature subset by the two feature selection method which presented the best performance results in Chapter 6 are presented. Table 35 shows the 85 selected features by Relief-F. Table 36 present the 65 main features, ranked by the PCA feature selection method.

Table 35: Complete list of the first 85 ranked features selected by the feature selection method Relief-F, wich produced the best results for K-NN classifier in Chapter 6, with feature label, its description and the corresponding reference.

Feature Ranking	Name	Description	Reference
1	CorrelationAcc_V_AP	Correlation Between Accelerantion Vertical and Antero Posterior axis	[134]
2	CorrelationGyr_V_AP	Correlation Between Angular Velocity Vertical and Antero Posterior axis	[134]
3	CorrelationGyr_ML_AP	Correlation Between Angular Velocity Medial Lateral and Anteror Posterior axis	[134]
4	CorrelationGyr_V_ML	Correlation Between Angular Velocity Vertical and Medial Lateral axis	[134]
5	CorrelationAcc_ML_AP	Correlation Between Accelerantion Medial Lateral and Anteror Posterior axis	[134]
6	CorrelationAcc_V_ML	Correlation Between Accelerantion Vertical and Medial Lateral axis	[134]

7	RAAcc_AP	Resultant of Average Acceleration (Antero Posterior axis)	[135]
8	GravityComponentAcc_AP	Gravity component of Acceleration (Antero Posterior Axis)	N/A
9	SSC_Acc_AP	Slope Changes of Acceleration (Antero Posterior Axis)	[64]
10	SSC_Acc_ML	Slope Changes of Acceleration (Medio Lateral Axis)	[64]
11	SSC_Acc_V	Slope Changes of Acceleration (Vertical Axis)	[64]
12	MeanAcc_AP	Mean Acceleration (Anterior posterior axis)	[66]
13	MinSVMAcc	Minimum SVM of acceleration	[64]
14	RAC_AP	Resultant angle change (Antero Posterior axis)	[136]
15	sk_SVMAcc	Skewness of SVM of acceleration	[66]
16	sk_V_Acc	Skewness of acceleration (Vertical axis)	[66]
17	RAAcc_ML	Resultant of Average Acceleration (Medio Lateral axis)	[135]
18	SSC_SVMAcc	Slope Changes of Acceleration SVM	[64]
19	sk_AP_Gyr	Skewness of angular velocity (Antero Posterior axis)	[66]
20	sk_ML_Gyr	Skewness of angular velocity (Medio Lateral axis)	[66]
21	RAAcc_V	Resultant of Average Acceleration (Vertical axis)	[135]
22	RSTD_AP	Resultant of Standard Deviation (Antero Posterior axis)	[135]
23	ZC_Acc_ML	Zero Crossings of Acceleration (Medio Lateral Axis)	[64]
24	MeanGyr_ML	Mean Angular Velocity (Medio Lateral axis)	[66]
25	sk_AP_Acc	Skewness of acceleration (Antero Posterior axis)	[66]
26	kur_SVMAcc	Kurtosis of SMV of Accelerarion	[66]
27	sk_V_Gyr	Skewness of angular velocity (Vertical axis)	[66]

28	Pk_Freq_SVMAcc	Peak Frequency of Acceleration SVM	[64]
29	sk_ML_Acc	Skewness of acceleration (Medio Lateral axis)	[66]
30	MeanAcc_ML	Mean Acceleration (Medio lateral axis)	[66]
31	RAAcc	SVM of Resultant of Average Acceleration	[135]
32	RAC_ML	Resultant angle change (Medio Lateral axis)	[136]
33	RSTD	SVM of Resultant of Standard Deviation	[135]
34	RMS_Acc_AP	Root Mean Square of Acceleration (Antero posterior axis)	[64, 66]
35	ZC_Gyr_AP	Zero Crossings of Angular Velocity (Antero Posterior Axis)	[64]
36	RI_Gyr_ML	Ratio index of Angular Velocity (Medio Lateral axis)	[137]
37	ZC_Gyr_ML	Zero Crossings of Angular Velocity (Medio Lateral Axis)	[64]
38	kur_V_Acc	Kurtosis of acceleration (Vertical axis)	[66]
39	SSC_Gyr_V	Slope Changes of Angular Velocity (Vertical Axis)	[64]
40	kur_AP_Gyr	Kurtosis of angular velocity (Antero posterior axis)	[66]
41	MeanAcc_V	Mean Acceleration (Vertical axis)	[66]
42	RAC_V	Resultant angle change (Vertical axis)	[136]
43	SSC_Gyr_AP	Slope Changes of Angular Velocity (Antero Posterior Axis)	[64]
44	SSC_Gyr_ML	Slope Changes of Angular Velocity (Medio Lateral Axis)	[64]
45	RSTD_V	Resultant of Standard Deviation (Vercial axis)	[135]
46	kur_V_Gyr	Kurtosis of angular velocity (Vertical axis)	[66]
47	kur_ML_Gyr	Kurtosis of angular velocity (Medio Lateral axis)	[66, 138]
48	kur_ML_Acc	Kurtosis of acceleration (Medio Lateral axis)	[66, 138]
49	Pk_Amp_Acc_AP	Peak Frequency's Magnitude of Acceleration (Antero Posterior Axis)	[64]

50	kur_AP_Acc	Kurtosis of acceleration (Antero Posterior axis)	[66, 138]
51	sk_SVMGyr	Skewness of SVM of angular velocity	[66]
52	ZC_Acc_V	Zero Crossings of Acceleration (Vertical Axis)	[64]
53	SMA	Signal Magnitude Area	[139]
54	RMS_Acc_V	Root Mean Square of Acceleration (Vertical axis)	[64, 66]
55	SSC_SVMGyr	Slope Changes of Angular Velocity SVM	[64]
56	RMS_Acc_ML	Root Mean Square of Acceleration (Medio Lateral Axis)	[64, 66]
57	kur_SVMGyr	SVM of Kurtosis of angular velocity	[66, 138]
58	RMS_Gyr_ML	Root Mean Square of angular velocity (Medio Lateral Axis)	[64, 66]
59	Pk_Amp_SVMGyr	Peak Frequency's Magnitude of Angular Velocity SVM	[64]
60	StdSVMGyr	Standard Deviation of Angular Velocity SVM	[64, 66]
61	MAD	Magnitude of Angular Displacement	[139]
62	RMS_SVMGyr	Root Mean Square of SVM of Angular Velocity	[64, 66]
63	RMS_Gyr_AP	Root Mean Square of SVM of Acceleration	[64, 66]
64	ZC_Gyr_V	Zero Crossings of Angular Velocity (Vertical Axis)	[64]
65	MinGyr_V	Minimum angular velocity (Vertical axis)	[64]
66	MeanGyr_AP	Mean Angular Velocity (Anterior posterior axis)	[66]
67	Pk_Amp_Gyr_ML	Peak Frequency's Magnitude of Angular Velocity (Medio Lateral Axis)	[64]
68	Mean_PSD_Acc_AP	Mean Frequency of Acceleration (Antero Posterior Axis)	[64]
69	StdAcc_AP	Standard Deviation of Acceleration (Anteror Posterior axis)	[64, 66]
70	MeanSVMGyr	Mean SVM of Angular Velocity	[66]
71	Total_Angular_Change	Total angular change	[140]

72	StdGyr_V	Standard Deviation of Angular Velocity (Vertical axis)	[64, 66]
73	MinGyr_ML	Minimum angular velocity (Medio Lateral axis)	[64]
74	StdGyr_ML	Standard Deviation of Angular Velocity (Medio Lateral axis)	[64, 66]
75	Pk_Amp_Acc_ML	Peak Frequency's Magnitude of Acceleration (Medio Lateral Axis)	[64]
76	StdGyr_AP	Standard Deviation of Angular Velocity (Antero Posterior axis)	[64, 66]
77	RSTD_ML	Resultant of Standard Deviation (Medio Lateral axis)	[135]
78	RI_GyrSVM	Ratio Index of SVM of Angular Velocity	[137]
79	MaxGyr_ML	Maximum angular velocity (Medio Lateral axis)	[64]
80	RI_Gyr_AP	Ratio index of Angular Velocity (Antero Posterior axis)	[137]
81	MinSVMGyr	Minimum SVM of angular velocity	[64]
82	MaxSVMGyr	Maximum SVM of angular velocity	[64]
83	RI_Acc_V	Ratio Index of Acceleration (Vertical axis)	[137]
84	PPV_Gyr_AP	Peak-to-peak values of Angular Velocity (Antero Posterior axis)	[137]
85	RMS_Gyr_V	Root Mean Square of Angular Velocity (Vertical axis)	[64, 67]

Table 36: Complete list of the first 65 ranked features selected by the feature selection method PCA, wich produced the best results for the Ensemble Learning classifier in Chapter 6, with feature label, its description and the corresponding reference.

Feature Ranking	Name	Description	Reference
1	sk_V_Acc	Skewness of acceleration (Vertical axis)	[66]
2	RI_Gyr_AP	Ratio index of Angular Velocity (Antero Posterior axis)	[137]
3	Pk_Freq_SVMAcc	Peak Frequency of Acceleration SVM	[64]

4	MeanAcc_ML	Mean Acceleration (Medio lateral axis)	[66]
5	RAC_ML	Resultant angle change (Medio Lateral axis)	[136]
6	MeanAcc_AP	Mean Acceleration (Anterior posterior axis)	[66]
7	RAC_AP	Resultant angle change (Antero Posterior axis)	[136]
8	GravityComponentAcc_AP	Gravity component of Acceleration (Antero Posterior Axis)	N/A
9	SSC_Acc_V	Slope Changes of Acceleration (Vertical Axis)	[64]
10	RSTD	SVM of Resultant of Standard Deviation	[135]
11	SSC_SVMAcc	Slope Changes of Acceleration SVM	[64]
12	RAAcc_AP	Resultant of Average Acceleration (Antero Posterior axis)	[135]
13	Pk_Freq_Acc_V	Peak Frequency of Acceleration (Vertical Axis)	[64]
14	MinSVMAcc	Minimum SVM of acceleration	[64]
15	SSC_Acc_AP	Slope Changes of Acceleration (Antero Posterior Axis)	[64]
16	ZC_Gyr_ML	Zero Crossings of Angular Velocity (Medio Lateral Axis)	[64]
17	RSTD_V	Resultant of Standard Deviation (Vercial axis)	[135]
18	RMS_Acc_V	Root Mean Square of Acceleration (Vertical axis)	[64, 66]
19	MeanAcc_V	Mean Acceleration (Vertical axis)	[66]
20	RAC_V	Resultant angle change (Vertical axis)	[136]
21	RAAcc_ML	Resultant of Average Acceleration (Medio Lateral axis)	[135]
22	SSC_Acc_ML	Slope Changes of Acceleration (Medio Lateral Axis)	[64]
23	RSTD_AP	Resultant of Standard Deviation (Antero Posterior axis)	[135]

24	Pk_Freq_Acc_ML	Peak Frequency of Acceleration (Medio Lateral Axis)	[64]
25	Pk_Freq_Acc_AP	Peak Frequency of Acceleration (Antero Posterior Axis)	[64]
26	MaxAcc_V	Maximum acceleration (Vertical axis)	[64]
27	Mean_PSD_Acc_V	Mean Frequency of Acceleration (Vertical Axis)	[64]
28	MinGyr_AP	Minimum angular velocity (Antero posterior axis)	[64]
29	RAAcc	SVM of Resultant of Average Acceleration	[135]
30	RMS_SVMAcc	Root Mean Square of Angular Velocity (Vertical axis)	[64, 66]
31	SSC_SVMGyr	Slope Changes of Angular Velocity SVM	[64]
32	MeanSVMAcc	Mean SVM of Acceleration	[66]
33	ZC_Acc_ML	Zero Crossings of Acceleration (Medio Lateral Axis)	[64]
34	SMA	Signal Magnitude Area	[139]
35	ZC_Gyr_AP	Zero Crossings of Angular Velocity (Antero Posterior Axis)	[64]
36	RAAcc_V	Resultant of Average Acceleration (Vertical axis)	[135]
37	GravityComponentAcc_V	Gravity component of Acceleration (Vertical Axis)	N/A
38	RMS_Acc_ML	Root Mean Square of Acceleration (Medio Lateral axis)	[64, 66]
39	MeanGyr_V	Mean Angular Velocity (Vertical axis)	[66]
40	MinAcc_ML	Minimum acceleration (Medio Lateral axis)	[64]
41	SSC_Gyr_V	Slope Changes of Angular Velocity (Vertical Axis)	[64]
42	CorrelationGyr_V_AP	Correlation Between Angular Velocity Vertical and Antero Posterior axis	[134]
43	Mean_PSD_SVMAcc	Mean Frequency of Acceleration SVM	[64]

44	GravityComponentAcc_ML	Gravity component of Acceleration (Medio Lateral Axis)	N/A
45	RAC	SVM of Resultant angle change	[136]
46	SSC_Gyr_AP	Slope Changes of Angular Velocity (Antero Posterior Axis)	[64]
47	RSTD_ML	Resultant of Standard Deviation (Medio Lateral axis)	[135]
48	MinAcc_AP	Minimum acceleration (Antero posterior axis)	[64]
49	ZC_Gyr_V	Zero Crossings of Angular Velocity (Vertical Axis)	[64]
50	SSC_Gyr_ML	Slope Changes of Angular Velocity (Medio Lateral Axis)	[64]
51	EMA	Acceleration exponential moving average	[141]
52	Gyr_V	Vertical angular velocity	N/A
53	MaxAcc_AP	Maximum acceleration (Antero posterior axis)	[64]
54	RMS_Acc_AP	Root Mean Square of Acceleration (Antero posterior axis)	[64, 66]
55	MaxGyr_V	Maximum angular velocity (Vertical axis)	[64]
56	Acc_V	Vertical Acceleration	N/A
57	MaxAcc_ML	Maximum acceleration (Medio Lateral axis)	[64]
58	SVMAcc	SVM of Acceleration	[140]
59	RI_Acc_V	Ratio Index of Acceleration (Vertical axis)	[137]
60	MeanGyr_AP	Mean Angular Velocity (Anterior posterior axis)	[66]
61	MeanGyr_ML	Mean Angular Velocity (Medio Lateral axis)	[66]
62	Pk_Freq_Gyr_ML	Peak Frequency of Angular Velocity (Medio Lateral Axis)	[64]
63	Mean_PSD_Acc_ML	Mean Frequency of Acceleration (Medio Lateral Axis)	[64]
64	Pk_Amp_Acc_V	Peak Frequency's Magnitude of Acceleration (Vertical Axis)	[64]

65	RI_Gyr_ML	Ratio index of Angular Velocity (Medio Lateral axis)	[137]

Appendix E

Deep Learning Confusion Matrix Results

In this appendix are presented the confusion matrices that refer to the results (Chapter 6) attained with the different Deep Learning architectures referred in Chapter 5 - CNN, LSTM, BiLSTM and CNN-LSTM.

E.1 Relief-F Feature Selection Method

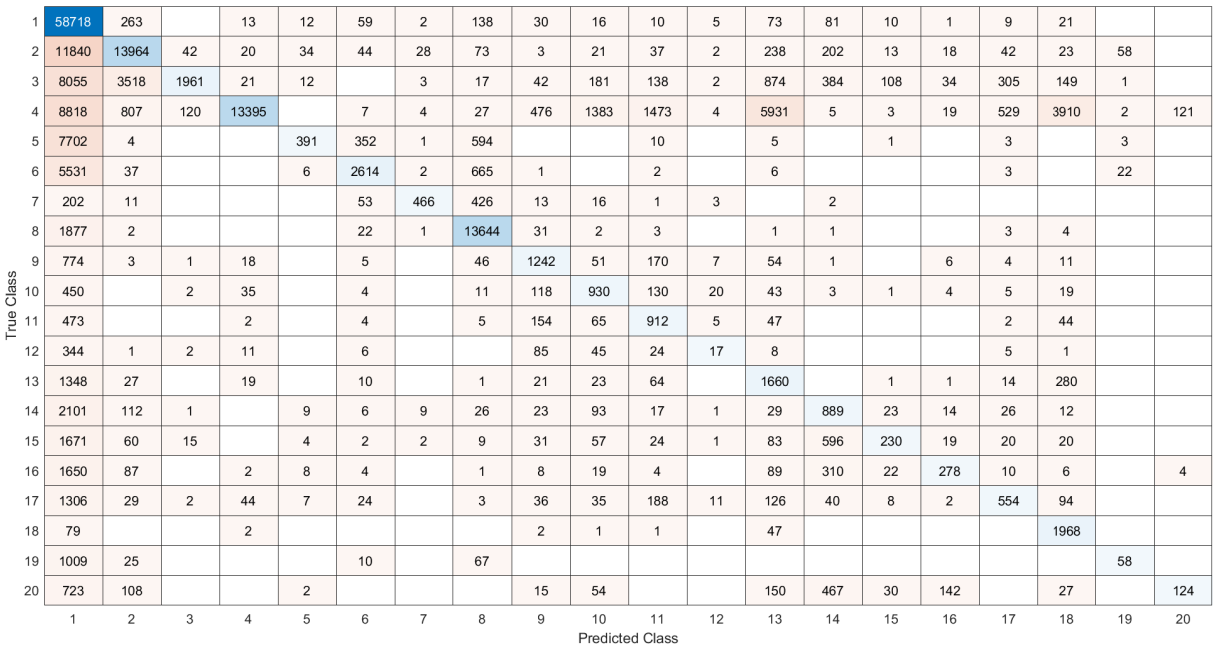


Figure 42: Confusion matrix results for the CNN architecture with Relief-F ranked features.

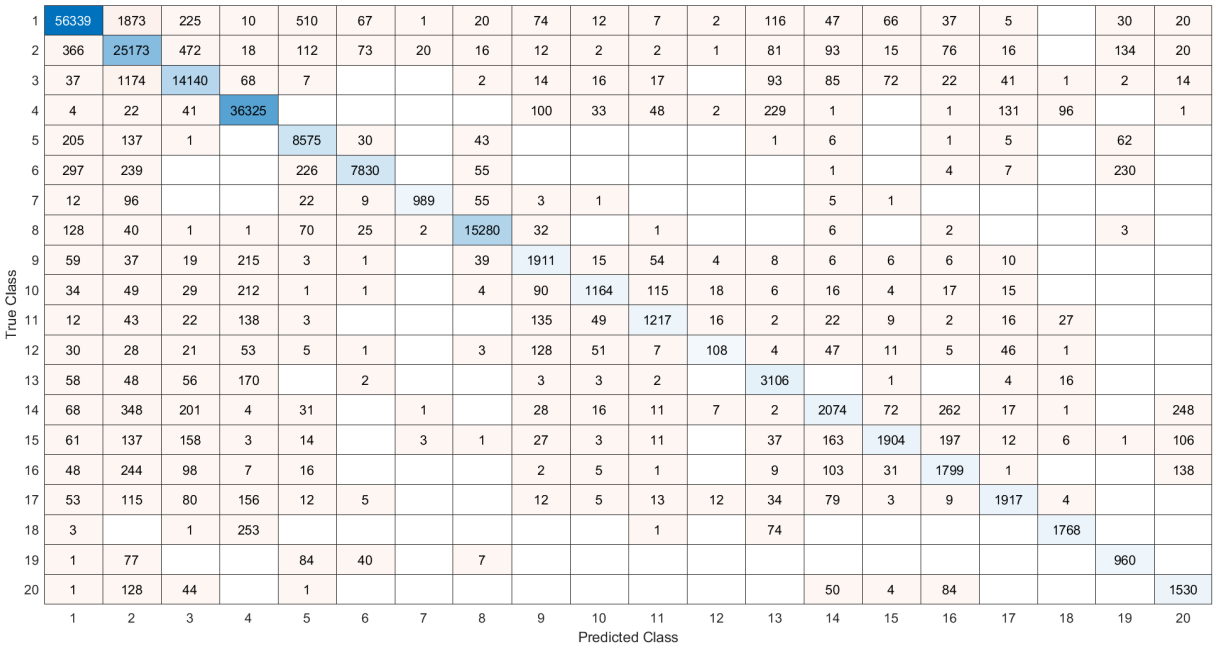


Figure 43: Confusion matrix results for the LSTM architecture with Relief-F ranked features.

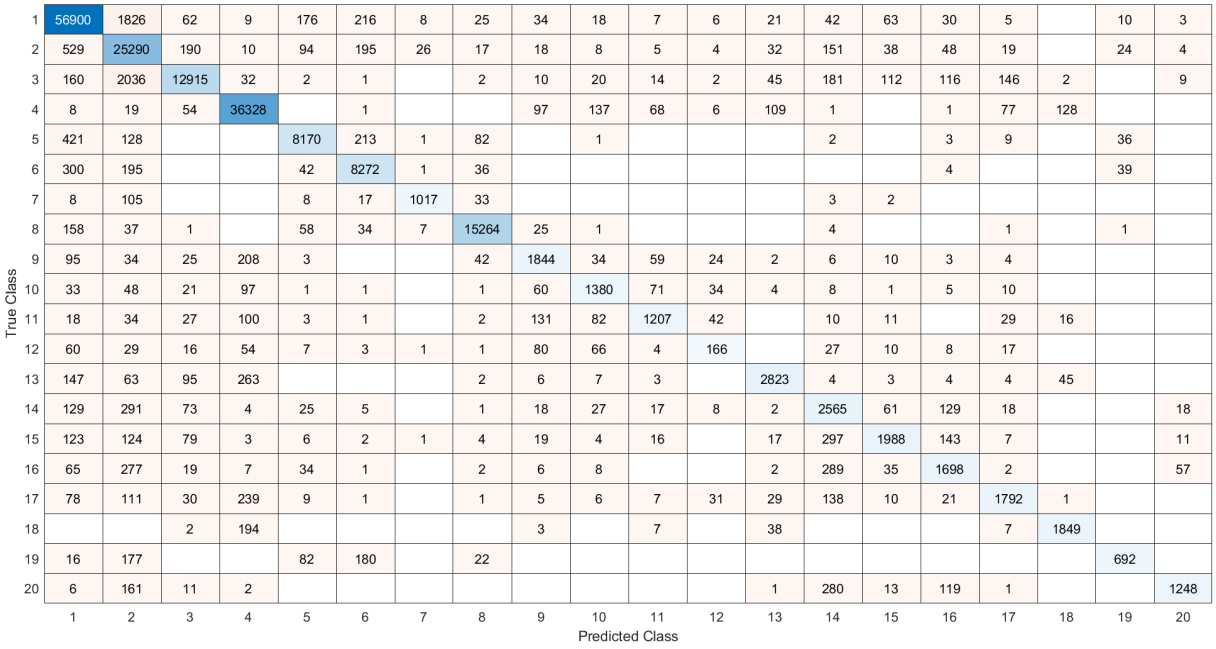


Figure 44: Confusion matrix results for the Bi-LSTM architecture with Relief-F ranked features.

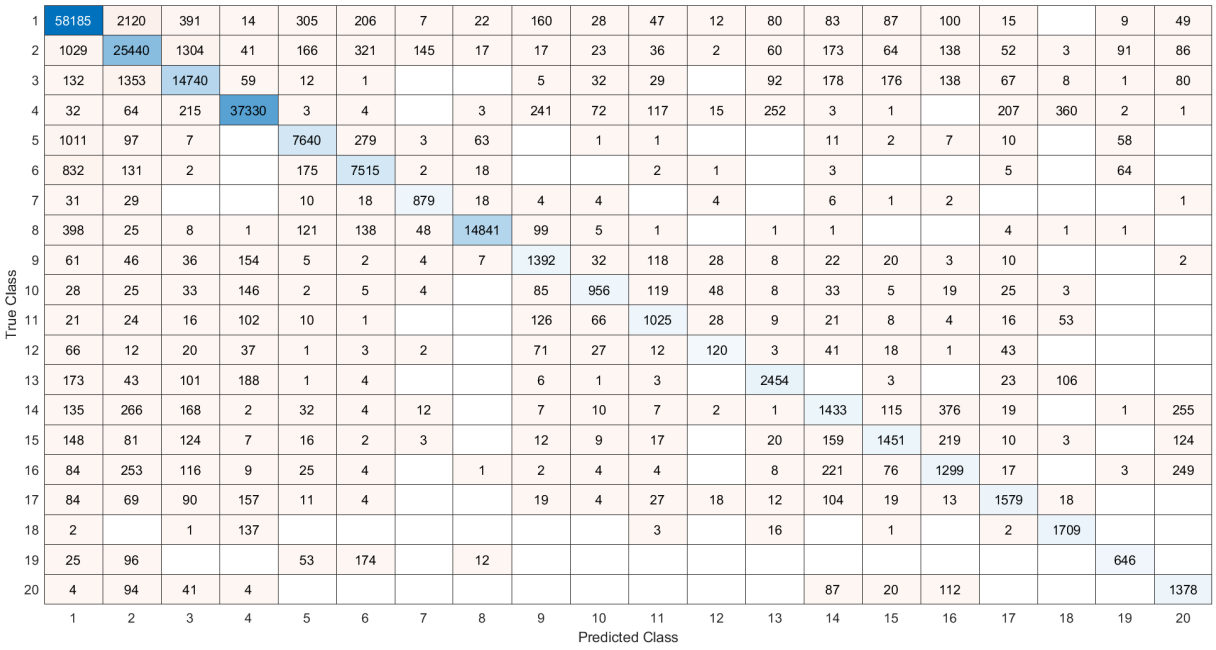


Figure 45: Confusion matrix results for the CNN-LSTM architecture with Relief-F ranked features.

E.2 PCA Feature Selection Method

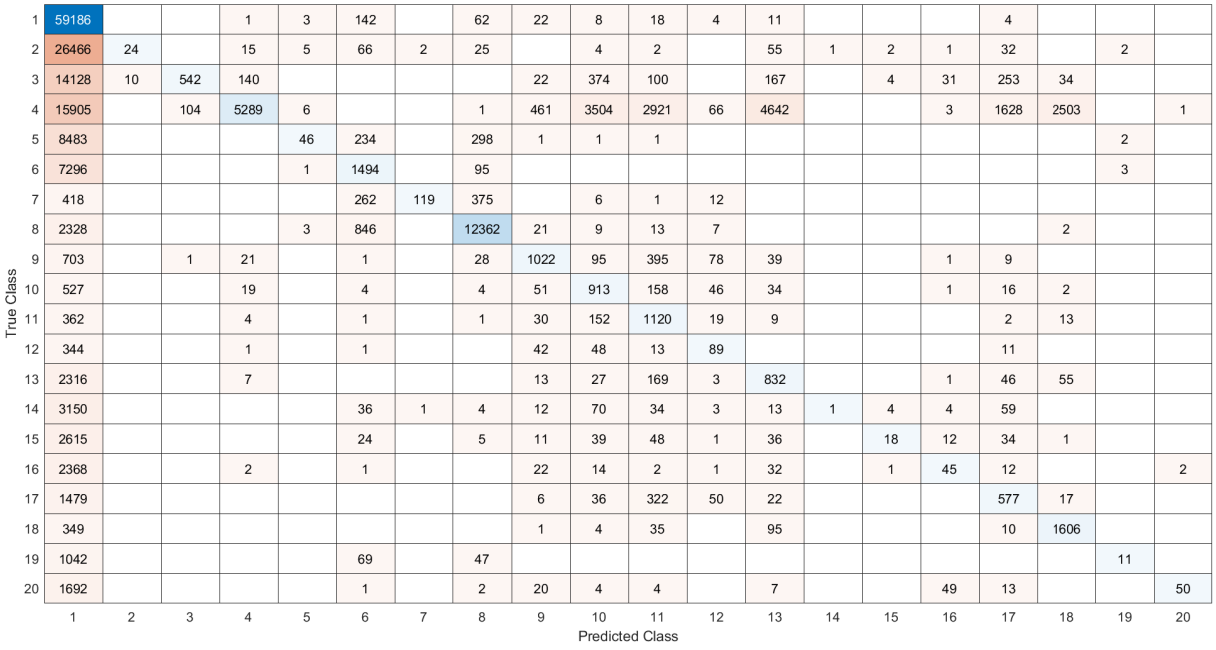


Figure 46: Confusion matrix results for the CNN architecture with PCA ranked features.

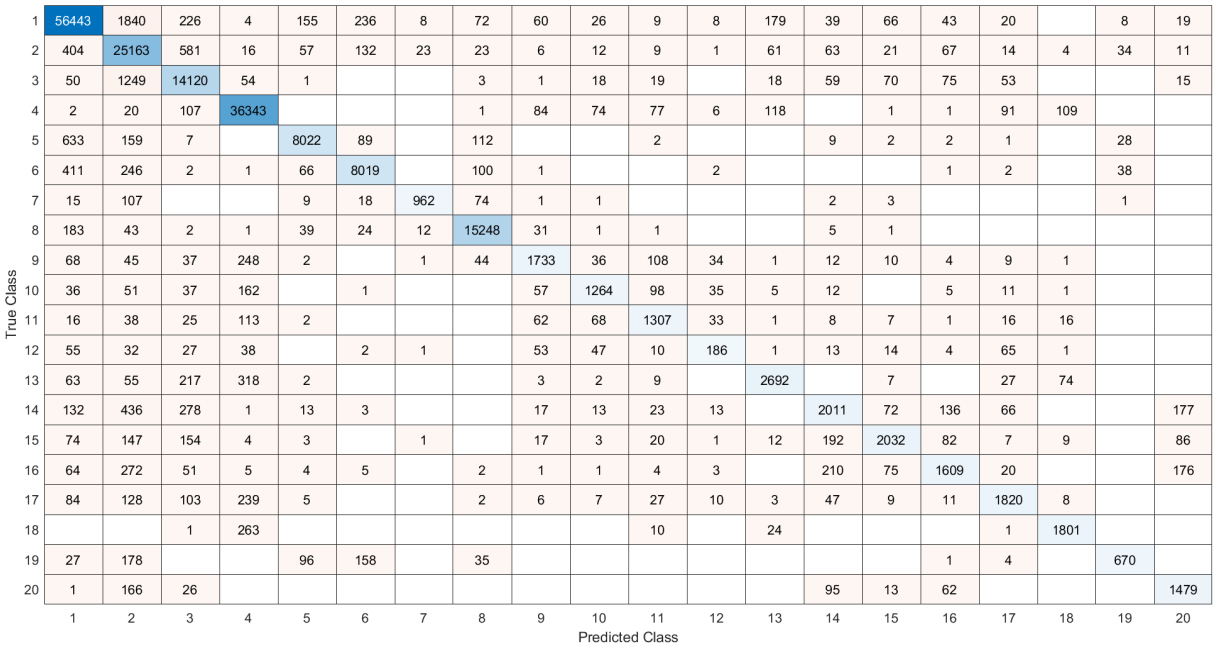


Figure 47: Confusion matrix results for the LSTM architecture with PCA ranked features.

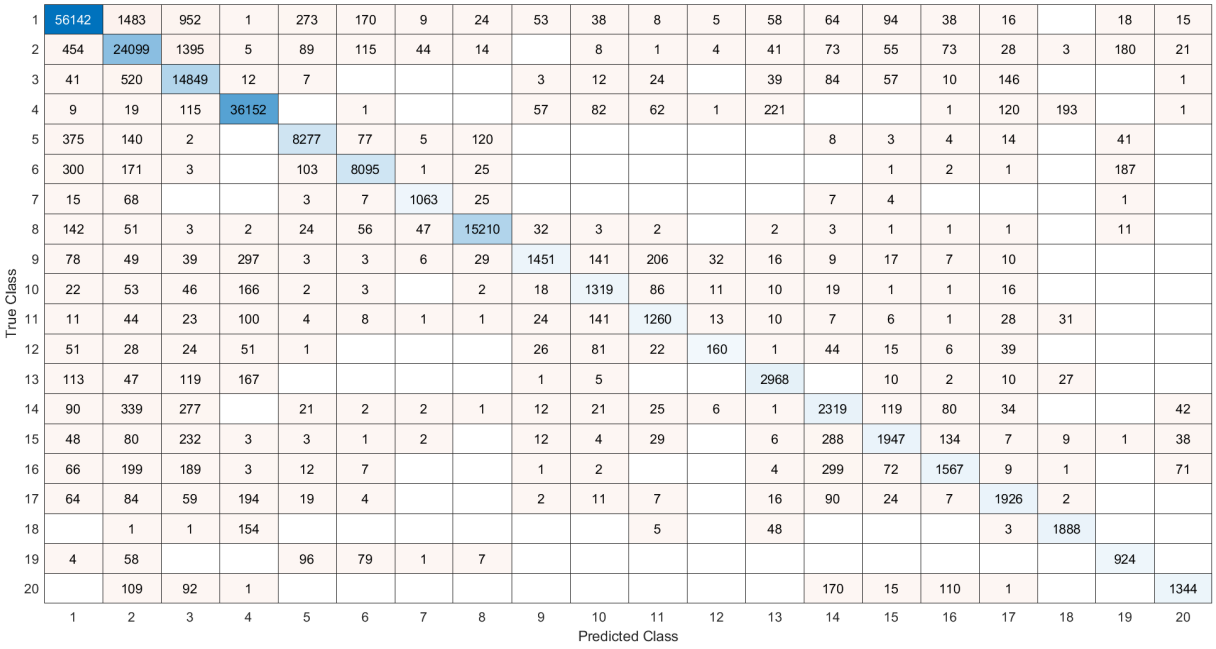


Figure 48: Confusion matrix results for the BiLSTM architecture with PCA ranked features.

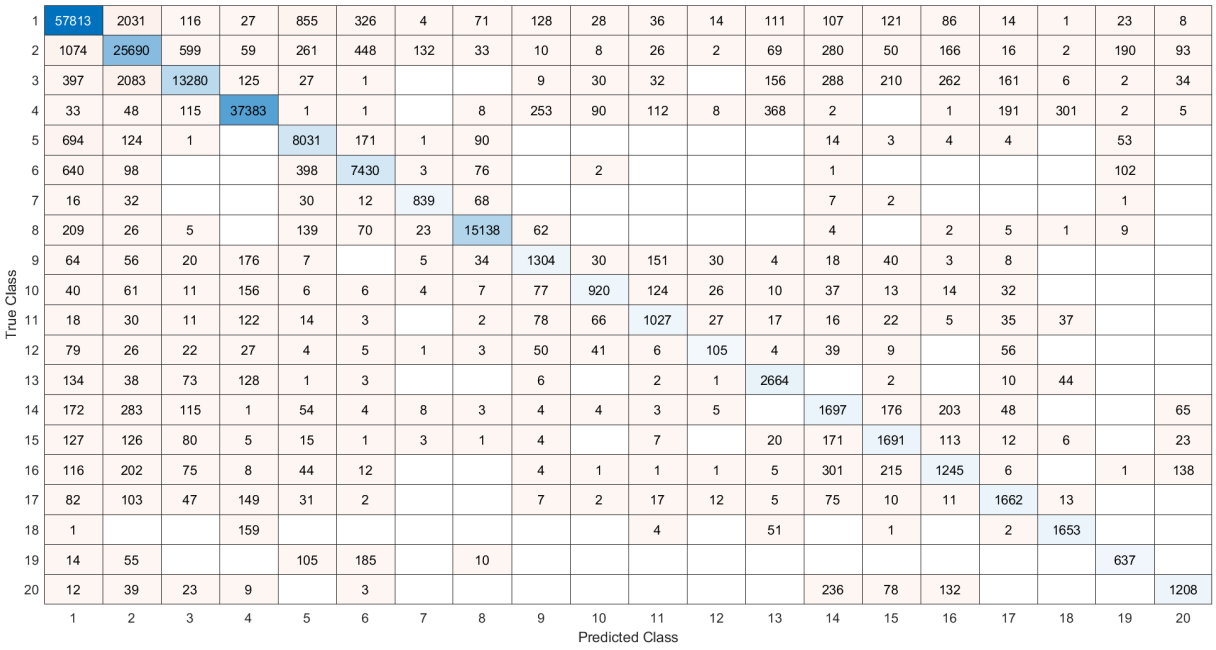


Figure 49: Confusion matrix results for the CNN-LSTM architecture with PCA ranked features.

E.3 Deep Learning Models Computation Time

Table 37: Results for the computation time of the 4 Deep Learning architectures with both feature subsets ranked by Relief-f and PCA

Architecture	Feature Selection Method	Number of Windows	Computation Time (min)
CNN	Relief-F	607600	1209
LSTM			388
CNN-LSTM			1406
BiLSTM			627
CNN	PCA		1277
LSTM			451
CNN-LSTM			1421
BiLSTM			592