Characterization and Analysis of the Cyclostationary Pressure Signals Generated During Walking : Predicting Falls for the Elderly

Reem Abou Marak Dit Brome

N°d'ordre NNT : 2022STET0055

THÈSE de DOCTORAT
DE L'UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE

Membre de l'Université de LYON

École Doctorale Ed SIS nº 488
JEAN MONNET SAINT-ETIENNE UNIVERSITY, LASPI

Spécialité / discipline de doctorat :
Sciences Ingénierie, Santé/ Image, Vision, Signal

Soutenue publiquement le 08/12/2022, par:
ABOU MARAK DIT BROME, Reem

Characterization and Analysis of the Cyclostationary Pressure Signals Generated during Walking: Predicting Falls for the Elderly

Caractérisation et Analyse des Signaux de Pression Cyclostationnaire Générés lors de la Marche : Prédiction des Chutes chez les Personnes Agées

---

Devant le jury composé de :

CARRAULT, Guy	Professeur à l'université Rennes 1	Rapporteur
ZANTOUT, Rached	Professeur à l'université Rafik Hariri	Rapporteur
LE BOUQUIN JEANNES, Régine	Professeur à l'université de Rennes 1	Présidente du jury
SERHAL, Dina	Maîtresse de conférences à l'université Rafik Hariri	Examinatrice
EL BADAOUI, Mohamed	Professeur à l'université UJM, Saint Etienne	Directeur
O. DIAB, Mohamad	Professeur à l'université Rafik Hariri	Co-directeur
BONNARDOT, Frédéric	Maître de conférences à UJM, Saint Etienne	Co-advisor
NASREDDINE, Jad	Senior Researcher à i2CAT Foundation	Co-advisor

Acknowledgments

I would like to acknowledge and give my warmest thanks to all my supervisors whose guidance, advice, and encouragement have carried me through all the stages of finishing my thesis. I would like to express my deepest gratitude to Dr. Jad NASREDDINE for his availability, his invaluable help, and his sound advice in all matters, especially related to machine learning. I am extremely grateful to Dr. Frédéric BONNARDOT for his deep experience in signal processing and his profound guidance and help in python coding. This endeavor would not have been possible without Prof. Mohamad O. DIAB, who has supported and encouraged my goals in academia since the beginning of my journey as a student. Words cannot express my gratitude to Prof. Mohamed EL BADAOUI, director of the LASPI laboratory, for having me accepted and warmly welcomed during my research stays at LASPI. I am also grateful to him for his great interest in my work, his relevant remarks during the various meetings organized within LASPI, and his sound advice.

I also would like to thank my committee members Prof. Guy CARRAULT, Prof. Rached ZANTOUT, Prof. Régine LE BOUQUIN JEANNES, and Dr. Dina SERHAL for allowing my defense to be an enjoyable and memorable moment and for all the insightful comments and suggestions. Thank you all so much.

I also would like to extend my appreciation to everyone from Rafik Hariri Univesity, Lebanon, and Universite Jean Monnet, France, that made this thesis co-direction possible.

I am also very grateful to Prof. Frédéric Roche, Laboratoire de Physiologie de l'Exercice (LPE) and Centre Hospitalo-Universitaire (CHU) of Jean Monnet St-Etienne

University for making the data-set available for our study.

I also would like to thank Mrs. Géraldine PECHARD for her help and guidance from the administrative standpoint and her motivation and support as well. I also would like to thank my colleagues in LASPI, Julian Alberto ESPEJO DIAZ, Gaetan VAUCHE, Dr. Anas HAD, Amjad AL KHATEEB, Ons SAIDI, Bilal EL YOUSFI, and Sharaf Eddine KRAMTI, for their encouragement and support.

I would also like to give special thanks to my husband, Wissam, my son Sam, and my family as a whole for their continuous support and understanding when undertaking my research and writing my report. Your prayer for me was what sustained me this far.

Résumé

Il existe un intérêt croissant pour le développement d'outils de prédiction du risque de chute chez les personnes âgées. Ces outils peuvent être utilisés à des fins préventives.

Les principaux objectifs de cette thèse sont d'étudier les aspects cyclostationnaires des signaux de pression de la semelle intérieure et d'extraire les caractéristiques essentielles indicatives du risque de chutes futures. D'autre part, cette thèse vise à mettre en œuvre et à comparer différentes méthodes d'apprentissage automatique supervisé pour classer les sujets âgés en sujets avec ou sans risque de chute dans le futur.

L'ensemble de données se compose de signaux de pression collectés à partir des semelles intérieures de 519 personnes âgées qui ont fait l'objet d'un questionnaire médical préalable incluant leur propension à la chute. Notre étude propose les caractéristiques indicatives des chutes futures, les modèles d'apprentissage automatique et les méthodes d'optimisation pour développer l'évaluation du risque de chute dans la communauté des personnes âgées.

Enfin, notre étude propose une nouvelle méthode pour représenter les signaux de la semelle intérieure de pression cyclostationnaire et les utiliser dans un modèle d'apprentissage en profondeur pour prévoir les chutes potentielles dans la communauté des personnes âgées.

Abstract

There is an increasing interest in developing elderly fall-risk prediction models that can be used as a preventive approach to prevent future risk of falling in the elderly community.

The primary objectives of this thesis are to study the cyclostationary aspects of the pressure insole signals of older adults and extract essential features indicative of the risk of future falls. In addition, this thesis aims to implement and compare different supervised machine-learning methods to classify elderly subjects as subjects with or without risk of falling in the future.

The data-set consists of pressure signals collected from the innersoles of 519 elderly people who reported whether they had experienced previous falls. Our study proposes the features indicative of future falls, the machine learning models, and optimization methods to develop fall risk assessment in the elderly community.

Finally, our study proposes a novel method for representing cylcostationary pressure insole signals and using them in a deep learning model to predict prospective falls in the elderly community.

Table of Contents

List of Figures

List of Tables

General Introduction

Introduction

French version

L'Organisation mondiale de la santé (OMS) a déclaré (en avril 2021) que les chutes étaient la deuxième cause de décès par blessures non intentionnelles dans le monde [OAU08]. Selon l'OMS une chute est un incident entraînant la chute d'un individu au repos par inadvertance sur le sol ou à d'autres niveaux inférieurs. Elles peuvent parfois avoir une issue fatale. Environ 646 000 personnes dans le monde meurent chaque année de chutes accidentelles et environ 37,3 millions de cas de chute nécessitent une assistance et des soins médicaux chaque année. La plupart des chutes mortelles concernait les personnes âgées de plus de 65 ans [OAU08].

Rubenstein a étudié les causes fréquente de chutes chez des personnes âgées [Rub06]. Ses résultats ont révélé que certains de ces facteurs de risque comprennent, sans toutefois s'y limiter : l'âge (80 ans ou plus), les anomalies musculaires, l'arthrite, la dépression, les chutes antérieures, la prise de plusieurs médicaments, l'utilisation d'un appareil

d'assistance, les troubles de la foulée, du mouvements, de l'équilibre, de la cognition et de la vision. La plupart des chutes accidentelles chez les personnes âgées surviennent en raison d'une combinaison de divers facteurs. Chez les personnes agées, l'importance des chutes est non seulement liée à leur fréquence qui augmente avec l'âge mais aussi à la gravité des blessures (plus élevé chez les sujets les plus âgés ayant des antécédents de multiples chutes antérieures). Cela entraîne un recours plus important aux services médicaux et des frais de ré-éducation. Plus important encore, cela augmente le risque d'invalidités et de décès [Rub06]. Par conséquent, la réduction du risque de chute accidentelle des personnes âgées est vitale d'un point de vue social et économique. Aussi, la mise en œuvre de stratégies de prévention devrait mettre l'accent sur l'éducation, la formation, la création d'environnements plus sûrs, l'établissement de politiques efficaces pour réduire ce risque et encourager la recherche liée aux chutes chez les personnes âgées [Rub06].

Dans de nombreux programmes de prévention une première étapes consiste a évaluer le risque de chute [PNG+01]. Les "Centers for Disease Control and Prevention" (CDC) et l'"American Geriatric Society" conseillent un dépistage annuel des chutes pour tous les adultes de 65 ans et plus [MM17]. L'évaluation du risque de chute est une procédure dans laquelle le risque de chutes futures est estimée [PNG+01] en effectuant une série de tests conçus pour examiner la force, l'équilibre et la démarche des patients. L'évaluation du risque de chute est généralement effectuée dans un environnement clinique et basée sur des questionnaires et des tests fonctionnels de mobilité [PNG+01], tels que le "Timed Up and Go" (TUG) [SCBW00] ou le test "Berg Balance Scale" [LRNP18]. Bien que ces tests donnent une bonne idée de la mobilité optimale et de l'exécution des tâches, leur capacité à prédire le risque chutes est limité [RLJ17] en raison de leur dépendance à l'analyse subjective du personnel effectuant les tests [RLJ17]. De plus, ces tests fonctionnels présentent d'autres inconvénients : ils sont long à réaliser et ne sont pas

standardisés [RLJ17].

Dans ce contexte, nous souhaiterions prévoir le risque de chutes des personnes âgées pour anticiper. Swanenburg et al. [SdBUM10] ce sont interessés aux données issues de plate-formes de force afin de prévoir le risque de chute à l'aide d'une analyse statistique. Ils ont noté que l'amplitude des mouvements médio-latéraux dans des conditions à tâche unique était un indicateur indépendant significatif chez les chuteurs âgés (en plus des antécédents de chutes multiples) [SdBUM10]. Howcroft et al. [HLKM17] a trouvé des différences statistiquement significatives dans les mesures de posturographie statique entre ceux qui ont subi des chutes accidentelles dans le passé et ceux qui ne l'ont pas fait. Dans leur étude, ils ont étudié la posturographie debout les yeux ouverts et les yeux fermés avec des adultes âgés et ont pu identifier les différences et déterminer les scores seuils de mesure pour classer les chuteurs potentiels, les chuteurs simples, les chuteurs multiples et les non-chuteurs. [HLK18] ont également étudié la prédiction des chuteurs âgés à l'aide d'algorithmes d'apprentissage automatique où leur meilleur modèle de classification a atteint une précision de 65% et une sensibilité de 59%, en utilisant la semelle intérieure à détection de pression et l'accéléromètre à tige gauche comme prédicteurs [HLK18]. Les semelles à détection de pression ont également été utilisées dans la littérature pour prédire ou analyser les conditions médicales chroniques et les maladies de la communauté des personnes âgées, telles que la maladie de Parkinson, la démence, et les chutes accidentelles.

Dans les études liées à la locomotion humaine, la marche humaine peut être décrite comme un mouvement qui consiste en une série répliquée et répétée d'actions physiques cycliques ou de foulées. La définition d'un processus cyclostationnaire est un signal qui a des propriétés statistiques variant cycliquement avec le temps. Ainsi, l'analyse des caractéristiques cyclostationnaires des signaux de pression exercé sur la semelle peut introduire de nouvelles caractéristiques significatives pour prévor le risque de chute chez

les personnes âgées. Les propriétés de cyclostationnarité dans les signaux de marche ont également été utilisées dans la modélisation et l'analyse des signaux de marche humaine et de force de réaction au sol (GRF) [SEBG+10, ZTEB+14, TEBS15]. Sabri et al. [SEBG+10] a proposé un cadre alternatif pour l'étude des signaux GRF, basé sur des caractéristiques cyclostationnaires plutôt que sur l'analyse traditionnelle du domaine temporel, qui suppose des composantes de signal statistiquement stationnaires. Le cadre proposé a pu modéliser la périodicité des statistiques de signal et a montré des résultats améliorés en démontrant le développement de la détection de la fatigue des coureurs. Zakaria et al. examiné et exploité les propriétés et caractéristiques cyclostationnaires (CS) telles que la fonction d'autocorrélation cyclique. Leurs travaux ont démontré qu'il existe une différence significative dans l'autocorrélation cyclique des chuteurs et des non-chuteurs [ZTEB+14]. Un indicateur de cyclostationnarité est le degré de cyclostationnarité (DS) [ŽG91] [ZEBM+13] qui s'est avéré être un bon indicateur pour la prévision des chutes futures [BNB+21]. De plus, dans cette thèse, nous avons montré que l'utilisation d'images de cartes thermiques de la corrélation spectrale comme entrées dans des réseaux de neurones convolutifs d'apprentissage en profondeur montre des résultats prometteurs à proposer comme nouvelle méthode pour évaluer le risque de chutes futures dans la communauté des personnes âgées.

English version

The World Health Organization's (WHO) facts sheet (updated in April 2021) declared falls as the second leading cause of unintentional injury deaths worldwide [OAU08]. The WHO defines falling as an incident resulting in an individual falling at rest inadvertently on the ground or at other lower levels. Unfortunately, fatality can occur due to injuries caused by falls. In summary, an estimated number of 646 thousand people worldwide die annually from accidental falls and round 37.3 million fall cases demand medical assistance and attention yearly. The most significant number

of fatal falls involved older adults over 65 years old [OAU08].

Rubenstein studied the common causes for falls in the elderly community [Rub06]. His results revealed that some of these risk factors include, but are not limited to, being 80 years old or above, muscle defects, arthritis, depression, prior falls, the use of multiple medications, the use of an assistive device, impairments in stride movements, balance, cognition, and vision. Most inadvertent elderly falls occur due to a combination of various factors. He also mentioned that the significance of falls within the elderly community is not just limited to the point that the frequency of number of falls increases as age increases, but also that the severity of the injury is highest among the oldest subjects with a history of multiple prior falls. This leads to an increase in medical services and rehabilitation expenses. More importantly, it increases the possibility of disabilities and fatalities [Rub06]. Therefore, reducing the risk of accidental elderly falls is vital from a social and economic point of view. Hence, implementing prevention strategies should emphasize education, training, building safer environments, establishing effective policies to lower susceptibility, and encouraging elderly fall-related research [Rub06].

In many intervention programs suggested to deter future falls, fall risk assessment is conducted as the first step to determine individuals at risk of falling[PNG+01]. In addition, the Centers for Disease Control and Prevention (CDC) and the American Geriatric Society advise yearly fall assessment screening for all adults 65 years and older [MM17]. Fall risk assessment is a procedure where the possibility of future falls is estimated [PNG+01] by performing a series of tests designed to examine patients' strength, balance, and gait. The assessment of fall risk is typically done in a clinical environment and based on questionnaires and functional tests of mobility [PNG+01], such as the Timed Up and Go (TUG) [SCBW00] or the Berg Balance Scale test [LRNP18]. Although these tests have been found to give a good sense of optimal mobility and task execution, their capability to predict the occurrence of future falls

is limited [RLJ17] due to their reliance on the subjective analysis of the personnel conducting the tests [RLJ17]. In addition, these functional tests have other drawbacks, such as requiring much time and the lack of a standardized approach with concise and straightforward instruction for both patients and providers [RLJ17].

In this context, there is growing interest in the prediction of future elderly falls to help reduce its risk. Swanenburg et al. [SdBUM10] studied whether force plate variables can be employed to predict multiple fallers using statistical analysis. They noted that the amplitude of medial-lateral movements in single-task conditions was a significant independent predictor of elderly fallers, along with having a history of multiple falls [SdBUM10]. Howcroft et al. [HLKM17] found statistically significant differences in the static posturography measures between those that experienced accidental falls in the past and those who had not. In their study, they investigated eyes open, and eyes closed standing posturography with elderly adults and were able to identify differences and determine the measure cut-off scores for classifying prospective fallers, single-fallers, multi-fallers, and non-fallers [HLKM17]. Howcroft et al.[HLK18] also studied the prediction of elderly fallers using machine-learning algorithms where their best classification model achieved 65% accuracy and 59% sensitivity, using pressure-sensing-insole and left-shank-accelerometer as predictors [HLK18]. Pressure sensing insoles have been also used throughout the literature to predict or analyze chronic medical conditions and diseases in the elderly community, such as Parkinson's disease, dementia, and accidental falls.

In studies related to human locomotion, the human walk can be described as a movement that consists of repeated replicated series of cyclic physical actions or strides. The definition of a cyclostationary process is a signal that has statistical properties varying cyclically with time. Thus, analyzing the cyclostationary characteristics of the foot plantar pressure signals can introduce new features that signify the risk of

falling in the elderly. Properties of cyclostationarity in gait signals have been also used in modeling and analyzing human walk and ground reaction force (GRF) signals [SEBG+10, ZTEB+14, TEBS15]. Sabri et al. [SEBG+10] proposed an alternative framework for the study of GRF signals, based on cyclostationary characteristics rather than the traditional time domain analysis, which assume statistically stationary signal components. The proposed framework was able to model the periodicity of the signal statistics and showed improved results in demonstrating the development of runners' fatigue detection. Zakaria et al. examined and exploited the Cyclostationary (CS) properties and features such as the cyclic autocorrelation function. Their work demonstrated that there is a significant difference in the cyclic autocorrelation of fallers and non-fallers [ZTEB+14]. One indicator of cyclostationarity is the degree of cyclostationarity (DS) [ŽG91] [ZEBM+13] that has shown to be a good indicator for prediction of future falls [BNB+21]. In addition, in this thesis we showed that utilizing heat map images of the spectral correlation as inputs to deep learning convolutional neural networks shows promising results to be proposed as a novel method to assess the risk of future falls in the elderly community.

Objectives

The main objective of the work carried out in this thesis is to propose a method for assessing the risk of future falls in older adults using the cyclostationary aspect of pressure insole signals and machine learning.

Cyclostationary analysis is conducted on the pressure insole signals of participants after proper noise filtering, normalization, and speed fluctuation compensation. Then the degree of cyclostationarity is extracted as a feature to be included among other time and frequency domain features used in the literature.

Feature selection methods such as Relief-F and Sequential backward selection (SBS) are explored and compared. In addition, hyperparameter optimization methods are used to set the optimal hyperparameters for each classification model.

Finally, the novel approach of using heat map images of the average spectral correlation of the pressure insole signals is explored to be used as inputs to a deep learning convolutional neural network (CNN) for predicting future falls in the elderly.

Motivations and Contributions of this Doctoral Thesis

This doctoral thesis proposes methods for predicting future falls in the elderly community using cyclostationary components of pressure insole signals. Nevertheless, we cannot achieve this purpose without a good understanding of the existing literature, the data-set provided, and the physical phenomena that generate these signals. Therefore, the first task to be accomplished is exploring the previous similar work, performing analytical modeling, and describing the cyclostationary aspect of the signals. Then, after a deep understanding of the mentioned two issues, we can proceed with the task of developing prediction methods and experimenting with the data-set provided.

The primary interest behind the methods developed is having an automated assessment method that provides an early diagnosis of elderly patients, whether they are at risk of falling or not. The existing fall risk assessment tools (such as the Berg Balance Scale) require a long time to perform, and several qualified staff to overlook the process. The results are sometimes considered subjective depending on the team conducting the tests. The use of machine learning will support specialists in their decision-making. The existing literature on utilizing machine learning to predict future falls has not achieved high enough accuracy to be used in the medical field. Our study shows that including the cyclostationary aspect of walking signals can help improve system performance

(around 94% accuarcy, 92.5% sensitivity, 96.3% specificity and 96% precision) as it provides essential features for predicting future falls in the elderly.

Organization of the Manuscript

This thesis manuscript is organized as follows:

— Chapter 1 presents the existing literature on the prediction of elderly fallers using machine learning with the comparison of the different techniques and their performances. The work in this chapter will be submitted to IEEE Transactions on Biomedical Engineering.

— Chapter 2 introduces cyclostationary signals, their analysis, and how the are related to gait signals.

— Chapter 3 describes the data, and the experiment conducted to gather the pressure insole signals of the 519 participants above the age of 65.

— Chapter 4 is dedicated to describing data preparation, which includes signal pre-processing, extraction of time and frequency domain features, cyclostationary analysis and related features, heat map representation of the spectral correlation, and feature selection methods.

— Chapter 5 clarifies in detail the different classification methods used throughout this thesis, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Decision Trees, and Convolutional Neural Networks (CNN).

— Chapter 6 addresses the optimization of the classification models used in chapter 5 using Grid Search and ADAM Optimization Algorithm for Deep Learning.

— Chapter 7 showcases the results of the models built, their analysis, and discussion.

— Finally, the conclusions section lists our general findings, conclusions, future

perspectives, and recommendations.

Contributions

The work carried out as part of this thesis has been highlighted by publications in international journals and conferences:

- R. Brome, J. Nasreddine, F. Bonnardot, M. Diab, M. El Badaoui.
  Prediction of Elderly Falls Using the Degree of Cyclostationarity of Walk Pressure Signals.
  IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 1-3 March 2021.
- R. Brome, J. Nasreddine, F. Bonnardot, M. Diab, M. El Badaoui.
  Identifying Elderly Patients At Risk of Falling using Time-Domain and Cyclostationarity Related Features.
  International Journal of Integrated Engineering, 13(5), 57–66.,20 May 2021
- R. Brome, F. Bonnardot, J. Nasreddine, M. Diab, M. El Badaoui.
  Estimation du Risque de Chute chez les Personnes Âgées.
  Congres National de la Recherche des IUT (CNRIUT), 9-10 June 2022.
- R. Brome, F. Bonnardot, J. Nasreddine, M. El Badaoui, M. Diab
  Elderly Fall Risk Prediction Using Cyclostationay and Time-Domain Features.
  GRETSI (Groupe d'Etudes du Traitement du Signal et des Images), 2022, 6-9 September 2022.
- R. Brome, J. Nasreddine, F. Bonnardot, M. El Badaoui, M. Diab.
  Fall Risk Assessment Using Pressure Insole Sensors and Convolutional Neural Networks.
  IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 7-9 December 2022.

Chapter 1

Literature Review

Ce chapitre résume les études existantes sur la prédiction et l'évaluation des chutes des personnes âgées utilisant principalement des techniques d'apprentissage automatique. De plus, nous analysons les différentes méthodes de collecte de données, les différentes approches statistiques et d'apprentissage automatique utilisées pour prédire les chutes, et les caractéristiques qui contribuent au risque de chute.

This chapter summarizes existing studies on the prediction of elderly falls mainly using machine learning techniques. Furthermore, we analyze the different methods for collecting data-sets, the various statistical and machine learning approaches used to predict elderly falls, and the features that contribute to the risk of elderly falls.

1.1 Elderly Falls and Their Prediction

The World Health Organization's (WHO) facts sheet (updated April 2021) identifies falls as the second leading cause of unintentional injury deaths worldwide [OAU08]. WHO defines falling as an event that results in an individual becoming at rest involuntarily on the ground or lower levels. Unfortunately, injuries caused by falls can lead to fatality. In 2008, approximately 680 thousand people worldwide die each year from accidental falls. In addition, around 37 million fall cases demand medical assistance and attention yearly. The most significant number of fatal falls involved adults above 60 years old [OAU08].

Early studies of human anatomy and physiology demonstrate that the *aging process* in humans, with the recession of nervous system function and physiological function, reduces older people's walking ability. Therefore an aged person may easily be injured during walking, leading to other physical impairments and senile diseases [ASS18]. Therefore, fall risk evaluation and prediction are indispensable while the elderly population surges. It is also asserted that effective physiotherapy and fall prevention programs are cost-effective and appropriate to maximize the quality of life and support the independence of older adults [GRG+12]. The foremost action to take for an effective fall prevention program is identifying those at risk of falling and determining the most appropriate interventions to reduce or eliminate falls [ASS18]. Unfortunately, there are currently no convenient solutions to identify the risk of falling in individuals. Doctors mostly rely on *visual examination and experience* to inspect patients' gait stability and symmetry in a clinical setting. This assessment is time-consuming and lacks neutrality and effectiveness in detecting gait abnormality and the future risk of falling. Furthermore, wrong diagnosis in these cases can lead to not taking the appropriate preventive measures for accident falls, which can often lead to serious physical injury or

even death. It is therefore imperative to, as early as possible, try to identify unstable gait to apply the necessary preventive measures.

There is growing interest in predicting future elderly falls to help reduce their risk of occurrence. More than *400 risk factors* for elderly falls have been reported in [DLB+10]. Most risk factors have been evaluated in laboratory settings or clinical test situations, and fall risk assessment systems have been designed to serve as screening tools for fall risk [OBS+97, RHPC00, TPS+01, NPC+04]. Considerable studies have demonstrated a relationship between falls and risk factors such as being a female, old age, diminished cognitive functions, muscle weakness, and gait abnormalities. In particular, gait abnormalities in the elderly population, including slow walking speed, more significant gait variability, and shorter steps, are considered the most significant risk factors for elderly falls.

Currently, there is no single *measurement device* that can evaluate all elements of fall risk factors. However, it may be possible that a single tool can optimally identify people at high risk for falling [ASS18]. The Berg Balance Scale (BBS) [LRNP18] is a valid and reliable scale for assessing the balance state of elderly people. However, there are a few inconveniences involved in relying on the BBS: it requires 14 balance tasks, takes a long period of assessment (12 months), and demands much work following up with each patient and collecting data for each balance task. The British STRATIFY tool [OPG+08] was developed to predict falls in a hospital setting. Although the tool exhibits several strengths, specific limitations exist which may not be acclimatized to other countries such as reported limited accuracy for predicting falls in hospital and rehabilitation settings [Kas09, PR91]. Considerable clinical balance and mobility performance tests, such as the Timed Up and Go (TUG) [SDH+13], sit-to-stand, and alternate step tests, have been proposed as fall risk screening tools [Boh06, TLS10]. Also, in Australia, four fall risk assessment tools have been studied, and the conclusion

is that the tools have poor clinical properties for identifying fall risk factors for those most at risk of falling [MHB+12].

More recently, *statistical analysis and machine learning techniques* have gained attention for recognizing essential features for predicting elderly falls and building classification models. Proper experimentation design (criteria for participants, involvement of secondary tasks or not), data collection (labeling of data using retrospective or prospective approach), and feature extraction (clinical data or data collected from wearable or non-wearable sensors) must be done to build reliable classification models. The machine learning techniques use features extracted from clinically relevant data, allowing computer algorithms to form a predictive model. In addition, the machine learning algorithms can be used to extract the optimal features affecting the risk of falls from the gait features.

In this chapter, we present our review of recent research papers investigating the various techniques in the process of classifying elderly fallers and non-fallers. This process includes data collection experimentation, feature extraction, feature selection, and classification algorithms.

1.2 Methods for Search Strategy and Criteria

This review was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement (PRISMA) [MSC+15]. In addition, a systematic electronic search was conducted in IEEE, Scopus, Web of Science, and PubMed databases in April 2022. The search algorithm was designed to include all possible combinations of keywords from the following: fall risk, fall prediction, fall assessment, gait, old, aged, and elderly. Finally, reference lists from the identified publications were reviewed to

identify additional research articles. The criteria set for the studies to be included in this review were (i) peer-reviewed publication in the English language, (ii) published after 2010, (iii) studying fall risk prediction, (iv) age of participants is above 60 years old, and (v) the fall risk prediction involves a structured, supervised experiment and not free daily-living activity tracking. In addition, studies were excluded from the review if they met any of the following criteria: (i) studies that focused on momentarily fall detection, (ii) studies that investigated fall risk specifically for neurologically impaired patients, and (iii) studies that did not include proper performance metrics to report. The results of this search are illustrated in Figure 1.1.

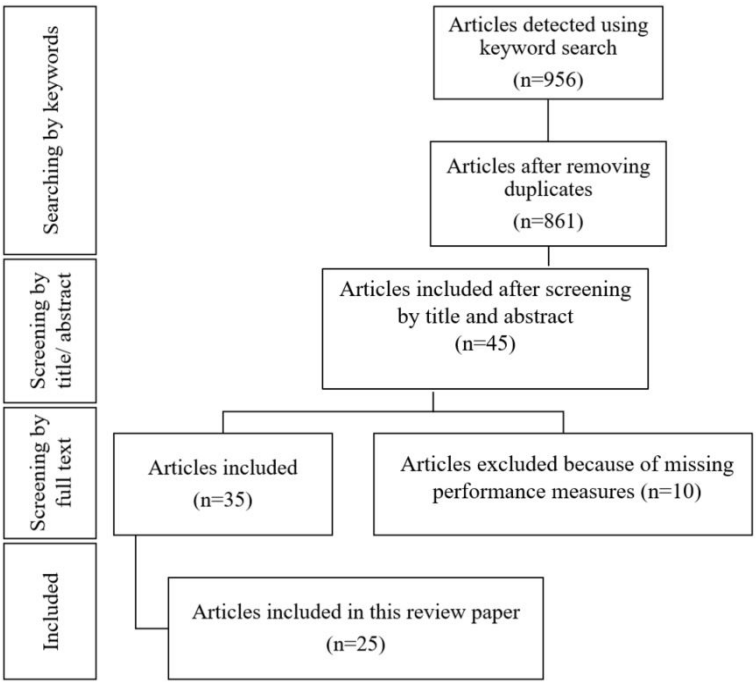

Figure 1.1 – Articles Selection Flow Chart

1.3 Datasets used for Prediction of Fallers

Existing studies in fall prediction are different in many aspects, such as the choice of the dataset, whether using a pre-existing dataset or starting with designing an

experiment, recruiting participants, and collecting data from scratch. They also vary in the population size and the ratio of fallers (F) and non-fallers (NF). In addition, the studies were different in terms of data used, whether it is purely clinical data or measurement-driven data. Clinical data involve the patients' medical records, medication consumption, gender, diseases, and functional test scores. In contrast, measurement-driven data are signals or data extracted from sensors used during the experiment. The experiments conducted to investigate falling in the elderly community involved different approaches, such as having the participants perform a single walking task or perform dual tasks (secondary tasks) while walking. The following sections explain further the different approaches in the experimentation strategies in the concerned studies.

1.3.1 Human Gait Analysis

Human gait is a manner of walking and a better-suited medical term to describe human locomotion [Wil14]. Walking gait is defined as repetitive cyclic gestures consisting of periodic movements of each foot from one position to another. In addition, human balance while walking is achieved by sufficient ground reaction forces exerted by the feet [Vau09].

This bipedal locomotion (means of movement by two limbs or legs) is enhanced by body parts such as the bones, the muscles, and the nervous system. Consequently, each limb contributes to braking and propulsive forces to maintain balance, the forward velocity, and the body's vertical support forces [SRA+]. Therefore, any defect in one of those parts may lead to a pathological gait and cause higher risks of future falls.

In addition, gait can be affected by many other variables, such as old age, past injuries, medications, lack of physical activities, etc.

The normal human gait cycle can be seen in figure 1.2 with two main phases: Stance

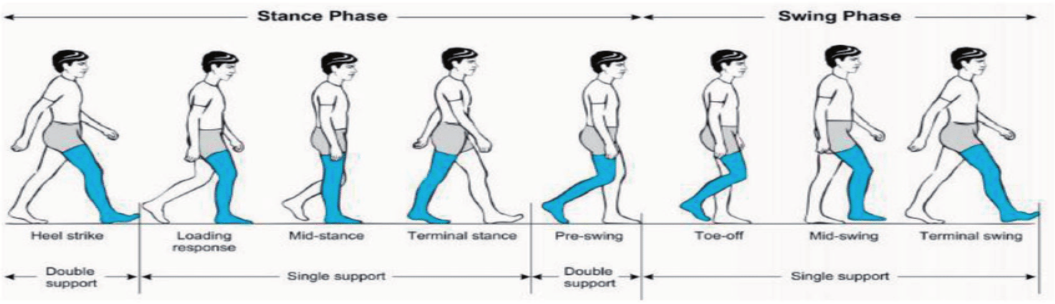

Figure 1.2 – Human Gait Cycle[NMMR18]

and Swing. The human gait cycle comprises two main phases: the stance phase and the swing phase. The stance phase, which occurs around 60% of the average human walking cycle, occurs when the foot is in contact with the ground, whereas the swing phase refers to the period when the foot is in the air. *Measuring and analyzing human gait can be done using clinical assessment tools or wearable and non-wearable sensors.[NMMR18]*

1.3.2 Clinical Fall Risk Assessments Tools

A few fall risk assessment tests are currently used in a clinical setting to determine if a person has a risk of falling in the future. During these assessments, the medical staff examines patients' strength, balance, and gait. Some of these fall assessment tools are explained briefly below.

The *Timed Up-and-Go* (TUG) is a test that checks the patient's *gait*. The patient starts the test sitting in a chair, then is asked to stand up and walk for about 10 feet at a regular pace. Then they sit down again. The duration of these tasks is of interest. If all these tasks take 12 seconds or longer, this indicates a risk of falling.[SDH+13]

The *30-Second Chair Stand Test* examines patients' *strength and balance*. This test starts with the subject sitting on a chair with arms crossed over the chest. They are later instructed to stand up and sit again. This task is repeated for 30 seconds. The

person conducting the test counts how many times the patient can perform the test. Lower numbers indicate a higher risk for future falls. The threshold that indicates a risk depends on the age of the patient [JSM+15].

The *4-Stage Balance Test* (4-stage) examines how well patients can *keep their balance*. The patient is asked to stand in four positions, holding each one for 10 seconds. The positions will get more challenging as the test resumes.

— Position 1: Standing with feet side-by-side.
— Position 2: Moving one foot halfway forward, so the instep touches the other foot's big toe.
— Position 3: Moving one foot in front of the other, so the toes touch the other foot's heel.
— Position 4: Standing on one foot.

If the subject cannot hold position 2 or position 3 for 10 seconds or they cannot stand on one leg for 5 seconds, the patient has a future risk of falling.[SKS+21]

1.3.3 Wearable and Non-Wearable Sensors

Various sensors were used throughout the literature for fall risk assessment or prediction. These sensors provide an opportunity to capture the fall-related biomechanics of elderly people. Sensors used in the studies reviewed were either wearable sensors, such as Inertial Measurement Units (IMUs) (as shown in figure 1.3: accelerometers (accel.) [RTP+13, HLK16], gyroscopes (as shown in Figure 1.4 [DWF+13], and pressure insoles [BND+21], or non-wearable, such as Microsoft Kinect [CRS+14, KMS+14], Wii balance boards (as shown in Figure 1.5 [KCP15, YAN+11], and pressure platforms [SMT+17].

Figure 1.3 – Example of IMUs Used for Gait Analysis

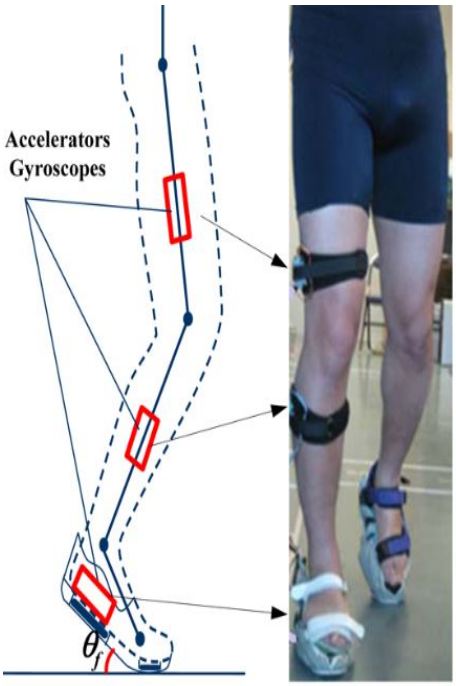

Figure 1.4 – Example of Accelerometers Gyroscopes Used for Gait Analysis

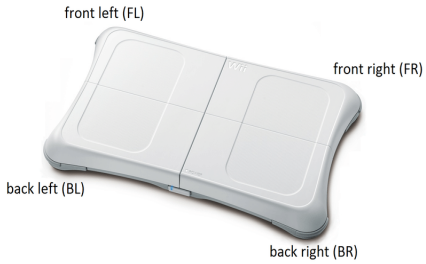

Figure 1.5 – The Wii Balance Board Used for Gait Analysis

To this end, several studies have demonstrated the ability of such sensors to capture biomechanical measures associated with fall risk in the elderly community, such as walking speedand stride time variability, and demonstrate correlations with clinical assessments.

While considerable devices that measure gait data exist, triaxial accelerometers have several desirable characteristics for screening purposes. It has been shown that raw data can provide accurate gait measurements when motion sensors sample at 30 Hz or higher.

1.3.4 Retrospective and Prospective Faller/Non-Faller Study

Most studies in this field have adopted the retrospective (retrospec.) approach in faller/ non-faller data [HKL13, SRNL11, GRC16, HLK16], in which each subject self-reported his/her history of falling prior to the fall risk prediction. Thus, the data was labeled as faller and non-faller and underwent supervised machine learning to classify between fallers and non-fallers. In contrast, only few works have adopted a prospective (prospec.) approach. The prospective strategy is when the experiment for collecting measurement first takes place. After a certain period, the subjects were followed up to report whether they experienced falling during this period. Once that information is collected, the data would be labeled and ready to input into a supervised

machine learning algorithm [SdBUM10, GFG+14, MRW+11, GDW+12].

In [GMC14], Greene et al. compared the retrospective and prospective fall data approaches using the same data-set. Interestingly, their results suggest that the retrospective procedure in falls risk assessment outperforms the prospective one in identifying fallers and non-fallers.

1.3.5 Experiments Involving Secondary Tasks

There has been a proliferation of research reporting various sensor-based methods for assessing fall risk in older adults in recent years. These are usually based on measured quantitative data obtained during an experiment of prescribed tasks, such as regular walking, walking while de-counting from 50, or a TUG (timed-up-and-go) test.

[CTPC13] presented a review to examine the influence of the type and complexity of a secondary task on falls prediction in the elderly population. They divided the secondary tasks into five groups: reaction time tasks, discrimination and decision-making tasks, mental tracking tasks, verbal fluency tasks, and manual tasks.

— *Reaction time tasks* refer to tasks that involve measuring the *elapsed time* between a sensory stimulus and a behavioral response (such as time spent to stand up and sit down). These tasks have typically measured processing speed when slowed processing might underlie a general attentional deficit.

— *Discrimination and decision-making tasks* refer to tasks that require selective *attention and response* to a specific stimulus or feature (such as answering mathematical questions). These tasks have typically been used to examine attention and response inhibition.

— *Mental tracking tasks* refer to tasks that require *holding information in the mind*

while carrying out a mental process, such as de-counting from 50. These tasks have typically been used to examine sustained attention, information procession speed, and working memory.

— *Verbal fluency tasks* refer to tasks that require spontaneous *word production under prespecified search conditions*, such as calling out loud names of animals that they can remember.

— *Manual tasks* refer to *balancing tasks of one or both arms*, such as cup- or tray-taking tasks. Results of this meta-analysis[CTPC13] showed that the mental tracking task yielded significant dual-task-related changes for fall prediction.

Most studies successively used an appropriate level of task complexity specific to the specified population of interest and the objectives of the study.

1.4 Features Involved in Fall Risk Prediction

Raw signal from datasets can not be exploited directly and some processing should be done to create or extract features. In the literature on predicting future falls in the elderly community, two types of features are mainly used:

1. clinical features from patients' medical profiles and clinical tools test scores, and,
2. features extracted from wearable and non-wearable sensors.

1.4.1 Clinical Fall Risk Factors

The clinical fall risk factors for elderly include, but are not limited to, the following: prior falls, balance impairment, decreased muscle strength, visual impairment, use of multiple medications, use of multiple psychoactive drugs, impairment or difficulty in walking, depression, dizziness or orthostasis, functional limitations, old age (above 80

years old), female gender, cognitive impairment, arthritis, diabetes, and physical pain [AB21].

The risk of falling escalates with the increase in the number of risk factors found in an older person. A recent meta-analysis recognized the following risk factors as having the strongest association with falling: previous falls, Parkinson's disease, the need to use walking aids, gait problems, and antiepileptic drugs.

[DLB+10] showed that the risk of falling increases with the severity of pain caused by chronic musculoskeletal, the number of damaged joint groups, and the extent of their interference with patients' lifestyles. In addition to data related to the medical profile of patients, results of clinical assessment tools are also considered clinical features. 17 of the 25 reviewed studies included in their features evaluated the test scores of clinical tests, such as TUG, Tinetti, Palliative Performance Scale (PPS), STS5 (Five Times Sit to Stand Test), Functional Reach, Berg Balance Scale (BBS), Five Chair Stands, and Romberg test scores.

Table 1.1 outlines the clinical and demographic features included in the reviewed studies.

Clinical and Demographic Features	Number of Studies out of the 25
Age	8
Gender	6
Height	4
Weight	3
Body Mass Index (BMI)	5
Previous Falls	4
Muscle Performance	1
Visual Function	2
Physical State	4
Cognitive Function	4
Cardiovasclar Health	2
number of medicines	2
number of health conditions	1
fear of falling	1
Education Level	1
TUG Scores	14
Tinetti test scores	3
PPS test scores	1
STS5 test scores	3
Functional reach test	1
Berg Balance Scale BBS	2
FCS Five Chairs Stands	1
Romberg test scores	1

Table 1.1 – Clinical and Demographic Features Used in the Studies Reviewed

1.4.2 Features Extracted from Sensor-Measured Data

The use of features extracted from sensor data permits the generalization of the assessment of elderly fall risks and the overcoming of the drawbacks of known threshold-based techniques in which several parameters need to be manually estimated according to the specific features of the end-user. Another advantage of the use of sensor data is limiting the workload required for the process [RLS13].

Tables 1.2, 1.3, 1.4, and 1.5 summarize the attributes extracted from the four types of sensors used in the papers reviewed.

Accelerometer features were the most popular as they have been used for gait analysis in older studies than other sensors (16 of the 25 studies). In addition, 5 studies out of the 25 included pressure insole extracted features. On the other hand, Kinect and laser infrared sensors are relatively new sensors used for elderly faller assessment. They offer ease of use as they do not require wearable sensors.

1.4.3 Cyclostationary Properties in Pressure Insole Signals

In the research field of human locomotion, the human walk can be thought of as a movement that consists of replicated sequences of cyclic physical movements or strides[ZTEB+14].

Figure 1.6 shows various measurements associated with the walk (pressure signal) and the run (force and acceleration).

The *force signal* comes from an instrumented treadmill that measures Ground Reaction Force during walking. The *pressure signal* comes from an instrumented sole that measures the pressure of the left heel. The *acceleration signal* comes from a MEMS

Sensors	Features	Number of Studies
Accelerometers	Gait Speed	14
	Gait Symmetry	3
	Step Variability	3
	Stride Variability	3
	Step Root Mean Square (RMS) Stride (RMS)	1
	Step count	6
	Stride Length	4
	Step Length	2
	Sway Length	1
	Stride time	6
	Stance time	3
	Stance phase	2
	Percentage Stance time	1
	Swing time	4
	Sway Velocity	2
	Sway Area	1
	Maximum and minimum Sway	1
	Sway Frequency	1
	Foot Angle	1
	Frequency of steps	3
	Frequency of steps	1
	Variation of Step Period	1
	Max, min, mean, median, std, and RMS acceleration	16
	Max, min, mean, median, std, and RMS angular velocity	15
	AP ratio of even to odd harmonics	1
	Fundamental harmonic of FFT	2
	Fundamental amplitude of FFT	2

Table 1.2 – Features Extracted from Accelerometers Used in the Studies Reviewed

Sensors	Features	Number of Studies
Pressure Insoles and Platforms	Energy	5
	Entropy	5
	Skewness	5
	Kurtosis	5
	Mean and std of the Stride Time	1
	Max, mini, mean, median, std, and RMS pressure signal	5
	CoP position coordinates	2
	mean and std CoP positions	2
	displacement of CoP	1
	Mean and std velocity of CoP displacement	1
	ML CoP path stance phase CoV	1
	Pulse Width	1
	Undershoot	1
	Duty Cycle	1
	Slew Rate	1
	Range	1
	Pressure Difference between different pressure points	2
	Average Degree of Cyclostationarity	1

Table 1.3 – Features Extracted from Pressure Sensors Used in the Studies Reviewed

Sensors	Features	Number of Studies
Kinect	Center of Mass (CoM)	2
	Human movement tracking	2
	ML Standard Deviation	2
	AP Standard Deviation	2

Table 1.4 – Features Extracted from Kinect Used in the Studies Reviewed

Sensors	Features	Number of Studies
Infrared Laser Device	Reaction time	1
	Stepping time	1
	Step length	1
	Percentage of correct steps	1

Table 1.5 – Features Extracted from Sensors Used in the Studies Reviewed

accelerometric sensor placed on the tibia of the subject. Scales are unit less.

The left column shows amplitude variation against the time, and the right column shows the Power Spectral Density (i.e., the power associated with each frequency). The stride period is indicated by alternative black and white patterns on the left. Vertical lines in the right figures materialize the stride rhythm and its harmonics. Since the sensor is fixed on one leg, there is only one shape per stride for the pressure and acceleration sensor. With the treadmill, both left and right foot shapes appear in a stride.

A slight difference in the shape enables us to distinguish the two feet. On each signal, a periodic pattern can be identified. This pattern is associated with a cyclic process of the human walk. If the signals are observed during a sufficiently short time to have a constant walking speed, the process associated with the signal will be cyclostationary.

Considering the process as stationary (i.e., not taking care of the periodicity) is equivalent to ignoring all the fluctuations in the cycle. Considering the process as a non-stationary (more general approach) enables us to take care of the signal variations but without considering that these variations are periodic.

The cyclostationary approach consists of taking advantage of the cycles present in the signal to analyze it. Cyclostationary results of a coupling of a periodic process (the walk) and a random process that introduces some variation (physical state, disease, additional task during the walk, type of ground).

1.5 Data Analysis and Prediction of Falls

The prediction methods mainly used for the prognosis of future falls in the elderly are statistical analysis or machine learning strategies. Statistical approaches are used

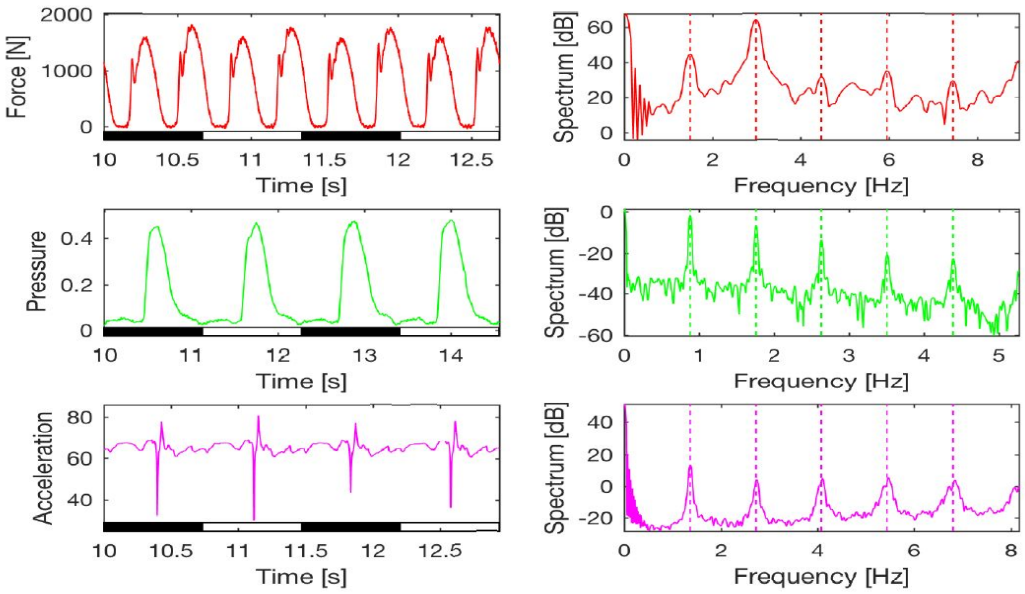

Figure 1.6 – Walking Signals in Time and Frequency Domain [BEB10]

in older research studies in this field. However, more recent studies rely on machine learning and are moving towards more improvement and robust techniques with the advancement in machine learning.

1.5.1 Statistical Analysis

In the literature on assessing the risk of falling in the elderly, several studies use statistical analysis to identify essential input features to the classification models. In contrast, some used statistical analysis as a classification method to identify elderly fallers and non-fallers by finding the proper threshold.

In [SdBUM10], they investigated the ability of force platform variables in single and dual tasks (with eyes closed) to predict multiple falls in 270 older people. Force platform is an instrument commonly used in gait analysis that gives the total force applied by the foot to the ground. Seven force plate variables were assessed to predict the risk of multiple falls. Falls were prospectively recorded during the following year.

The force platform variables such as the root-mean-square amplitude in media-lateral directions (RMS-ML), in the single-task condition predicted multiple falls together with the following covariables: history of multiple falls, use of medications (fall-risk medications or multiple medicine use), and gender. In addition, multiple fallers had a narrower stance width than non-fallers.

[OBS+97] suggested a simple assessment tool to be used in a hospital setting to target prevention programs for elderly patients at high risk of falling. First, they conducted a case-control study to investigate 21 possible risk factors for falling in the elderly. As a result, five factors were identified as significantly associated with falls. Next, they compared the fallers and controls; they used Student's t-tests for continuously distributed data, Wilcoxon's non-parametric test for categorical scales (such as abbreviated mental test score), and logistic regression for categorical data. Finally, they calculated the odds ratios for all differences to select variables for constructing the risk assessment tool. The obtained five factors were: report of previous falls by patients; a transfer and mobility score (judged by the nursing staff); judgment that the patient was agitated (visual assessment done by the nurse); required frequent toileting; and has visual impairment (as indicated by the physician).

A risk assessment score (range 0-5) was developed by scoring one point for each factor. In the final phases, a risk assessment score >2 was used to define the high risk of falling: the sensitivity and specificity of the score to predict falls during the following week were 93% and 88%, respectively, in phase 2 and 92% and 68% in phase 3. The results look promising based on the performance; however, the number of subjects is low, the data used is subjective, and the method requires a lot of time for each subject in need of a diagnosis.

1.5.2 Machine Learning Classification Methods

Using clinical data and sensor data as inputs to machine learning algorithms was introduced as an alternative to statistical analysis to quantify kinematics precisely and predict future falls.

Building the machine learning models requires a long-term evaluation of large samples of subjects' locomotion and complex feature engineering of sensor kinematics. Once the machine learning model is built, it provides a faster and more objective diagnosis than the statistical methods that require a lot of time to spend with each patient to collect data. In addition, clinical tools and statistical methods have the disadvantage of being subject to a clinician's judgment. Therefore, creating an objective fall-risk detection model that can efficiently measure risk factors with minimal costs is critical.

Supervised Machine learning techniques have been used to classify fall risk based on fall history using data from wearable sensors during gait. The classification methods used in the 25 studies reviewed were: Radial basis function network (RBFN), Support Vector Machines (SVM), k-nearest neighbors (kNN), Naive Bayes (NB), Multilayer perceptron (MLP), Locally Weighted Learning (LWL), neuroevolution of augmenting topologies (NEAT), Decision Trees (DT), Linear Discriminant, Discriminate analysis, discriminate classifier, Majority Classifier, Random Forest, Artificial Neural Networks (ANN), XGBoost, and Convolutional Neural Networks (CNN). Table 1.6 indicates the classification model used in these papers.

Ref. F:NF	Labeling Method	Sensors	Experiment Protocol	Clinical Features	Feature Selection	Classification Model	Model Valid.	Acc. %	Sens. %	Spec. %	AUC
[BJVKM11] 40:41	retrospec.	accel.	walking straight, TUG	TUG metrics	-	Logistic Regression	-	77	78	78	0.83
[CKdSH+11] 15:05	retrospec.	accel.	6 functio -nal tests	functional test metrics	Forward Wrapper	RBFN,SVM kNN, NB	Leave-one out cross-valid.	95	100	80	-
[JTB+11] 40:40	retrospec.	accel.	walking straight		-	NB, MLP, SVM, LWL DT NEAT	10 fold cross-valid.	61-82	58-80	62-84	-
[MRW+11] 19:27	prospec.	accel.	walking straight, TUG	TUG metrics	Forward Wrapper	DT, Logistic Regression	10 fold cross-valid	80	74	96	0.87
[LRNL11] 22:46	prospec.	accel.	TUG, STS5	gender, age, tests' metrics	SFFS	MLP, linear discriminant	Leave-one out cross-valid.	78	59	90	-
[PHL11] 54:43	prospec.	accel.	walking on a circuit		-	Logistic Regression	-	67	74	58	-
[WHP+11] 23:18	retrospec.	accel.	TUG	TUG metrics	-	Logistic Regression	-	63-88	65-91	50-83	-
[YAN+11] 16:29	retrospec.	Wii Balance Board	balance while standing	game scores	-	Discriminate Analysis	-	89			-
[GDW+12] 83:143	prospec.	accel.	TUG	TUG metrics	-	Discriminate Classifier	-	73-83	56-90	73-96	-
[GMW+12] 65:55	retrospec.	accel.	standing balance	balance metrics		SVM	10-fold cross-valid.	63-72	68-82	59-67	-
[SSG+12] 50:50	Tinetti test	accel.	walking straight			Linear Regression	-	-	-	-	0.85
[DWF+13] 19:20	retrospec.	accel. gyros.	STS5	STS5 scores		Logistic Regression	Leave-one out cross-valid.	74	69	80	0.7
[HOT+13] 16:57	prospec..	accel.	walking straight			Logistic Regression	-	69	84	-	0.81
[RTP+13] 44:90	retrospec.	accel.	treadmill walking			Logistic Regression	-	73	21	97	-
[NYU+13] 41:111	retrospec.	laser range finder	choice stepping test			Logistic Regression	-		70	-	0.73
[CRS+14] 22:44	Tinetti BBS BESTest	Kinect	standing sitting balance	postural sway	Relief-F	Majority Classifier, DT SVM, kNN, NB	bootstrap	48-84	48-91	48-83	-
[KMS+14] 07:05	physician's assessment	Kinect	GUGT	GUGT metrics and posture	-	SVM	Leave-one out cross-valid.	67	67	68	-
[KCP15] 18:55	prospec.	Wii Balance Board	balance and TUG	TUG metrics and BMI	-	Logistic Regression	-	-	-	-	0.71
[HLK16] 24:76	retrospec.	accel. pressure insoles	normal walk and with cognitive load		-	MLP, NB, SVM	-	72-84	33-100	74-100	-
[SMT+17] 51:245	retrospec.	accel. pressure platform	TUG, STS 4-stage	age, gender, BMI, medication, and health	Forward Selection	kNN,NB,Random Forest,DT ANN,SVM, Linear, and Logistic Regression	10 fold cross-valid. with random split	88	-	-	-
[GRC16] 11:11	retrospec.	accel.	TUG	gender, height, weight, age vision, health TUG metrics	nested cross-valid.	Logistic Regression	10x 10-fold cross-valid.	72.7	90.91	54.50	-
[HLK18] 28:47	prospec.	accel. pressure insoles	normal walk and with cognitive load		Relief-F	ANN, NB, SVM	10,000 random 75:25 train :test stratified holdouts	65	59		
[NYG+21] 290:456	retrospec.	accel. pressure insoles	slower, faster, normal walking	age, height weight, BMI, education level	-	XGBoost	10-fold cross-valid.	67-70	43-53	77-84	0.7
[RJBB21] 53:45	physician's assessment	accel.	TUG, 4-stage balance	age, gender, BMI functional test scores	-	CNN	-	66	87	42	0.75
[BND+21] 54:54	retrospec.	pressure insoles	normal walk and with cognitive load	gender	Relief-F	kNN, SVM, ANN, DT, Logistic Regression	100x 10-fold cross-valid.	81	80	78	-

Table 1.6 – Characteristics of the Reviewed Studies for the Prediction of Fall Risk in the Elderly

1.6 Discussion

This literature review investigated 25 existing studies in the period from 2011 to 2022 concerning machine learning techniques using clinical features and features extracted from sensors in fall risk prediction in the elderly community. Table 1.6 summarizes our findings.

The population size ranged from 12 to 746. It should be noted that *the choice of the population and its size affect the classification models' performance and generalization potential*. Larger sample sizes provide more reliable, precise, and dependable results, but with the cost of more computational energy, time and money. The authors in [CKdSH+11] (accuracy of 95%), [KMS+14] (accuracy of 67%), and [GRC16] (accuracy of 72.7%) used a low population size of 20, 12, and 22, which makes their suggested approach unlikely to achieve their reported performance in a clinical setting.

Another issue is the *population age*. In [BJVKM11], they studied 40 fallers versus 41 non-fallers which is good population size. However, the average age between the two groups was very different: the fallers had an average age of 21.6 while the non-fallers had an average of 79.1. They achieved high accuracy in their classification model (77%). However, it is also unlikely to perform that well dealing with fallers and non-fallers in the same elderly age group.

The studies were also different in terms of *labeling methods*. For example, 6 out of 25 studies involved prospective fall event tracking, and 14 out of 25 studies applied retrospective fall history. In comparison, only 4 out of the 25 studies used clinical assessment tests and physician's assessments.

Even though the retrospective fall recall is relatively widely used and has been proven to lead to a better performance in classification models of fallers and non-fallers,

it may suffer from a lack of reliability because of the patient's poor memory. In addition, a long history of falling may lead to a change in the original gait pattern due to injury or psychological fear of falling. It is also believed that the clinical assessment tools (TUG, Tinetti, BBS, and STS5) still do not achieve perfect diagnostic accuracy. While prospective fall occurrence follow-ups are relatively time-consuming and inconvenient to achieve. They are recommended to be employed in future fall risk assessment research with a follow-up period of at least 6 months after the initial assessment.

The studies also varied in the *sensor technologies* used, whether wearable sensors such as accelerometers, gyroscopes, pressure insole, and non-wearable sensors such as Wii Balance Boards and Microsoft Kinect. Non-wearable sensors have the benefits of making the participants feel in a more normal setting, walking freely without having anything attached to them. However, they have just recently entered this field of research. That said, they have yet to prove if they can provide equal or better accuracies than wearable accelerometers and pressure sensors, which have been used extensively in measuring human gait for a long time. Sensor technology has proven to be a practical, affordable, and somewhat accurate tool for fall risk assessment.

Some studies included in their *input features clinical factors* such as measurements made from traditional functional tests and physiological aspects of patients such as gender, BMI, age, diseases, and medications. Including more clinical risk factor as features in the classification model in future studies is believed to improve performance [GRC16].

Moreover, each study's experimental *movement protocol during data collection differed*. Some adopted the task of walking straight, others while performing functional tests or walking on a treadmill, with or without cognitive tasks. The involvement of cognitive tasks has shown to bring out more significant statistical difference between fallers and non-fallers [HLK18]. The metrics extracted during the experiments included

the duration and smoothness of specific movements and the continuous signals extracted from the sensors utilized. However, a large pool of variables may not be clinically relevant or confounding to other existing variables [SWL+15]. Including many features in the classification model can also be very exhaustive. Thus, selecting valuable features based on research evidence and feature selection methods is essential.

Therefore, some studies included in their classification models some *feature selection methods*, especially those dealing with large feature sets [NYG+21, BND+21, HLK18]. Applying feature selection methods when dealing with a large feature may help in decreasing the computational cost of the model significantly and thus leads to a faster diagnosis. After feature selection, the classification methods were applied.

12 out of 25 studies did *not apply proper model validation procedures*, which can lead to exaggerated diagnostic accuracy and are unlikely to maintain their reported performance during practical use in the concerned elderly population [SWL+15]. This observation emphasizes the need for proper guidelines and standardized model construction/validation procedures in future research [SWL+15].

Consequently, future work in this field needs to consider improved classification model performance, minimal preparation time, simple instructions, user-friendliness, and real-time result display have been reported as critical factors for the continued use of technologies [50]. However, the disconnection between clinical functionality and user experience evaluation remains a significant gap that needs to be addressed.

1.7 Conclusion

Implementation of accurate fall risk assessment and prediction methods in the elderly community has the potential to improve the quality of care and lead to a reduction in

associated hospital costs due to fewer admissions and reduced injuries due to falling.

However, the literature on the prediction and assessment of elderly fallers shows that not enough research covering the topic is available. Furthermore, the existing studies require further improvement in terms of performance and convenience to be used in a clinical setting with simplicity in data collection and reduced duration. In addition, the gap between functional evaluation and user experience with this technology should be addressed.

Therefore, the extension of such research is crucial. Furthermore, the highly extraordinary predictive power associated with the help of direct measurements of cognitive function (such as mental tracking as a dual task) highlights an essential avenue for future research in fall risk prediction.

Chapter 2

Cyclostationary Analysis

Après la bibliographie détaillée présentée sur la prédiction des chuteurs âgés, nous commençons par donner quelques définitions sur la cyclostationnarité pour mieux comprendre le mécanisme à l'origine des phénomènes qui seront abordés dans les chapitres suivants. Ce chapitre vise à introduire le concept de cyclostationnarité et les outils nécessaires au développement de modèles analytiques pour la prédiction de chutes chez les personnes âgées.

After the in-depth literature review presented on the prediction of elderly fallers, we start with giving some definitions of cyclostationarity to understand better the mechanism behind the phenomena, which will be considered in the following chapters. This chapter aims to introduce the concept of cyclostationarity and the tools necessary for developing analytical models to predict future falls in the elderly.

2.1 Introduction

Diagnostic tools based on signal processing are increasingly sophisticated and provide exciting results in several fields, such as mechanics and biomechanics. However, in the beginning, many of these tools were established, assuming that the process being treated is stationary, whereas most processes encountered in nature, such as walking, have *time-evolving parameters*. This makes the tools, assuming stationarity, unable to extract specific information related to the non-stationary character of these processes resulting from nature. Thus, the diagnostic tools dedicated to non-stationary processes seem adequate to extract the maximum information.

Classical analysis tools approach the signal from a stationary point of view. Therefore, they are unable or insufficient to deal with the non-stationary processes encountered during this thesis. Several algorithms and methods have been proposed to deal with *non-stationary signals and processes*. These tools include the sliding Fourier transform, the spectrogram, the Wigner-Ville representation, quadratic time-frequency distributions, and methods based on wavelet theory.

The signals processed in our work have a *cyclostationary* character, a particular case of non-stationarity. Cyclostationarity is defined as a pairing between a deterministic periodic phenomenon and a random one. This particularity makes it possible to add dimension to the traditional indicators characterizing its cyclical evolution. Thus, to exploit the cyclostationarity, it is essential to analyze this cyclical evolution through the periodicity of the statistical moments of the process.

2.2 History of cyclostationarity

Cyclostationary signals from nature are sometimes characterized by periodic and random behavior. In other words, cyclostationary signals are random processes with a periodic structure in time. A signal is characterized as cyclostationary if its *statistical properties vary periodically over time.* However, it is impossible to detect these periodicities through classical tools based on stationarity, hence their name of hidden periodicities. The processing tools developed within the framework of cyclostationarity have made it possible to have information on these hidden periodicities despite the presence of considerable noise in specific processes. Thus, one of the advantages of cyclostationary analysis lies in its robustness and performance in noisy environments.

The notion of cyclostationarity appeared in 1958 in a theoretical context in the field of telecommunications with Bennett [Ben58]. Soon after, several Russian mathematicians introduced vital concepts necessary to represent cyclostationary processes. The first was Gudzenko [Gud59], who studied nonparametric spectral estimation of cyclostationary processes. Then, between 1961 and 1963, Gladyshev conducted a study on the analysis of spectral properties and the relationship between cyclostationary processes and stationary vector sequences [Gla61]. The latter also introduced the concept of nearly-cyclostationary processes in [Gla63]. Subsequently, the notion of cyclo-ergodicity for cyclostationary processes, having a single periodicity, was introduced by Nedoma in 1963 [Ned63]. Finally, this notion was generalized, in 1983, by Boyles and Gardner for cyclostationary processes with multiple periodicities [BG83].

Over time, cyclostationarity has gained its place in the field of signal processing, so several works, theses, and books have addressed the new concept of cyclostationarity. In 1962, Franks devoted a whole part of his book to cyclostationary processes in

the theory of telecommunications [Fra69]. Hurd's thesis [Hur70] is considered a good introduction to continuous-time cyclostationary processes after studying the nature of the spectral support in this case of processes. During this work, he established a link between cyclostationary processes and processes having the possibility of being stationary by applying a time shift (*i.i.d.*). In 1975, Gardner and Franks [GF75] studied the advantages of different representations of cyclostationary processes in the context of optimal filtering. Indeed, Gardner made significant contributions to the theory of cyclostationarity and applications of cyclostationary analysis. Two of his works are dedicated to the treatment of cyclostationary processes following two different approaches, namely a (classical) probabilistic approach [Gar86] and an approach called *Fraction Of Time* [Gar88]. A detailed synthesis of the two approaches was established by Gardner himself in 1991 [Gar91].

In 2006, Gardner, Napolitano, and Paura published a work gathering the developments of cyclostationarity during the previous fifty years [GNP06]. Since then, cyclostationarity and its tools have been widely used in different fields. Indeed, cyclostationarity has also been applied in climatology [LHBS95], in mechanics [ABRB04], in econometrics [PP79], in biology [FJ91], and more recently in biomechanics [CSL17]. Serpedin, Panduru, Sari, and Giannakis have published an excellent bibliographic article [SPSG05] containing a huge part of the work done on cyclostationarity and its applications in many fields.

2.3 Definitions and properties

2.3.1 Stochastic processes

A stochastic approach is essential for modeling certain processes resulting from nature, often non-stationary. A random (stochastic) process, denoted by $\{X(t,w), t \in \mathbb{R}, w \in \mathbb{Z}\}$, is defined as a set of continuous-time processes. This set is called realizations of $X(t)$ and indexed by w. Furthermore, stochastic processes are characterized by their probability densities.

The first stochastic processes studied are stationary processes, given their ease of processing and their constant statistical properties over time. Despite their efficiency in dealing with a few processes, they remain insufficient to extract much of the information related to the non-stationarity of several processes. Hence the usefulness of cyclostationary processes, which include stationary processes as a particular case. Unlike stationary processes, the statistical properties of cyclostationary processes depend on time. This allows access to additional information across time-varying moments and cumulants.

Consider a continuous-time random process with real values $X(t)$ admitting as distribution function $F_{X(t)}(x)$ (with $x \in \mathbb{R}$) and defined by:

$$F_{X(t)}(x) = Prob\{X(t) < x\} \tag{2.1}$$

The random process $X(t)$ is said:

— Cyclostationary in the strict sense at order n, if these statistical properties up

to order n are periodic ($\forall$ the period $\mathcal{T}$) according to time [Gar94]:

$$F_{X(t)}(x) = F_{X(t+\mathcal{T})}(x) \tag{2.2}$$

— Cyclostationary in the broad sense at order n, if these statistical properties up to order n are periodic for the same period $\mathcal{T}_0$ according to time [Gar94].

$$F_{X(t)}(x) = F_{X(t+\mathcal{T}_0)}(x) \tag{2.3}$$

2.3.2 Cyclostationarity of order one and two

The statistical properties of a cyclostationary process are periodic in time. The distribution function $F_{X(t)}(x)$ makes it possible to describe a stochastic process fully. Unfortunately, it is usually impossible to determine the distribution function in an experimental setting with natural processes. To overcome this problem, most stochastic processes are often characterized by the process's first-order and second-order statistical properties. Thus, the periodicity of the latter is necessary to characterize the processes as first-order and second-order cyclostationary.

2.3.2.1 Cyclostationarity of order one

The first moment of a random process $X(t)$ corresponds to the mean. A random process is said to be cyclostationary of order one if its average is periodic in time:

$$\mu_X(t) = \mu_X(t + T_0) \tag{2.4}$$

Where $\mu_X(t) = \mathbb{E}\{X(t)\}$ represents the first-order statistic of a random process $X(t)$, and T_0 denotes the cyclic period.

2.3.2.2 Cyclostationarity of order two

The cross-correlation function makes it possible to measure the correlation between two real random variables X and Y at two different instants t_1 and t_2. Thus, we can understand the intensity of the connection between these variables. This function is expressed as follows:

$$R_{X,Y}(t_1, t_2) = \mathbb{E}\{X(t_1)Y^*(t_2)\} \tag{2.5}$$

Where $(.)^*$ denotes the conjugate. By replacing $Y(t)$ by $X(t-\tau)$, we obtain the instantaneous autocorrelation function of the random process $X(t)$ as a function of the delay τ. The advantage of this function is that it allows the detection of repeated profiles in a random process. There are two versions of the instantaneous autocorrelation function, namely:

— The symmetric version $R_X(t,\tau) = \mathbb{E}\{X(t-\tau/2)X^*(t+\tau/2)\}$

— The asymmetric version $R_X(t,\tau) = \mathbb{E}\{X(t)X^*(t-\tau)\}$

It is important to note that the instantaneous autocorrelation functions $R_X(t,\tau)$ represent a statistic of order 2. Thus, we say that a random process $X(t)$ is cyclostationary of order two if its instantaneous autocorrelation function is periodic according to time:

$$R_X(t,\tau) = R_X(t+T_0,\tau) \tag{2.6}$$

where T_0 represents the cyclic period. Moreover, if the mean $\mu_X(t)$ and the instantaneous autocorrelation $R_X(t,\tau)$ are periodic at the same period T_0, the process X(t) is said to be cyclostationary in the broad sense and T_0 is its cyclic period.

If a stochastic process is cyclostationary of order two, its instantaneous autocorrelation function is periodic with time. Consequently, the instantaneous autocorrelation

$R_X(t,\tau)$ admits a Fourier series decomposition according to time t:

$$R_X(t,\tau) = \sum_{\alpha\in\mathcal{D}} R_X^\alpha(\tau)e^{j2\pi\alpha t}, \text{with } \mathcal{D} = \{\alpha = \frac{k}{T_0}, k \in \mathbb{Z}\} \tag{2.7}$$

Here the sum is performed over the set $\mathcal{D}$ of cyclic frequencies α that are multiples of the fundamental cyclic frequency $\alpha_0 = \frac{1}{T_0}$. Thanks to this decomposition, we have access to the Fourier coefficients $R_X^\alpha(\tau)$, called the cyclic autocorrelation function [Gar86]. Indeed, this function makes it possible to see the fundamental cyclic frequency and its harmonics clearly, and is expressed as follows:

$$R_X^\alpha(\tau) = \lim_{T\to+\infty} \frac{1}{T} \int_{-T/2}^{T/2} R_X(t,\tau)e^{-j2\pi\alpha t}dt \tag{2.8}$$

It is important to note that the cyclic autocorrelation function $R_X^\alpha(\tau)$ is discretized along α and continuous along τ. Indeed, it is non-zero only for the fundamental cyclic frequency α_0 and its harmonics in addition to $\alpha = 0$. This is due to the periodicity of the instantaneous autocorrelation function $R_X(t,\tau)$ according to time. For $\alpha = 0$, the cyclic autocorrelation function is reduced to the classical autocorrelation function. Thus for a stationary process, $R_X^\alpha(\tau)$ is zero everywhere except at $\alpha = 0$.

2.3.3 Second-order frequency descriptors

According to the Wiener-Khintchin theorem, a stationary process's power spectral density (PSD) is defined as the Fourier transform of its autocorrelation function. The extension of this theorem in the cyclostationary case [Gar88] makes it possible to establish the following relation:

$$S_X(t,f) = F_\tau\{R_X(t,\tau)\} \tag{2.9}$$

Where $S_X(t,f)$ denotes the instantaneous power spectral density (Wigner Ville spectrum) and F_τ represents the Fourier transform for the delay τ. In the case of a second-order cyclostationary process, the instantaneous autocorrelation functions $R_X(t,\tau)$ is periodic according to time which implies that the instantaneous PSD $S_X(t,f)$ is also periodic over time. Thus, the quantity $S_X(t,f)$ admits a Fourier series decomposition according to time:

$$S_X(t,f) = \sum_{\alpha \in \mathcal{D}} S_X^\alpha(f) e^{j2\pi\alpha t}, \text{with} \quad D = \{\alpha = \frac{k}{T_0}, k \in \mathbb{Z}\} \tag{2.10}$$

Where T_0 represents the cyclic period. Furthermore, the Fourier coefficients $S_X^\alpha(\tau)$ are called the spectral correlation function and are expressed as follows:

$$S_X^\alpha(f) = \lim_{T \to +\infty} \frac{1}{T} \int_{-T/2}^{T/2} S_X(t,f) e^{-j2\pi\alpha t} dt \tag{2.11}$$

By simple calculations and substitutions using the equations 2.7, 2.9, and 2.10, we arrive at the following relation:

$$S_X^\alpha(f) = F_\tau\{R_X^\alpha(\tau)\} \tag{2.12}$$

Thus, to calculate the spectral correlation function $S_X^\alpha(\tau)$ it suffices to calculate the Fourier transform of the cyclic autocorrelation function $R_X^\alpha(\tau)$ with respect to the delay τ. Therefore, $S_X^\alpha(\tau)$ inherits the same properties as $R_X^\alpha(\tau)$. As for the cyclic autocorrelation function, the spectral correlation $S_X^\alpha(f)$ is discretized on the axis of the cyclic frequencies α and non-zero only for the fundamental frequency $\alpha_0 = \frac{1}{T_0}$ and its harmonics. When $\alpha = 0$, the spectral correlation is reduced to the classic DSP defined for stationary processes.

2.3.4 Ergodicity and cycloergodicity

Computing first- and second-order statistical properties requires computing the process's mean. Furthermore, the estimation of statistical properties requires the probability densities obtained by exploiting several identical systems called realizations. However, in reality, it is often very difficult to have access to a large number of achievements. On the other hand, observing a process for a long time is possible and exploitable. This makes it possible to calculate temporal averages, which can replace ensemble averages. Thus, a stochastic process is said to be ergodic if the temporal and ensemble averages are equal for a single realization [CL90]. Theoretically, the temporal average is calculated over an infinite window in time. However, in practice, this is impossible, and we must content ourselves with measuring the average over sufficiently long windows. Consequently, for an ergodic process, the ensemble means are approximated by the time means as follows:

$$\mathbb{E}\{X(t)\} = \langle X(t)\rangle_t \tag{2.13}$$

Where $\langle . \rangle_t$ represents the time average operator and is expressed by:

$$\langle . \rangle_t = \lim_{T \to +\infty} \frac{1}{T} \int_{-T/2}^{T/2} (.) dt \tag{2.14}$$

With the appearance of the notion of cyclostationarity, Gardner and Boyles [BG83] became interested in studying ergodicity in the case of cyclostationary processes. Thus, the concept of cycloergodism was born through the extension of the notion of ergodism to the case of cyclostationary processes. This concept implies that asymptotically the time averages of a process and the time averages of the statistical properties of the same process are equal if they are multiplied by a complex sinusoid. This is expressed

mathematically by:

$$\langle \mathbb{E}\{X(t)e^{j2\pi\alpha t}\}\rangle_t = \langle X(t)e^{j2\pi\alpha t}\rangle_t \tag{2.15}$$

In the case where the process is assumed to be cyclostationary and cycloergodic, the relationship can be exploited to estimate the cyclic autocorrelation function via the following two expressions:

— The symmetrical version $R_X^\alpha(\tau) = \langle X(t-\frac{\tau}{2})X^*(t+\frac{\tau}{2})e^{-j2\pi\alpha t}\rangle_t$

— The asymmetrical version $R_X^\alpha(\tau) = \langle X(t)X^*(t-\tau)e^{-j2\pi\alpha t}\rangle_t$

With the cycloergodicity hypothesis, one can estimate the statistical descriptors of a cyclostationary process from the synchronous time averages (or synchronous averaging) of a single realization. This is very useful in estimating second-order statistical properties for processes assumed to be cyclostationary and cycloergodic.

2.4 Conclusion

Through this chapter, we have defined the concept of cyclostationarity and its analysis tools. For the study of cyclostationary processes, several cyclostationary tools are proposed in the literature. In the following chapters, cyclostationary tools are exploited for modeling cyclostationary properties of pressure insole signals for fall prediction.

Chapter 3

Experiment and Data-set Description

Les précédents chapitres ont présenté le contexte de cette thèse a travers une étude bibliograpique et l'analyse cyclostationnaire. Dans ce chapitre, nous décrivons en détail l'ensemble de données utilisé dans cette thèse: la conception de l'expérience, la collecte de données et les propriétés statistiques de la population.

The background of this thesis related to previous literature and cyclostationary analysis has been covered in the previous chapters. In this chapter, we describe in detail the data-set used in this thesis, including the experiment design, the data collection, and the statistical properties of the population.

3.1 Introduction

In clinical research, the main goal is to design a study that would be able to derive a valid and meaningful scientific conclusion using appropriate methods that can be translated to the "real world" setting [LH07]. Prior to designing an experiment, one must establish the aims and objectives of the study and choose an appropriate target population that is most representative of the studied population. The research study's conclusions aim to improve health care and diagnosis procedures. Hence, this requires a well-designed clinical research study that rests on a solid foundation of a detailed methodology and is governed by ethical principles [UMG11].

This study aims to explore the cyclostationary properties of the pressure insole signals of the elderly and develop methods to diagnose risks of future falls. Therefore, the targetted population is the older people. The total number of participants was 519 people. They have all provided consent to participate in this study. This database is from the original series of the study by the LPE (Laboratoire de Physiologie de l'Exercice), and CHU (Centre Hospitalier-Universitaire) of Jean Monnet St-Etienne University. The below sections explain the steps dedicated to experimental design and data collection.

3.2 Experiment Design

519 elderly adults were recruited to partake in this experiment at the Hospital University of Saint Etienne. They were instructed to *walk in the same straight 20 meters hallway while wearing insole pressure sensors*. They follow a protocol :

— At first, they walked for a *test trial*.

— After that, each participant was asked to walk this same distance three times.
  — The first time, which formed the baseline in this study, they simply walked without performing secondary tasks, which is denoted the MS walking condition.
  — The second time, they walked the same distance again, but while *de-counting from the number 50*. It is denoted the MD walking condition. The counting backward task depends on *working memory*. Working memory is a system for temporarily holding and manipulating information while performing various cognitive tasks such as comprehension, learning, and reasoning.
  — The third time, they walked while *enumerating aloud as many animal names as they could remember*, which is denoted the MF walking condition. The verbal fluency task relies on *semantic memory*. Semantic memory refers to general world knowledge accumulated throughout our lives.

  The last two types of walks included secondary tasks to simulate an everyday life situation. Typically, when one walks, they are not focusing all their attention on walking but rather are occupied by other thoughts. Also, it has been shown in the literature that involving secondary tasks brings out a more significant statistical difference between fallers and non-fallers.

In addition, specific measures were taken while collecting the data to block other factors that could influence participants' walk. These measures included ensuring proper lighting, a quiet area, and the use of comfortable flat shoes.

3.3 Data Collection

The data collected during the experiment were clinical data and sensor data. Clinical data were information regarding the participants' demographics, lifestyle, and medical

profiles. On the other hand, the sensory data were the signals recorded by the insole pressure signals during the three types of walks that the participants completed in the experiment.

3.3.1 Clinical Data

In the database, 410 subjects out of 519 reported whether they had previous falls or not. Moreover, out of 410, only 54 reported that they had fallen in the past, while the rest reported that they had not. One of the challenges faced in this thesis is the case of *highly unbalanced data*.

The demographic data collected about the subjects include gender and age. At the same time, lifestyle information included alcohol consumption and smoking. Physical information included height, weight,and Body Mass Index (BMI). The below figure 3.1 shows the distribution of labeled data between fallers and non-fallers according to gender, BMI class, alcohol consumption and smoking status.

3.3.2 Sensor Data

The participants wore a pair of innersole pressure sensors fitted inside their shoes. Each innersole contains two independent SMTEC pressure sensors. These sensors were fixed in specific positions to record the pressure signals from the heels and toes of the participants' left and right feet while walking. These sensors were connected to a portable data logger that the participants wore at their waist. In this setup, a pressure above 4 kilo-Pascal (kPa) activates the sensors and defines the contact with the ground. For each step, the activation of the heel sensors establishes the first contact, whereas the last contact with the ground corresponds to when the toes sensor stops measuring.

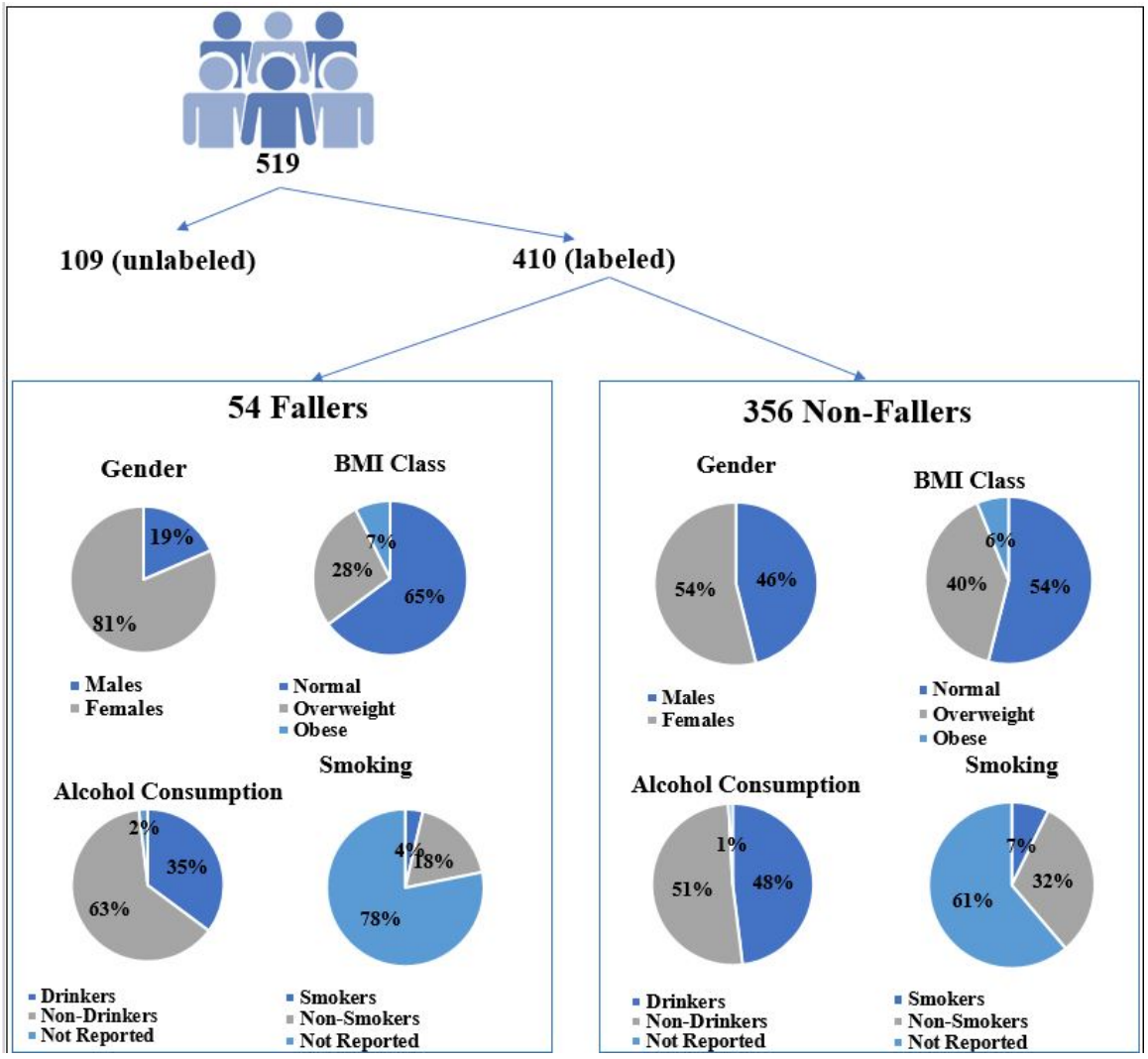

Figure 3.1 – Data-set Description

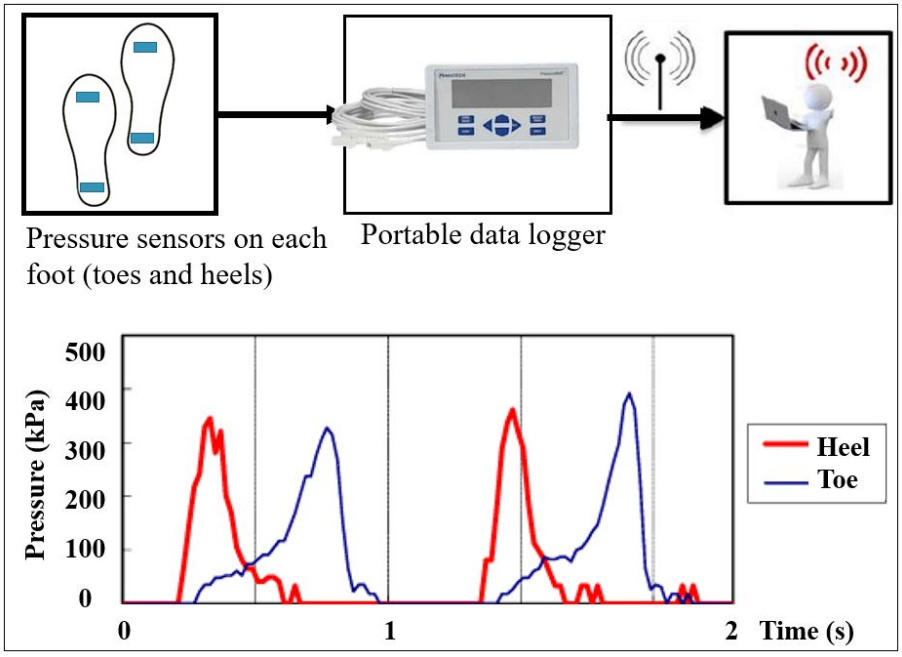

Figure 3.2 – The sensor system and pressure signal recorded from a participant's foot.

As mentioned by the manufacturer, the data from foot switches is ***sampled at a frequency of 100 Hz*** allowing a temporal resolution of 10 ms. The signals collected were processed using software designed specifically for the task by SMTEC software. The system measures four independent pressure signals: left heels, left toes, right heels, and right toes. Figure 3.2 displays the setup along with the obtained signal from the pressure sensors while the subject is walking.

Some patients do not necessary succeed to walk 20 meters due to elderly and health problems. It means a poor number of steps.

3.4 Conclusion

This chapter describes the data-set studied throughout this thesis from the experimental design perspective. The different characteristics recorded about participants will be studied in later chapters to see which of them qualify to serve as essential features for

future fall prediction. The setup used to record the signals studies was also explained. In addition, the data-set brings out one main challenge addressed in this thesis: the unbalance of data.

Chapter 4

Feature Extraction and Selection

L'extraction d'indicateurs correspond au processus de transformation des données brutes en caractéristiques qui peuvent être traitées pour classer ou prévoir un résultat. Ce pré-traitement permet d'avoir de meilleurs résultats pour l'apprentissage automatique par rapport à l'utilisation de données brutes. La sélection d'indicateurs, est une méthode de réduction de variable d'entrée conservant uniquement les données pertinentes et éliminant ainsi le bruit dans les données. Il s'agit du processus de sélection automatique des fonctionnalités pertinentes pour le modèle d'apprentissage en fonction de l'objectif de l'étude. Ce chapitre présente les techniques utilisées pour l'extraction et la sélection des caractéristiques afin de prévoir le risque de chutes chez les personnes âgées en mettant l'accent sur les caractéristiques cyclostationnaires.

Feature extraction refers to the process of transforming raw data into features that can be processed to classify or predict an outcome. As a result, it yields better results than applying machine learning directly to the raw data. As for feature selection, it is the method of reducing the input variable to a model by using only relevant data and getting rid of noise in the data. It is the process of automatically choosing relevant features for the machine learning model based on the study's objective. This chapter presents the techniques used for feature extraction and selection to predict the risk of future falls in the elderly with focus on cyclostationary features.

4.1 Introduction

Machine learning helps to find complex and potentially useful patterns in data automatically. These patterns are condensed in a machine learning model that can then be used on new data points—a process called making predictions or performing inference. Building a machine learning model is a multi-step process. Each step presents its own technical and conceptual challenges.

This thesis focuses on the supervised learning task and the process of selecting, transforming, and augmenting the source data to create a powerful predictive ability for the target variable, which is the risk of future falls in the elderly. These operations combine domain knowledge with machine learning techniques and are the essence of feature engineering. When a machine learning model is used to make predictions, one should apply the same transformations used for the training data on the new subjects that require a diagnosis. By applying the same transformations, the machine learning model receives the data as expected.

Pre-processing the data for machine learning involves both data engineering and feature engineering. Data engineering converts raw data (the data in its source form, without any prior preparation for machine learning) into prepared data (this involves signal pre-processing such as removal of outliers). Feature engineering or feature extraction then tunes the prepared data to create the features expected by the ML model. Engineered features are created by performing specific machine learning-specific operations on the columns in the prepared data-set and creating new features for the model during training and prediction. These operations include scaling numerical columns to a value between 0 and 1 and clipping values. The following diagram, Figure 4.1, shows the steps involved in preparing the data before feeding it to the machine

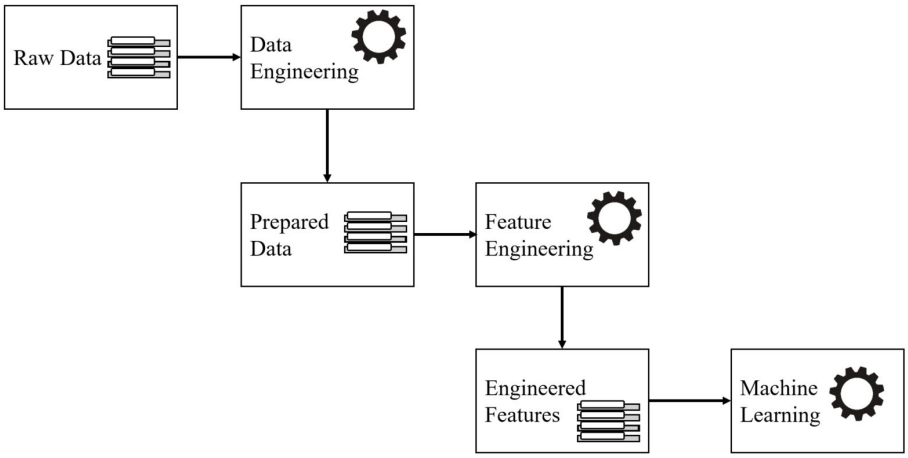

Figure 4.1 – The flow of data from raw data to prepared data to engineered features to machine learning.

learning models.

This chapter addresses the methodology of transforming raw data into engineered feature sets ready to be used in a machine learning model. This process comprises data pre-processing, filtering, normalization, feature extraction, and selection of feature sets.

4.2 Data Engieering and Pre-Processing

This section addresses the raw data preparation and pre-processing steps taken before the feature engineering process.

4.2.1 Removal of Signal Outliers

As a first step in pre-processing the signals and preparing them for feature extraction, the first 3 seconds of all recordings were cropped to avoid possible noise caused by the system or its surroundings upon the beginning of the recording.

The second step in data pre-processing is dealing with outliers. Outliers are a

common problem in data science and signal processing. There are many popular methods for identifying outliers, such as the standard score method [GV12] or John Tukey's fences [Tuk93]. This thesis identified outliers as a point beyond the low and high percentiles used to define a given threshold. After identifying the outliers, two standard approaches are used to deal with them: Trimming or Winsorizing [GV12]. Winsorizing data means replacing the extreme values of a data set with a specific percentile value from each end, while Trimming or Truncating involves removing those extreme values. Winsorizing refers to changing the outliers to the minimum and maximum percentiles or the average between the previous and the next sample. Although trimming is the most common method for dealing with outliers, Winsorizing is worth discussing in dealing with our data because trimming has a potential drawback: loss of data collected from other sensors. The most straightforward way to remove outliers is to drop the observations from the data-set. In addition, dropping the outliers from one signal leads to dropping those observations from the other signals recorded simultaneously from different sensors (left and right heels and toes). Since observations that are outliers from one sensor may be inliers for others, one could trim more inliers than outliers. Data loss can be avoided by setting outliers to null values instead of dropping them. However, since many modeling software cannot manage null values, this could lead to another problem. An essential advantage of Winsorizing is that there is no loss of data. Winsorizing outliers does not lead to dropping other inliers from other sensors. Another essential property of Winsorizing is preserving some of the original information. Since Winsorizing reduces the weight of outliers without eliminating them, the former outliers still influence models or statistical calculations. For example, the pseudo-code of the function that removes the outliers in Python is shown below. The outliers, once identified, are replaced by the average of the previous and preceding samples (the "smooth" option is chosen rather than the "saturate").

```
def removeOutliers(signal,q=[5,95],k=2,replace='smooth'):
1. Input the raw signal with outliers.
2. Define what are the low and high percentiles thresholds.
   5 and 95 are chosen as percentile[0] and percentile[1] respectively.
3. Set k which is a coefficient used to define outliers.
   For this application k is set to 2.
3. Calculate the pmid which is the midpoint between the two percentiles.
3. A sample is an outlier if:
    * it is greater than pmid+k.(percentile[0]-pmid)
    * it is lower than pmid+k.(percentile[1]-pmid)

4. replace the outlier:
           * 'saturate': replace by computed threshold value
           * 'smooth'  : replace by average of previous and
                         next sample
   In this application the option 'smooth' is only used.

5. Return a clean signal
```

4.3 Feature Engineering

This section explains the steps taken to extract features after cleaning the data and before the feature selection phase.

4.3.1 Extraction of Time-Domain Features

The classical features extracted from the pressure signals were previously explored in the literature and are: mean, rise time, fall time, pulse width, overshoot, undershoot, duty cycle, slew rate, mid-cross, autocorrelation, standard deviation, band power, median, root mean square, range, Pwelch, skewness, interquartile range, kurtosis, and 95 percentile of the signal distribution. The definition of each feature is explained below.

- **Mean**

  The mean, indicated by μ, is the average value of a signal. It can be calculated by adding all of the values for each sample together, and divide by the total number of samples (N). Its mathematical form is shown below:

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i \tag{4.1}$$

- **Standard Deviation**

  The standard deviation (σ) is a measure of how far the signal fluctuates from the mean (μ). A low standard deviation indicates that the values tend to be close to the mean of the set. In contrast, a high standard deviation indicates that the values are spread over a broader range. It is calculated using the below equation.

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2}. \tag{4.2}$$

- **Median**

  The median (Med) is the middle value in a sorted, ascending, or descending list of values (X). The median is sometimes used instead of the mean when there are outliers in the sequence that might skew the average of the values. It is the

point above and below which half (50%) of the observed data fall and represents the data's midpoint. If there is an odd amount of sample points (N), the median value is the number in the middle, with the same amount of sample points below and above. If there is an even number of sample points (N) in the data, the middle pair must be determined, added, and divided by two to find the median value. In a normal distribution, the median is the same as the mean and the mode.

$$Med(X) = \begin{cases} X[\frac{N+1}{2}], & \text{if N is odd} \\ \frac{X[\frac{N}{2}]+X[\frac{N+1}{2}]}{2}, & \text{if N is even} \end{cases} \tag{4.3}$$

- **Root Mean Square**

  The Root Mean Square (RMS) is the square root of the arithmetic mean of the squares of a set of values. Its formula is mentioned below.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} x_i^2} \tag{4.4}$$

- **Range**

  The range is defined as the difference between the signal's maximum and minimum values in amplitude.

$$Range(X) = Max(X) - Min(X) \tag{4.5}$$

- **Rise Time**

  The rise time is the time a signal crosses from a specified low value to a specified high value. For example, in analog and digital electronics, the specified lower and higher values are 10% and 90% of the final or steady-state value. So the rise time is how long it takes for the signal to go from 10% to 90% of its highest

value as shown in Figure 4.2.

- **Fall Time**

  The fall time is the opposite of the rise time. It is the time taken for the amplitude of a pulse to decrease from a specified value (90% of the peak value) to another specified value (10% of the maximum value)as shown also in Figure 4.2.

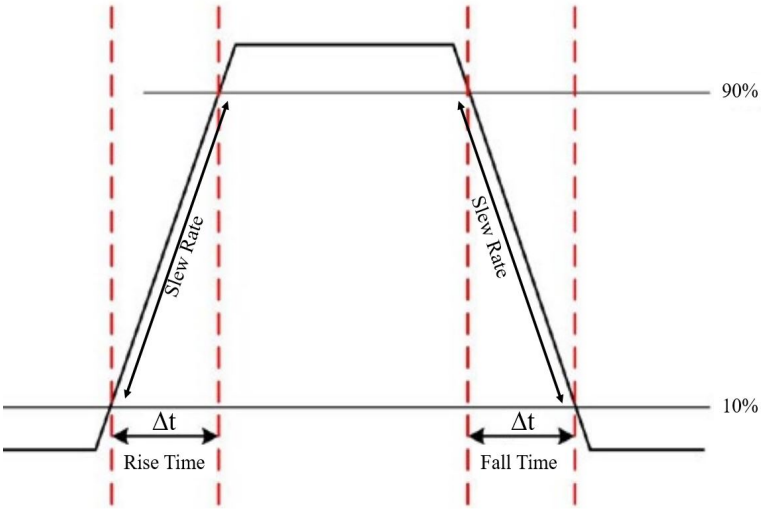

Figure 4.2 – Rise time and fall time indicate the length of time a signal takes to change voltage between the low level and high level. In addition the slew rate represents the the slope of the line connecting the 10% and 90% reference levels.

- **Pulse Width**

  The pulse width measures the elapsed time between a single energy pulse's leading and trailing edges. The below figure shows an example of measured pulse width in the signal.

- **Overshoot**

  In control theory and signal processing, the manifestation of a signal that exceeds its target is known as the overshoot. The overshoot is the greatest absolute deviation larger than the final state levels or the expected peak of each transition in the signal. An overshoot is expressed as the percentage difference between

the state levels. The overshoot length corresponds to the number of transitions in the input signal. The sample instants in the signal correspond to the vector indices.

- **Undershoot**

  Overshoot occurs when the transitory values exceed the target or expected value. However, when they are lower than their target, the phenomenon is called the undershoot. Undershoot is the greatest deviation below the final state levels of each transition in the signal. An undershoot is described as a percentage difference between the state levels. The length of the undershoot corresponds to the number of transitions detected in the signal.

- **Duty Cycle**

  The duty cycle of a signal computes the portion of time a given source is transmitting a signal. This portion of time determines the overall power delivered by the signal. The duty cycle is the pulse width ratio to the pulse period for each positive-polarity pulse in the signal. It identifies all regions that cross the upper-state boundary of the low state and the lower-state boundary of the high state.

- **Slew Rate**

  In electronics, the slew rate is defined as the maximum rate of change of the output voltage per unit of time. The slew rate is the slope of the line connecting the 10% and 90% reference levels as shown in Figure 4.2.

- **Mid-Cross**

  It is the time instants where each transition of the input signal crosses the 50% reference level as shown in Figure 4.3.

- **Autocorrelation**

  Autocorrelation is a mathematical depiction of the similarity between a given time series and a lagged version of itself over consecutive intervals of time. The

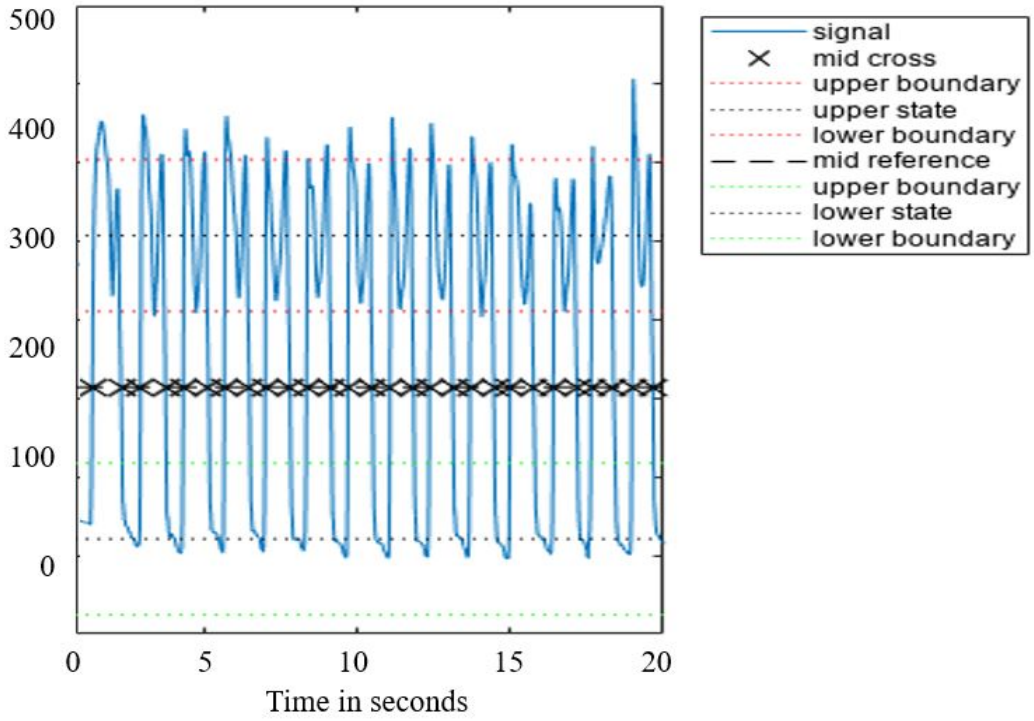

Figure 4.3 – The x defining the mid cross points in the left foot signal (heel sensor)

relationship between a variable's current value and its past values is measured. The average autocorrelation at different lags is calculated for each participant.

- **Band Power**

  Band power is the average power of the input signal. The power of a signal is the sum of the absolute squares of its time-domain samples divided by the signal length or, equivalently, the square of its RMS level.

- **P-Welch**

  The energy variation that takes place within a vibrating signal is known as power spectral density (PSD). It's derived as the frequency per unit of mass. In other words, for each frequency, the spectral density function shows whether the energy that is present is higher or lower. The average P-Welch power spectral density estimate is calculated for each participant in the data-set.

- **Skewness**

Skewness measures a signal's asymmetrical spread about its mean value. A signal is considered symmetric if it looks the same to the left and right of the center point.

When a data set has outliers, variability is often summarized by a statistic called the interquartile range, which is the difference between the first and third quartiles IQR=Q3-Q1.

- **Kurtosis** Kurtosis estimates whether the data distribution is heavy-tailed or light-tailed relative to a normal distribution. Data points with high kurtosis tend to have heavy tails or outliers. Whereas data sets with low kurtosis tend to have light tails or lack outliers. A uniform distribution would be the extreme case.

4.3.2 Extraction of Cyclostationary Features

4.3.2.1 Resampling: Walk Signal Synchronization Techniques

In order to process the walking pressure insole signals, we intended to benefit from the advantage of their repetitive or cyclic aspect. The technique for synchronization is demonstrated in this section.

Synchronizing the signals is equivalent to displaying the signals based on the stride number instead of time. Since then, the synchronized element will always appear in the same position inside the stride. To achieve this synchronization, it is essential to first identify the signals' significant and identical elements at each stride, for example, the maximum peaks. Once an event is identical to each stride, it is possible to re-adjust the signal by stretching or contracting it by interpolation to make all the events coincide periodically.

One technique consists of detecting the maximum peaks present in the signal. The

search for the maximum peaks is done by looking for the change of signs (positive or negative) in the derivative.

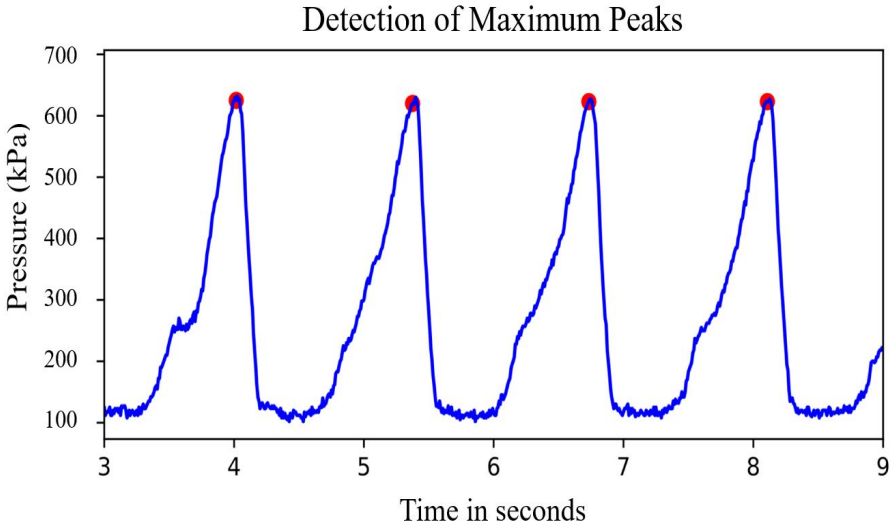

Figure 4.4 – Detection of Maximum Peaks to be used for synchronization of the signals

In order to combat the sensitivity of the derivative to noise, the derivative is estimated by performing a regression at each point polynomial by the method of Savitzky and Golay [SG64] on a neighborhood of 80 ms and eliminating slopes less than 30%. The results are materialized by the red dots shown in Figure 4.4.

This maximum peaks detection enables us to estimate the average walking period which is different for each person. Therefore, we can make automatic resampling.

The first step is to transform the signal $x(t)$ into an exponential :

$$\tilde{y}(t) = a(t).e^{j\phi(t)} \tag{4.6}$$

Where :

— a(t) is the amplitude of the envelope signal,

— $\phi(t)$ is the phasis.

Two steps are necessary to obtain $\tilde{y}(t)$:

— Band pass filter the signal $x(t)$ around the stride frequency in order to obtain a sine wave with a frequency modulated by the stride rhythm. We obtain the signal shown in Figure 4.5.

— Transform this sine wave into an exponential by using analytic signal (by simply removing negative frequency in Fourier Transform).

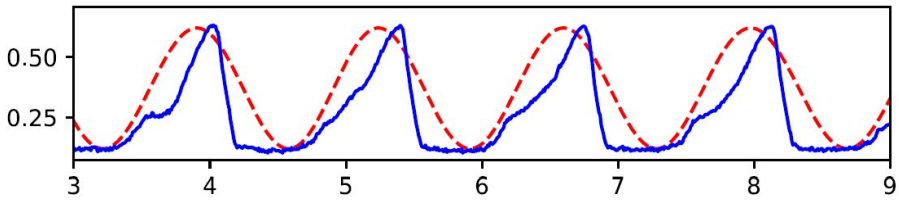

Figure 4.5 – Sine wave (red) vs. pressure insole signal (blue)

The interesting information is contained in the phase $\phi(t)$ of the sine wave and the exponential. This phase can be estimated by computing the angle of the analytic signal, and performing phase unwrapping.

Physically, the phase will increase to 2π at every stride and enables us to have the stride number according to the time (by dividing phase by 2π). Non-integer values from could be seen as a progression in the walk movement (physical interpretation could be difficult).

Since the stride number (i.e. $\phi(t)/(2\pi)$) and amplitude $x(t)$ are known for every instant t, it is possible to reconstruct the amplitude associated to a given walk stride n° as shown in Figure 4.6.

Estimating the signal against the stride number $x(\theta)$ could be done by using interpolation :

$$x(\theta) = interp1(\theta(t)/(2\pi), x(t), \theta) \tag{4.7}$$

Where :

— $\theta(t)/(2\pi)$ is the stride number,

— θ is a vector which indicate the stride number (it goes from 0 to the total number of strides with a constant step increment)

Working with walk "position" enables us to synchronized walking signal as shown in figure 4.7.

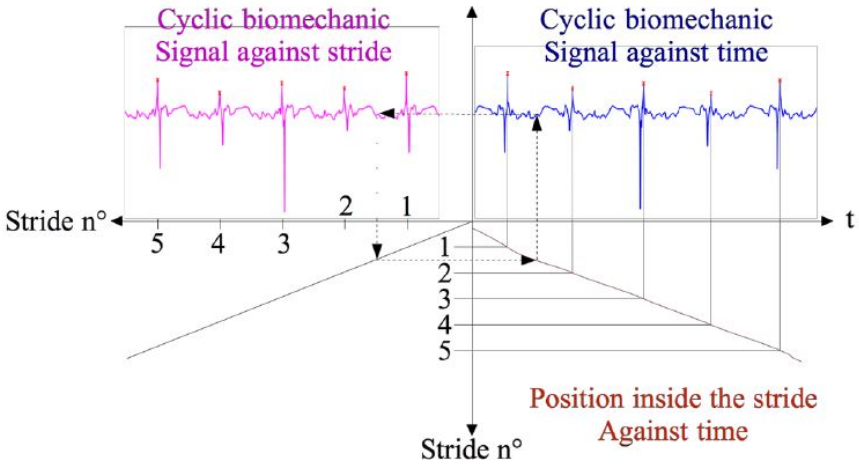

Figure 4.6 – Compensation of Speed Fluctuation

Once the maximum peaks are detected, the angular resampling algorithm is applied to align the peaks and obtain the synchronous average which displays the average pattern of the signal as shown in Figure 4.7.

4.3.2.2 Degree of Cyclostationarity

A cyclostationary process is a signal that holds statistical properties that varies cyclically with time. In the research field of human locomotion, the human walk can be thought of as a movement that consists of replicated sequences of cyclic physical movements or strides [ZTEB+14]. Analyzing the cyclostationary characteristics of the pressure insoles obtained during walking can can be used to assess the risk of prospective falls in the elderly community [ZTEB+14].

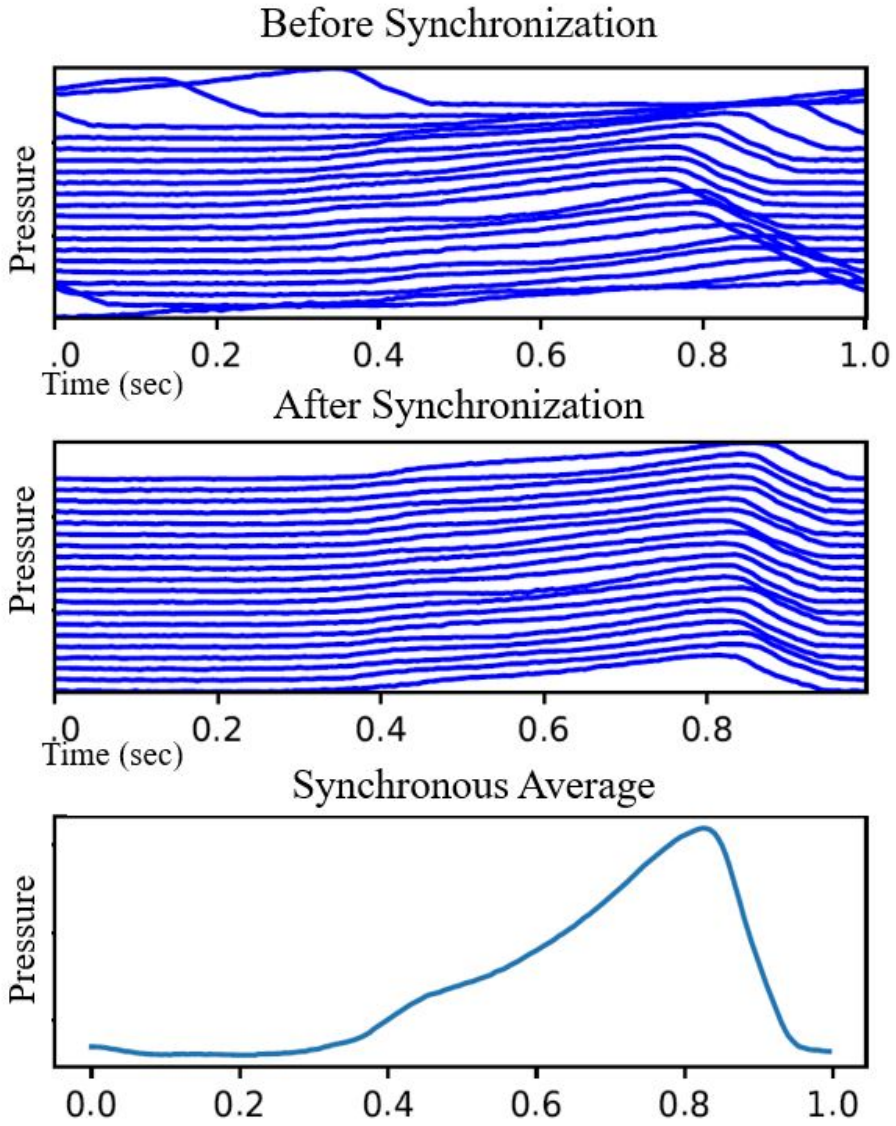

Figure 4.7 – The Effect of the Synchronization Technique and the Synchronous Average

Utilizing the properties of cyclostationarity demands a constant number of samples for every stride. Hence, it is crucial to preprocess the signal obtained to compensate for the speed oscillation. Preprocessing the signal can be done by estimating this fluctuation as described in the previous section[BEBR+05] and using interpolation to stretch the signal and offset speed fluctuation.

The work in [SG94] define the first two orders of cyclostationarity in a signal [SG94]. A signal $S(t)$ is regarded to be cyclostationary of order 1 with cycle T if the expectation $\mu_{S(t)}$ of $S(t)$, is periodic with period T:

$$\mu_{S(t)}(t) = \mu_{S(t)}(t+T) \tag{4.8}$$

$\mu_{S(t)}$ designates the repetitive pattern in the signal. The residual signal $r(t)$ can be calculated by subtracting $\mu_{S(t)}$ from the whole signal:

$$r(t) = S(t) - \mu_{S(t)}(t) \tag{4.9}$$

A signal is declared to be cyclostationary of order 2 if the autocorrelation $C_{S(t)}$ of the signal $S(t)$ is periodic with period T:

$$C_{S(t)}(t_1, t_2) = C_{S(t)}(t_1 + T, t_2 + T) \tag{4.10}$$

where t is defined as,

$$t = (t_1 + t_2)/2 \tag{4.11}$$

By defining τ as

$$\tau = t_2 - t_1 \tag{4.12}$$

We can write,

$$C_{S(t)}(t_2, t_1) = C_{S(t)}(t_1 + T, t_2 + T) = C_{S(t)}(t_2 - t_1) = C_S(t, \tau) \tag{4.13}$$

In cases where a signal is proven to be cyclostationary of order 2, such as the pressure insole signals analyzed in this study, the instantaneous autocorrelation function of the signal is periodic. It, therefore, can be represented as a Fourier series as shown below [ŽG91].

$$C_{S(t)}[t, \tau] = \sum_{\alpha} CAF_S[\alpha, \tau] e^{(-j2\pi\alpha t)} \tag{4.14}$$

where, α is the cyclic frequency that belongs to the set of cyclic frequencies such that, α=k/T and $k \in$ Z. The cyclic autocorrelation function (CAF) is defined as:

$$CAF_S[\alpha, \tau] = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} C_S[t, \tau] e^{(-j2\pi\alpha t)} \tag{4.15}$$

To minimize the effect of order 1 cyclostationarity in the cyclic autocorrelation, we can calculate the cyclic autocorrelation $CAF_R[\alpha, \tau]$ of the residual signal $r(t)$, from Equation 4.7, instead of the autocorrelation $CAF_S[\alpha, \tau]$ of the whole signal $S(t)$.

With the CAF being a 3-dimensions representation of the cyclostationarity of the pressure walking signal, it is also feasible to generate another 3-dimension model called the spectral correlation $SCD_R(\alpha, f)$. This is done by applying the Fourier Transform of τ to get the frequency f. Therefore, the entire data is held within the cyclic frequencies α associated with the characteristic cycles of the signal. At other cyclic frequencies, the spectral correlation should have null energy. In Fig.4.8, the cyclic frequencies unrelated to signal cycles are not entirely flat but considered negligible since this is an estimated

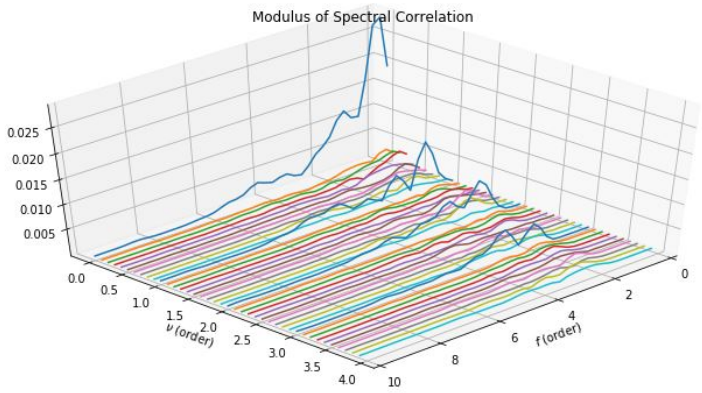

Figure 4.8 – The Spectral Correlation of the average insole pressure signals while walking without involving any secondary tasks (MS)

representation. The method use for the estimation of the spectral correlation is found in [AXH17].

4.3.3 Heat Map Representation of the Spectral Correlation

The analysis of the graph in Figure 4.8 of the spectral correlation is complicated. In addition, it is difficult to use this graph directly as input to machine learning algorithms. Therefore, we investigated the transformation of this graph into a 2-dimensions heat-map in which colors represents the magnitude of the spectral correlation. The x-axis in the heat maps obtained represents the cyclic frequencies α order and the y-axis represents the frequency f order. Figure 4.9 shows an example of heat-map images, which will be used as input to the CNN to classify individuals with high and low risk of prospective falling. Fig. 4.10 shows 5 examples of each of the two classes of participants with and without risk of falling. The differences between the two classes can be noticed visually as the straight lines corresponding to the high modulus of spectral correlation at $\alpha{=}k/T$ and $k \in \mathrm{Z}$ can be seen clearly in cases of participants without risk of falling and vaguely in cases of patients with risk of falling. but can be captured computationally using

CNN.

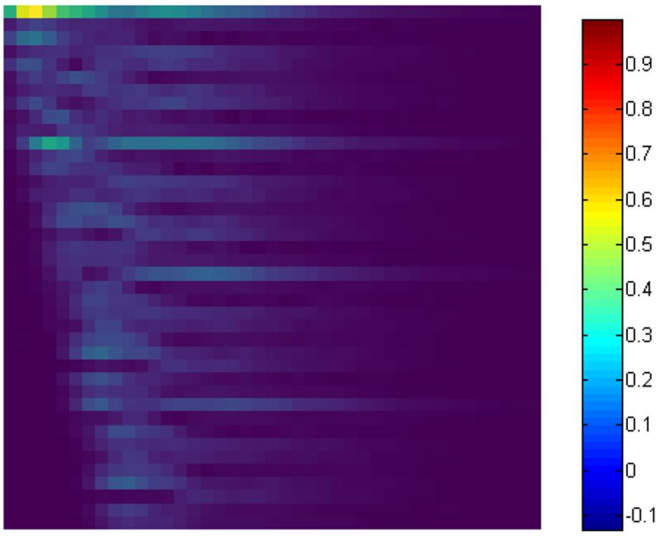

Figure 4.9 – The Heat Map representation of the Spectral Correlation in Figure 4.8

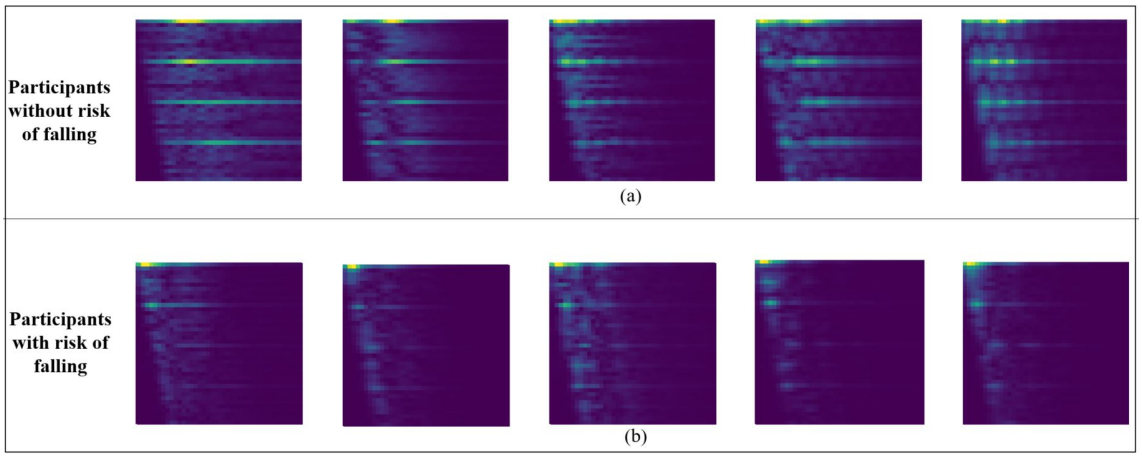

Figure 4.10 – 5 Examples of the Raw Heat Maps (colored) of the two classes: (a) without risk and (b) with risk of falling during MD walking condition

4.3.4 Heat Map Image Quality improvement

We sought to further improve the heat map image quality by trying to make the lines differentiating between cases of fallers and non-fallers more visible. In addition, we attempted to reduce their complexities before feeding them to the CNN model.

Therefore, two types of transformations were explored: the grey-scale and the log transformations.

The main reason why gray-scale representations are often used for CNNs and extracting descriptors instead of color images is that gray-scale simplifies the algorithm and reduces computational requirements, which is worth exploring.

The log functions are beneficial when the input grey-level values may have a vast range of values. This can be seen in the example in Figure 4.11 of the Fourier transform of an image is put through a log transform to reveal more details. And this can also be seen in the two images in Figure 4.12. The images obtained after performing the log transform on the colored raw heat map images are shown in Figure 4.13. The images obtained after only applying grey-scale on the heat map images are shown in Figure 4.14. The resulted images of applying both grey-scale and log transformation are shown in Figure 4.15.

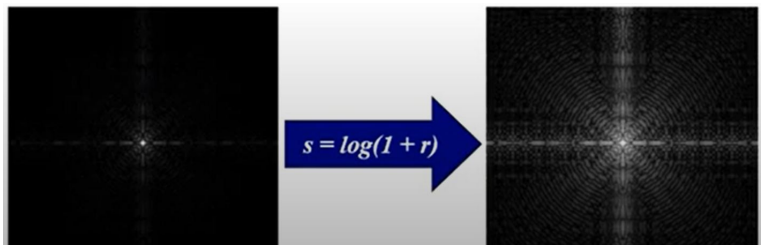

Figure 4.11 – An Example of a Log Transformation on a Grey-Scale Image

4.4 Statistical Analysis Techniques

4.4.1 Student t-test

Student's t-test, in statistics, is a method for testing hypotheses related to the mean of a small sample from a normally distributed population with an unknown standard

Figure 4.12 – An Example of a Log Transformation on a Colored Image

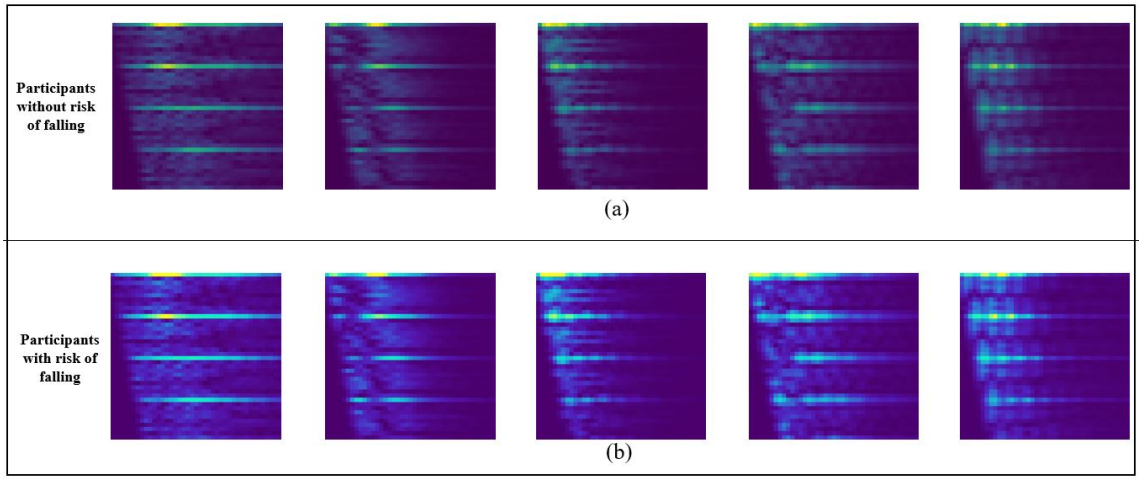

Figure 4.13 – 5 Examples of the Log Transformation of the Colored Heat Maps of the two classes: (a) without risk and (b) with risk of falling during MD walking condition

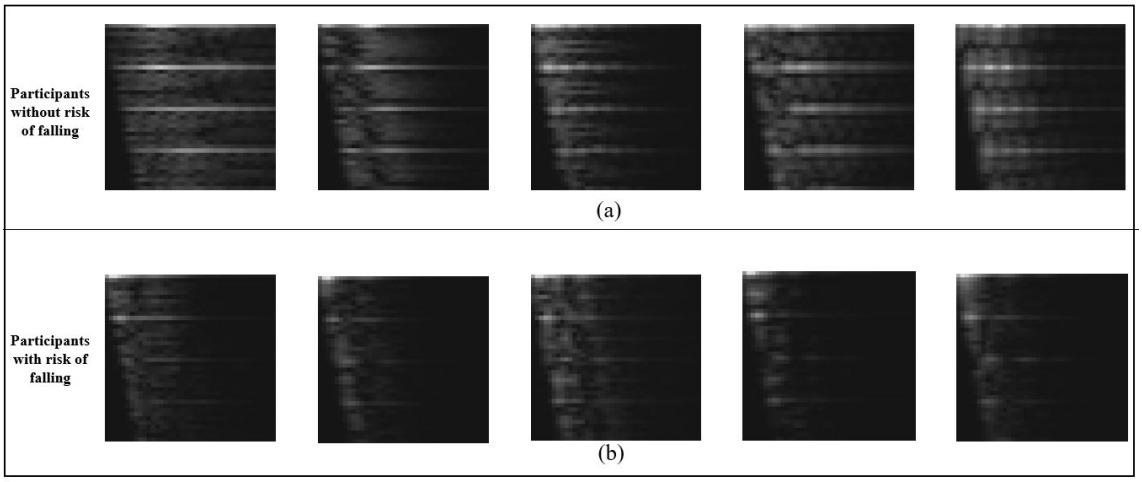

Figure 4.14 – 5 Examples of the Grey-Scale Heat Maps of the two classes: (a) without risk and (b) with risk of falling during MD walking condition

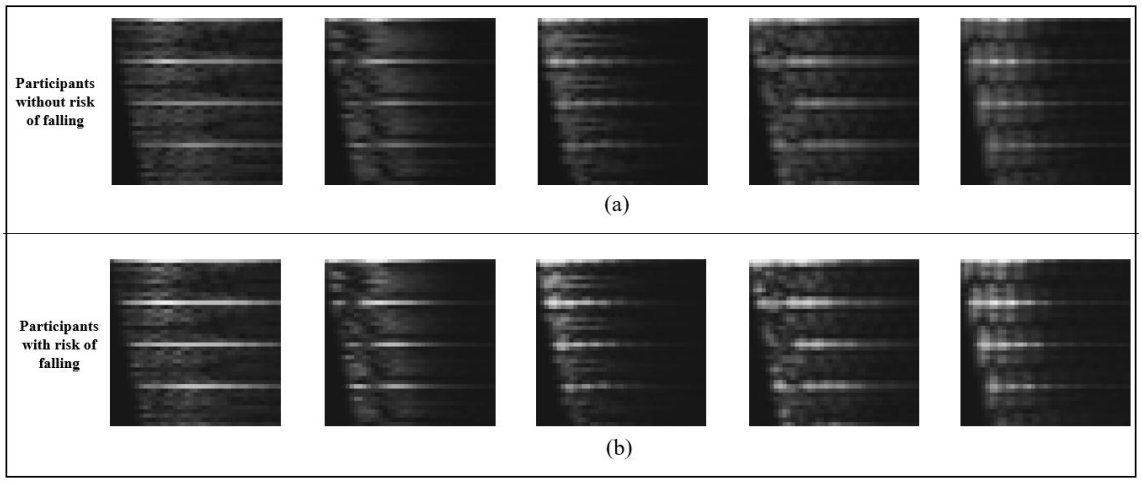

Figure 4.15 – 5 Examples of the Log Transformation of the Grey-Scale Heat Maps of the two classes: (a) without risk and (b) with risk of falling during MD walking condition

deviation It is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related [Rux06].

In 1908 William Sealy Gosset developed the t-test and t-distribution [FL01]. The t distribution consists of a family of curves where the number of degrees of freedom (number of independent observations in the sample minus one) determines a specific curve. As the sample size (and the number of degrees of freedom) increases, the t distribution approaches the bell shape of the standard normal distribution [FL01]. Technically, the normal distribution is usually applied for tests concerning the mean of a sample size greater than 30 [Rux06].

It is usually first to formulate a null hypothesis, which indicates that there is no practical difference between the observed sample mean. The hypothesized or stated population means that any measured difference is due only to chance. Generally, a t-test may be either two-sided (two-tailed), stating that the means are not equal, or one-sided, determining whether the observed mean is different from the hypothesized mean. The test statistic t is then computed [Rux06]. The null hypothesis is rejected when the observed t-statistic is more extreme than the critical value specified by the relevant reference distribution. The relevant reference distribution for the t-statistic is the t distribution. The critical value relies on the tests' significance level (the probability of wrongly rejecting the null hypothesis) [Rux06].

A second application of the t-distribution tests the hypothesis that two independent random samples have similar means. The t distribution is also used to produce confidence intervals for the actual mean of a given population or the difference between two sample means [Rux06].

4.4.2 Analysis of variance (ANOVA)

Analysis of variance (ANOVA) is a statistical method used to split an observed aggregate variability within a data set into two parts: systematic and random factors. The systematic factors statistically influence the given data set, while the random factors do not. Analysts use the ANOVA test to determine independent variables' influence on the dependent variable in a regression study [SW+89].

The ANOVA test is considered the first step in analyzing factors that influence a given data set. Once the test is complete, an analyst conducts further testing on the methodical factors that measurably affect the data set's inconsistency. Finally, the analyst uses the ANOVA test results in an f-test to develop additional data that aligns with the suggested regression models [SW+89].

The ANOVA statistical test permits a comparison of more than two groups simultaneously to resolve whether a relationship between them exists or doesn't exist. The result of the ANOVA formula, the F statistic (F-ratio), allows the analysis of multiple data groups to determine the variability between samples and within [SW+89].

If no difference exists between the tested groups (null hypothesis), the outcome of the ANOVA's F-ratio statistic will be near 1. The distribution of all potential values of the F statistic is the F-distribution. This is a group of distribution functions with two characteristic numbers: the numerator and denominator degrees of freedom [SW+89].

4.5 Feature Selection Methods

In machine learning, feature selection, also called attribute selection, variable selection, or variable subset selection, includes techniques to reduce the input variable

to the machine learning model by using only relevant data and eliminating noise in data. It is the process of automatically choosing relevant features for the classification or prediction model depending on the problem that needs to be solved. The motivation behind using feature selection techniques is that the data comprises some redundant or irrelevant features and should be removed without causing a loss of information.

Feature selection techniques are different from feature extraction. Feature extraction produces features from the input data, whereas feature selection outputs a subset of the extracted features as shown in Figure 4.16. Feature selection techniques are often used in cases with many features and comparatively few samples, such as the analysis of handwritten words and genetic sequencing [SH12], where there are many thousands of features and just a few tens to hundreds of samples.

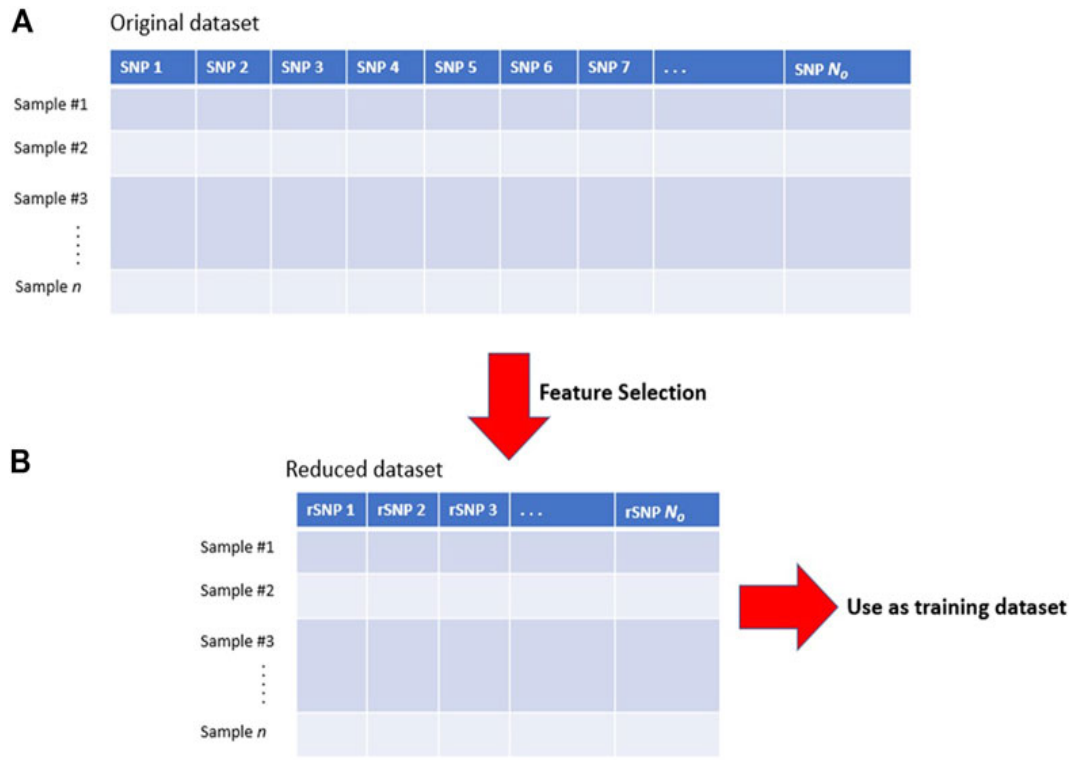

Figure 4.16 – An Illustration of the feature selection process. (A) The original dataset may contain excessive number of features (SNPs) that are irrelevant. (B) Feature selection reduces the dimensionality of the data-set by excluding irrelevant features and including only those relevant for prediction [PFKLO22].

There are many benefits to including feature selection in a machine-learning algorithm; this includes:

— Simplification and reduction of machine learning models's complexity, which makes them easier to interpret, faster to train, and less computationally expenssive.

— Avoiding the curse of dimensionality. The curse of dimensionality is a phenomenon observed in machine learning fields. It describes the explosive nature of increasing data dimensions and its resulting exponential increase in computational efforts required for its processing and analysis. Bellman [BK59] first introduced this term to explain the increase in the volume of Euclidean space associated with adding extra dimensions in the area of dynamic programming. An increase in the dimensions can, in theory, add more information to the data, thereby improving the quality of data but practically increasing the noise and redundancy during its analysis as shown in Figure 4.17.

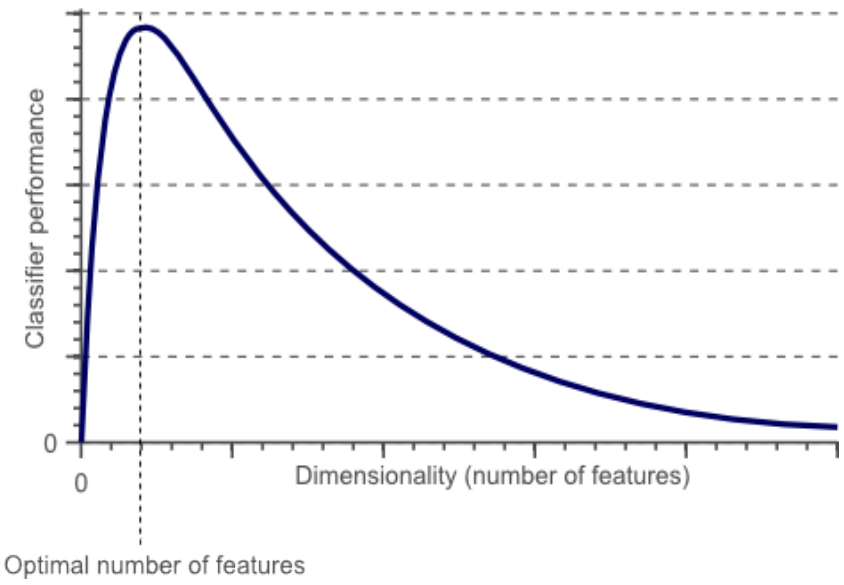

Figure 4.17 – The classifier's performance continues to increase with the number of features until the optimal number is attained. Then, further increasing the number of features without increasing the number of samples leads to a decreased classifier performance [Spr14].

4.5.1 Relief-F

The *Relief* algorithm was proposed in 1992 by Kira and Rendell [KR92a][KR+92b]. According to Kira and Rendell, the *Relief* algorithm weights each feature according to its relevance to the class. Initially, all weights are set to *zero* and then updated iteratively. In each iteration, this non-deterministic algorithm chooses a random instance i in the data-set and estimates how well each feature value of this instance distinguishes between instances close to i. In this process, two groups of instances are selected: some closest instances belonging to the same class and some belonging to a different category. With these instances, *Relief* will iteratively update each feature's weight, differentiating data points from other classes while simultaneously recognizing data points from the same category. In the end, a certain number of features with the highest weights is selected. An alternative version sets a threshold, so only the features with weights above this value are selected.

The *Relief* algorithm was described as a simple, fast, and practical approach to attribute weighing [KR92a][KR+92b]. The output of the *Relief* algorithm is a weight between -1 and 1 for each attribute, with more positive weights indicating more predictive attributes. The pseudo-code for *Relief* is shown in Figure 4.18. The weight of an attribute is iteratively updated as follows:

1. A sample is selected from the data.
2. The nearest neighboring sample that belongs to the same class (nearest hit) and the nearest neighboring sample that belongs to the opposite class (nearest miss) are identified by Euclidean distance.
3. A change in attribute value accompanied by a change in class leads up to the weighting of the attribute based on the intuition that the attribute change could

be responsible for the class change.

4. On the other hand, a change in attribute value accompanied by no change in class leads to a down weighting of the attribute based on the observation that the attribute change did not affect the class.

This procedure attribute weight updating is performed for a random set of samples in the data or every sample in the data. The weight updates are then averaged so that the final weight is in the range $[-1\ ,\ 1]$. The attribute weight estimated by *Relief* has a probabilistic interpretation. It is proportional to the difference between two conditional probabilities, namely, the probability of the attribute's value being differently conditioned on the given nearest miss and nearest hit, respectively [KR92a][KR+92b].

Algorithm Relief
Input: for each training instance a vector of attribute values and the class value
Output: the vector W of estimations of the qualities of attributes

1. set all weights $W[A] := 0.0$;
2. **for** i := 1 **to** m **do begin**
3. randomly select an instance R_i;
4. find nearest hit H and nearest miss M;
5. **for** $A := 1$ **to** a **do**
6. $W[A] := W[A] - \text{diff}(A,R_i,H)/m + \text{diff}(A,R_i,M)/m$;
7. **end**;

Figure 4.18 – The Pseudo-code of the basic Relief algorithm [KR92a][KR+92b].

Kononenko et al. proposed several updates to *Relief* to the method now called *Relief-F* [KRSP96] with the pseudo-code shown in Figure 4.19.

1. *Relief-F* used the near-hit and near-miss instances using the Manhattan (L1) norm rather than the Euclidean (L2) norm. Furthermore, it utilized the absolute differences between xi and near-hit and xi and near-miss to be sufficient when updating the weight vector (rather than the square of those differences).

2. In *Relief-F*, the contribution of missing values to the feature weight is determined using the conditional probability that two values should be the same or different,

approximated with relative frequencies from the data set. This can be calculated if one or both features are missing.

3. Instead of using Kira and Rendell's proposed *Relief* decomposition of a multinomial classification into several binomial problems, *Relief-F* searches for k near misses from each class and averages their contributions for updating W, weighted with the prior probability of each class.

Algorithm ReliefF
Input: for each training instance a vector of attribute values and the class value
Output: the vector W of estimations of the qualities of attributes

1. set all weights $W[A] := 0.0$;
2. **for** i := 1 **to** m **do begin**
3. randomly select an instance R_i;
4. find k nearest hits H_j;
5. **for** each class $C \neq class(R_i)$ **do**
6. from class C find k nearest misses $M_j(C)$;
7. **for** $A := 1$ **to** a **do**
8. $W[A] := W[A] - \sum_{j=1}^{k} \text{diff}(A,R_i,H_j)/(m \cdot k) +$
9. $\sum_{C \neq class(R_i)} \left[\frac{P(C)}{1-P(class(R_i))}\right] \sum_{j=1}^{k} \text{diff}(A,R_i,M_j(C))]/(m \cdot k)$;
10. **end**;

Figure 4.19 – The Pseudo-code of the extended Relief-F algorithm [KRSP96].

4.5.2 Sequential Backward Propagation

Sequential backward selection is a wrapper method to select the most relevant features for optimal model performance. Its evaluation uses criteria related to the classification algorithm used. The objective function is a pattern classifier, which evaluates feature subsets by their predictive accuracy (recognition rate on test data) by statistical re-sampling or cross-validation.

1. The criterion function is computed for all n features.
2. Each feature is deleted one at a time, the criterion function is computed for all subsets with n-1 features, and the worst feature is discarded.

3. Each feature among the remaining n-1 is deleted one at a time, and the worst feature is discarded to form a subset with n-2 features.

4. This procedure continues until a predefined number of features are left.

An example of using SBS is shown in Figure 4.20.

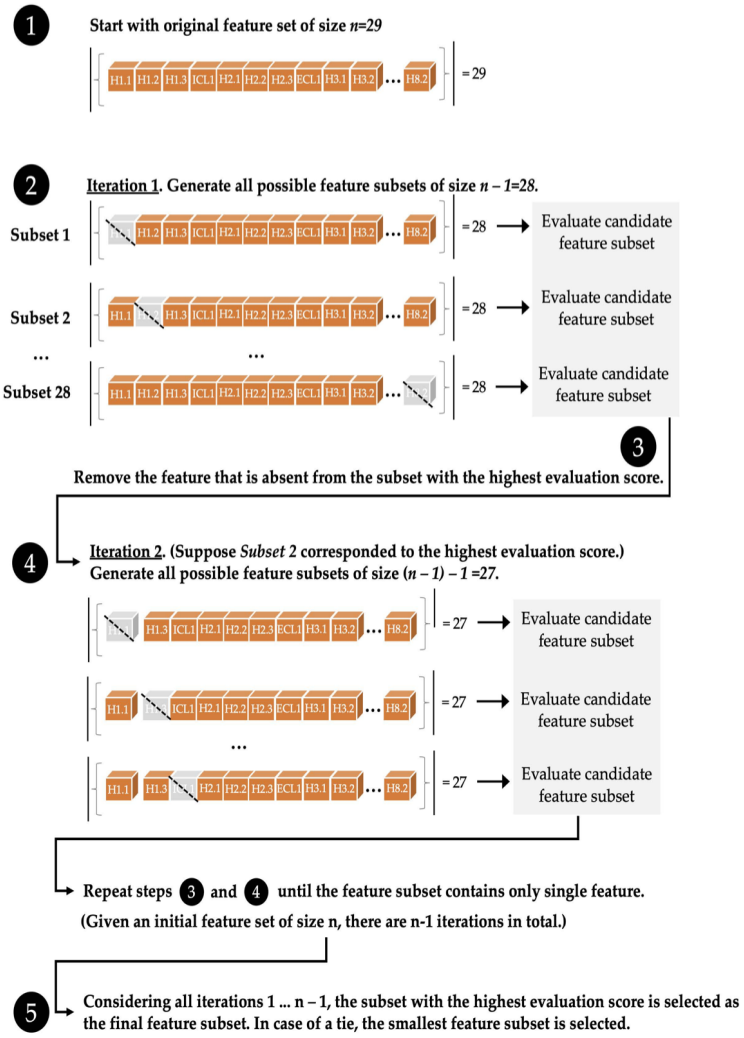

Figure 4.20 – An example of SBS for identifying feature subsets that maximize the performance of a machine learning model.

4.6 Conclusion

This chapter illustrated the steps taken before feeding the data to the machine learning models in the following chapter. The first steps included data cleaning from noise and outliers. Then, the feature extraction step is executed, including classical features previously explored in the literature and cyclostationary features. Statistical and machine learning feature selection is then applied to reduce dimensionality, improve the model's performance, and reduce computational costs.

Chapter 5

Machine Learning Classification Methods

Les outils de classification vont faire correspondre à chacune des donnée la catégorie à laquelle elle appartient. Une mise en correspondance des données d'entrée et des variables de sortie discrètes est réalisée. L'objectif principal est de déterminer à quelle classe/catégorie les nouvelles données appartiendront. Ce chapitre explique les différents outils de classification supervisée basés sur l'apprentissage automatique utilisés dans cette thèse.
Machine learning classification models categorize a given set of data into classes. The classification predictive modeling task approximates the mapping function from input data to discrete output variables. The main objective is to determine which class/category the new data will belong to. This chapter explains the different supervised classification machine learning classification models used in this thesis.

5.1 Introduction

Machine learning is a subfield of artificial intelligence (AI). It generally seeks to understand the structure of data and fit that data into models that people can easily understand and use [Zho21].

Machine learning can be used for a large variety of applications and is a constantly evolving area. Even though machine learning is a field of computer science, it is different from traditional computational techniques. In conventional computing, algorithms are packs of explicitly programmed instructions used by computers to compute or predict [Zho21]. Instead, machine learning algorithms allow computers to train on data inputs and use statistical analysis to predict outputs within a precise range. Machine learning simplifies computers' ability to build models from sample data to automate decision-making procedures from input data [Zho21].

In machine learning, methods are generally classified into broad categories. These categories are based on the learning process or how the feedback is given to the system developed. For example, two widely adopted machine learning categories are supervised learning and unsupervised learning. *Supervised learning* trains algorithms based on example input and output data humans label. On the other hand, *unsupervised learning* provides algorithms with no labeled data, allowing it to find structure within its input data [Zho21].

5.1.1 Unsupervised Machine Learning

Unlike in supervised learning, in unsupervised machine *learning the input data is without labels* (Figure 5.1), so the learning algorithm is forced to find commonalities

within its input data points [AAJM+20].

The objective of unsupervised learning may be as straightforward as discovering hidden patterns within a data-set. However, it may also have a feature learning goal, allowing the computational machine to discover the representations needed to classify raw data automatically [AAJM+20].

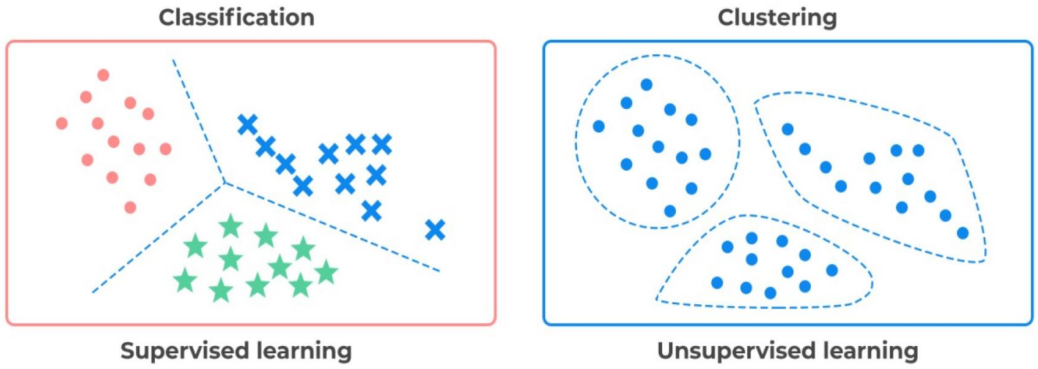

Figure 5.1 – Supervised and Unsupervised Machine Learning

5.1.2 Supervised Machine Learning

In supervised machine learning, *the algorithm is given example data inputs with labeled classes* [Kad19]. The algorithm then learns by comparing the data's actual category with the predicted output to find errors and adjust the model accordingly. Supervised learning, therefore, uses patterns to predict labels for new unlabeled input data [Kad19]. The data-set used in this thesis is provided with labels considering that participants' responses to whether they have fallen n the past or not are accurate.

In this thesis, five classical supervised machine learning algorithms were used on the data-set to classify elderly people with risk and no risk of future falls. These five classification methods were chosen based on the methods used in the literature of elderly fallers prediction detailed in the first chapter. The primary purpose was to compare the performances of each method.

Other types of well-known machine learning classification models were not explored in this thesis because of some characteristics they have that do not fit the type of data-set this thesis worked on. An example is the *Naive Bayes* machine learning algorithm, which was not used since it is better suited for categorical input variables than numerical variables [VSD15].

The below sections explain the algorithm of each supervised machine learning method used in this thesis: Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Artificial Neural Networks (ANN). The Table 5.1 summarizes the pros and cons of each method.

Table 5.1 – The Pros and Cons of Each Classification Algorithm

Method	**Pros**	**Cons**
Logistic Regression	Easy to implement,interpret, and efficient to train.	assumes linearity between the dependent and the independent variables.
Support Vector Machine (SVM)	Robust to overfitting and noise useful in non-linear problems	Computationally expenssive and runs slow.
K-Nearest Neighbors(KNN)	Easy to understand Robust to noisy data	Limitation of memory
Decision Trees	Interpretable, can handle missing data	Greedy (may not find the best tree)
Artificial Neural Networks (ANN)	Able to manage abundant number of data, appealing attributes of non-linear identification and control	requires a large amount of data for accuarte predictions, takes timein training

5.2 Logistic Regression

Logistic regression is a type of supervised machine learning that is often utilized for classification and predictive applications [HJLS13]. Logistic regression calculates the

probability of an occurrence of an event based on a given data-set with independent variables [HJLS13]. The output dependent variable is between 0 and 1 since it is a probability.

In Logistic Regression, a log transformation is applied to the odds, which is the probability of success divided by the probability of failure, which is also commonly known as the log odds or the natural logarithm of odds. The below equation represents the logistic function:

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{5.1}$$

Logistic Regression is named after the logistic function used at the method's core. The logistic function is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never precisely at those limits. For example, the below Figure 5.2 shows a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function [HJLS13].

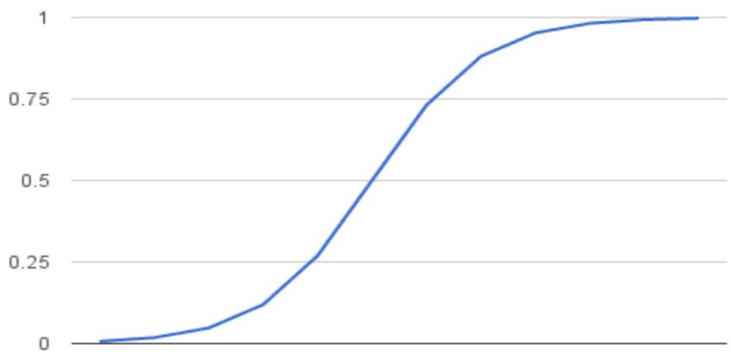

Figure 5.2 – The Plot of the Logistic Function.

Logistic regression uses an equation as the representation where the input values x are combined linearly using weights or coefficient values βs to predict an output value y. A key difference from linear regression is that the modeled output value is a binary

value (0 or 1) rather than a numeric value. Below is an example logistic regression equation [HJLS13]:

$$y(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \tag{5.2}$$

Where :

— y is the predicted output,

— β_0 is the bias or intercept term,

— and β_1 is the coefficient for the single input value x.

Therefore, each column in the input data has an associated β coefficient (a real constant parameter) that must be learned from the training data [HJLS13].

The actual model representation used for future inputs is the coefficients in the equation (5.2). This model's β parameters, or coefficient, is commonly estimated via *maximum likelihood estimation* (MLE). This method tries different beta values through multiple iterations to optimize for the best fit of log odds. All of these iterations produce the log-likelihood function, and logistic regression seeks to maximize this function to find the best parameter estimate. Once the optimal coefficients are found, the conditional probabilities for each observation can be estimated, logged, and summed together to generate a predicted probability. For example, for binary classification, a probability less than 0.5 will predict 0, while a probability greater than 0 will predict 1. After the model has been computed, it is best practice to evaluate how well the model predicts the dependent variable, called goodness of fit [HJLS13].

Logistic regression is a discriminative model which attempts to distinguish between classes (or categories). As mentioned earlier, logistic regression maximizes the log-likelihood function to estimate the beta coefficients of the model. Within the context of machine learning, logistic regression uses the negative log-likelihood as the loss function,

using a *gradient descent* approach to find the global maximum [SV21].

Logistic regression is considered prone to overfitting, mainly when there are many predictor variables within the model. Therefore, regularization penalizes parameter-significant coefficients when the model possesses high dimensionality [SV21]. Regularization is a term used in machine learning refering to modifications made to a learning algorithm intended to reduce its generalization error but not its training error. It is used to train models that generalize better on new data by preventing the algorithm from overfitting the training dataset.

5.3 Support Vector Machine (SVM)

Support Vector Machines (SVMs) were initially designed to solve binary classification problems and later extended and applied to regression and unsupervised learning [Nob06]. As a result, they have shown success in solving many complex machine-learning classification problems [Nob06]. SVM is used for classification and regression applications by *finding a hyperplane in an N-dimensional space that distinctly classifies the data points* with maximum marginal distances that contribute to more confidence in classifying new data points [Nob06].

The SVM assumes a linear decision boundary between two different classes and aims to find a hyperplane that delivers the maximum separation between the two categories as shown in Figure 5.3. For this reason, the alternate term maximum margin classifier occasionally refers to an SVM. The perpendicular distance between the nearest data point and the decision boundary is referred to as the margin. When the margin completely separates the positive and negative samples and does not accept any errors, it is called the hard margin [PS20].

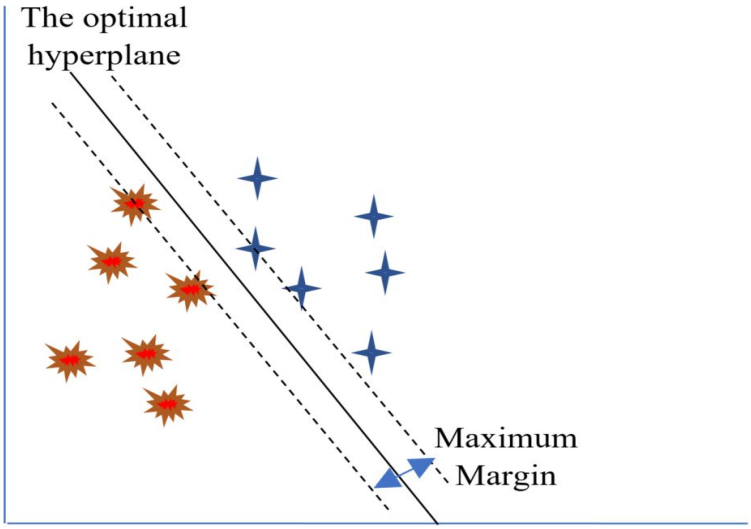

Figure 5.3 – SVM Example

Suppose a training dataset is available (with labeled data) $\{(x_1, y_1), \ldots, (x_n, y_n)\}$,

$$\text{where } y_i = \begin{cases} 1 & \text{if belonging to class 1} \\ -1 & \text{if belonging to class 2} \end{cases}$$

The mathematical expression for a hyperplane is shown below. w_j's are the coefficients, and w_0 is the arbitrary constant that determines the distance of the hyperplane from the origin [PS20].

$$w^T x_i + w_0 = 0 \tag{5.3}$$

For the ith 2-dimensional point (x_{i1}, x_{i2}), the above expression is reduced, as shown below.

$$w_1 x_{i1} + w_2 x_{i2} + w_0 = 0 \tag{5.4}$$

While looking to maximize the margin between positive and negative samples, the positive data points are required to satisfy the following constraint [PS20]:

$$w^T x_i^+ + w_0 \geq +1 \tag{5.5}$$

Similarly, the negative data points are required to satisfy the below constraint:

$$w^T x_i^- + w_0 \leq -1 \tag{5.6}$$

The below uniform equation can be written for both the positive and negative data points by using t_i being either -1 or $+1$ to predict the class label of data point x_i [PS20].

$$t_i(w^T x_i + w_0) \geq +1 \tag{5.7}$$

The below equation is for the perpendicular distance d_i of a data point x_i from the margin.

$$d_i = \frac{|w^T x_i + w_0|}{||w||} \tag{5.8}$$

To maximize this distance, we can minimize the square of the denominator to give us a quadratic equation shown below.

$$min \frac{1}{2}||w||^2 \tag{5.9}$$

which is subject to the below condition.

$$t_i(w^T x_i + w_0) \geq +1, \forall i \tag{5.10}$$

To maximize this distance, the square of the denominator needs to be minimized to give us a quadratic programming problem given by [PS20]:

$$L(w, w_0, \alpha) = \frac{1}{2}||w||^2 + \sum_i \alpha_i(t_i(w^T x_i + w_0) - 1) \tag{5.11}$$

To solve the above, the following need to be set:

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial w_0} = 0 \tag{5.12}$$

Plugging above in the Lagrange function gives us the following optimization problem:

$$L_d = -\frac{1}{2}\sum_i \sum_k \alpha_1 \alpha_k t_i t_k (x_i)^T (x_k) + \sum_i \alpha_i \tag{5.13}$$

The above needs to be maximized with subject to the following [PS20]:

$$w = \sum_i \alpha_i t_i x_i \tag{5.14}$$

and,

$$0 = \sum_i \alpha_i t_i \tag{5.15}$$

The classification of any test point x can be determined using the below expression:

$$y(x) = \sum_i \alpha_i t_i x^T x_i + w_0 \tag{5.16}$$

where a positive value of y(x) implies that x belongs to class +1 and a negative value of y(x) means that x belongs to class -1.

For w_0 , any support vector x_s can be selected and solve for the below [PS20]:

$$t_s y(x_s) = 1 \tag{5.17}$$

giving us the below equation:

$$t_s(\sum_i \alpha_i t_i x_s^T + w_0) = 1 \tag{5.18}$$

Non-linear SVM is used for linearly inseparable data. It is identical to the above algorithm, but with every dot product substituted with a nonlinear kernel function to better fit the maximum margin hyperplane in a transformed feature space [PS20].

5.4 K-Nearest Neighbors (KNN)

KNN algorithm is a pattern recognition, non-parametric method used for both classification and regression [Pet09]. The KNN algorithm operates by storing the previous known cases and classifies new instances based on a similarity measure of distance functions (such as Euclidean, Manhattan, and Minkowski) [Pet09]. After obtaining the K nearest neighbors, a *simple majority of these KNNs are selected in the prediction* of the new instance [Pet09].

The example below in Figure 5.4 demonstrates how KNN algorithm works as we change K. The point in question is the green point.

— Based on 1-nearest neighbor, the data point in question is classifies as red.

— Whereas, based on 2-nearest neighbor, KNN will not be able to classify the point since the second nearest point is blue. As for setting K to 5 leads to classifying the point in question to red as the number of votes for the red are 3 and the number of votes for blue are 2.

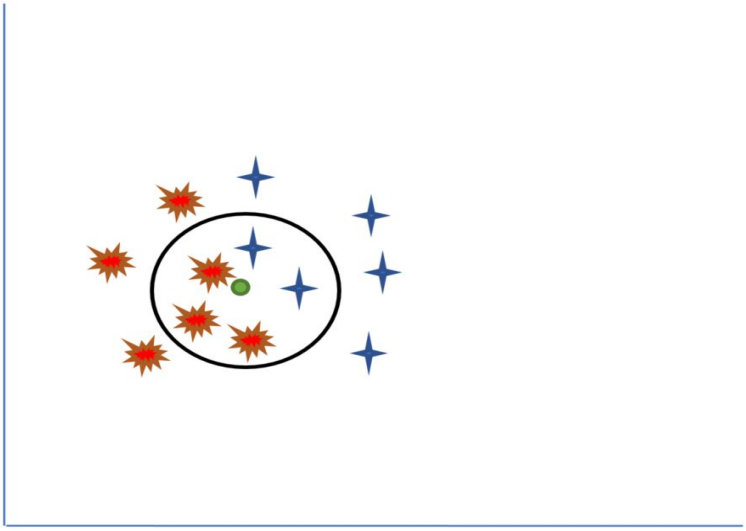

Figure 5.4 – A KNN Example

KNN models are considered straightforward to implement and handle non-linearities well. Fitting the model also tends to be fast as the computer does not have to calculate any particular parameters or values. The trade-off here is that while the model is fast and straightforward to set up, it is slower to predict since, to predict an outcome for a new input data, it will have to search through all the data points in its training dataset every time to find the nearest ones. As a result, KNN can be a relatively slow method for large datasets compared to other algorithms that may take longer to fit but will make their predictions with relatively simple computations [Pet09].

5.5 Decision Trees

Decision trees are used in machine learning for both classification and regression applications [KS08]. They have a structure that resembles a *flow chart mechanism*,

where each internal node represents an assessment of a feature, each branch denotes the result of the assessment, and each terminal node outputs the label [KS08].

Constructing a decision tree learning algorithm works from top to bottom in the sense of selecting a feature that would best split to the rest of the features as shown in Figure 5.5. The decision of how the architecture of the tree should be is achieved by the Gini impurity or using information gain[KS08].

— The Gini impurity would be an assessment of the likelihood of the wrong classification of a new instance of a random variable if that new instance was randomly classified according to the distribution of class labels from the data set. To calculate the Gini impurity for a set of data points with J classes, and p_i are the items that belong to class i [KS08].

$$I_G(p) = \sum_{i=1}^{J}(p_i \sum_{k \neq i} p_k) = \sum_{i=1}^{J} p_i(1-p_i) = \sum_{i=1}^{J}(p_i - p_i^2) = \sum_{i=1}^{J} p_i - \sum_{i=1}^{J} p_i^2 = 1 - \sum_{i=1}^{J} p_i^2 \tag{5.19}$$

Examples :

— If there is only item that belong to one class $\{50, 0, 0\}$: G=0,

— if there is the same number $\{50, 50, 50\}$: G=2/3,

— G will decrease when the elements of one class increase $\{70, 30, 50\}$: G=0.63.

— Information gain is a measure of how much "information" a certain feature can provide about a class. The decision trees algorithm tries to maximize information gain making the attribute with the highest information gain be split first [KS08].

$$I_G(T, a) = H(T) - H(T|a) = -\sum_{i=1}^{J} p_i log_2 p_i - \sum_{a} p(a) \sum_{i=1}^{J} -Pr(i|a) log_2 Pr(i|a) \tag{5.20}$$

Where, *H(T)* is the Entropy of parent and $H(\mathrm{T}|\mathrm{a})$ is the weighted sum of Entropy of children.

Entropy measures the discriminatory power of a feature for the classification task. It defines the amount of randomness in the attribute for a classification task. Information Gain is used for ranking features or attributes for filtering at a given node in a decision tree. The ranking is done based on high information gain entropy in decreasing order.

Some of the advantages of using decision trees for classification problems are that they can capture nonlinear relationships and do not require any transformation of the features. They are also considered fast and efficient compared to KNN and other classification algorithms. In addition, they are easy to understand, interpret, and visualize. Moreover, it provides information about the relative importance of attributes [KS08]. On the other hand, decision trees take much time to train, especially with large datasets and many features. In addition, overfitting is the main problem of the Decision Trees. In order to fit the data (even noisy data), it generates new nodes, and ultimately the tree becomes too complex to interpret. As a result, it loses its generalization capabilities. As a result, it performs well on the trained data but makes many mistakes on future input data [KS08].

5.6 Neural Networks

5.6.1 Artificial Neural Networks(ANN)

Artificial Neural Networks (ANN) is a branch of Artificial intelligence that has been accepted as a new technology in computer science [GVRP10]. It is a mathematical model that *tries to simulate the structure and functionalities of biological neural networks.*

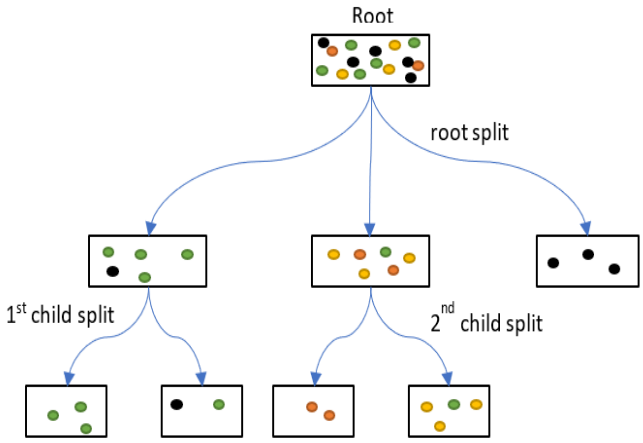

Figure 5.5 – Decision Trees Example

ANNs design and functionalities are derived from observing a biological neuron that is the basic building block of biological neural networks, including the brain, spinal cord, and peripheral ganglia [KBK11].

Artificial Neural Networks (ANN) have been widely employed in a large variety of applications in various fields in the last two decades owing to their remarkable capabilities, such as; pattern recognition in complex biological data sets that cannot be detected with conventional linear statistical analysis [AAIO18], modeling non-linear systems and making a generalized conclusion from data similar to the human brain reaction. Some examples of applications that use ANNs are in pharmaceutical applications [AKB00], diagnostic medical applications [Suz11], skin diseases, sclerosis, anesthesia, and cardiovascular; in clinical and other applications such as remote telemedicine [Hua11] and medical image processing [AAIO18]. Furthermore, technology has rapidly evolved in recent times [GVRP10]. This development has led to a new approach to solving many data processing-based problems [KAM13].

An Artificial Neural Network is a group of interconnected nodes, similar to the network of neurons in an animal brain. ANNs are characterized in principle by network

topology, a connection pattern, neural activation properties, and a training strategy to process data [AD11]. Every Artificial Neural Network has a fundamental building block called artificial neuron organized in layers, which is a simple mathematical model composed of three simple sets of rules, namely; multiplication in which every individual input is multiplied with individual weight at the entrance, summation of all weighted inputs and bias (bias in machine learning refers to the erroneous assumption that some aspects of a dataset are given more weight and/or representation than others.) in the middle section, and lastly, activation, where the result obtained from the middle section is taken through a process called transfer or activation function as illustrated in the Figure 5.6 below.

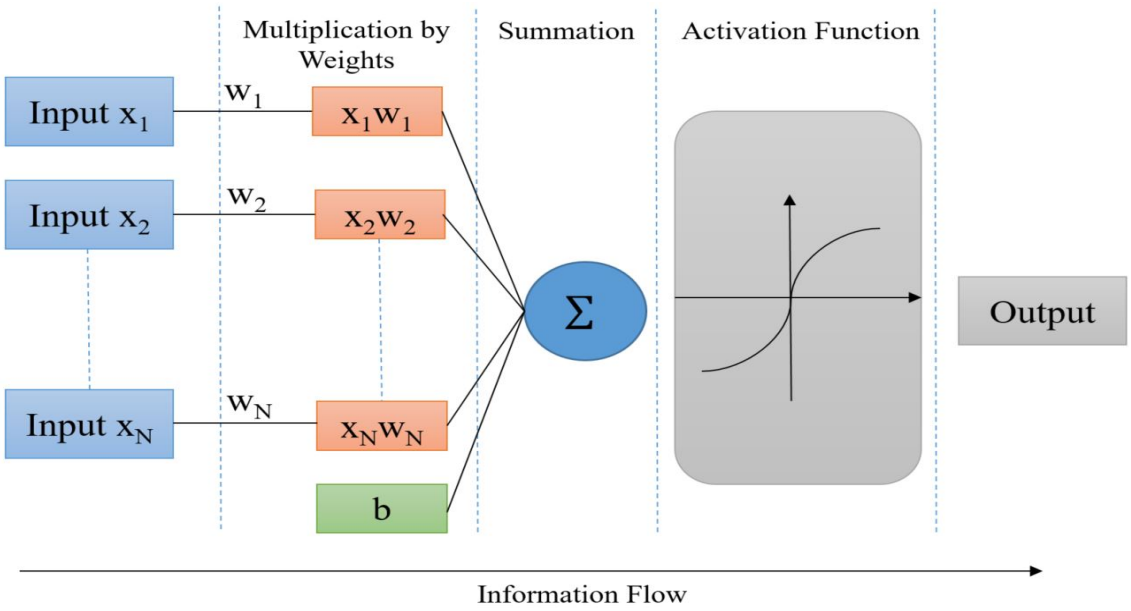

Figure 5.6 – Principle of An ANN

A biological neuron is composed of branched structures known as dendrites; a processing unit called the cell body and a long axon through which the impulses are transmitted as an output to the peripheral [EA13]. Thus, the dendrites receive impulses from neighboring dendrites, transmit the impulses to the processing unit and pass the processed impulses via the long axon to give output [EA13] as shown in Figure 5.7.

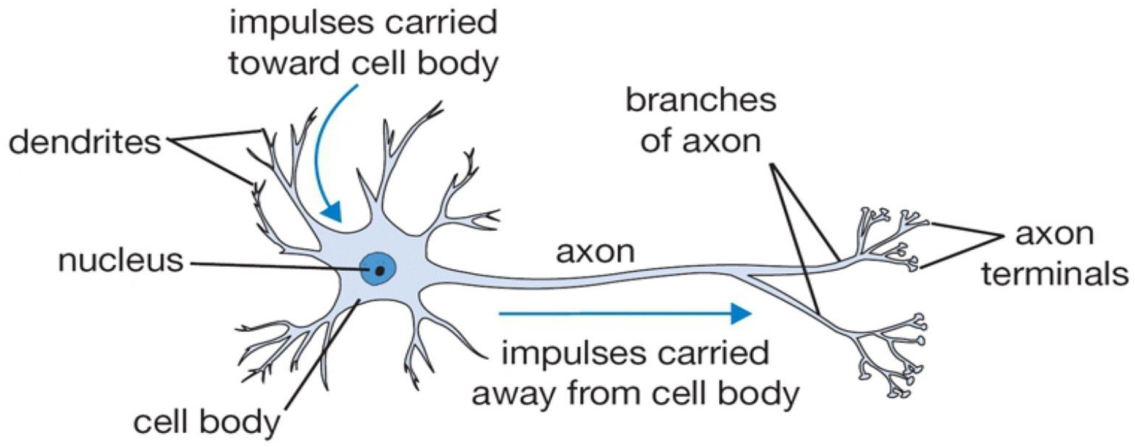

Figure 5.7 – Model of a Biological Neuron [EA13]

Similarly, an artificial neuron receives information via weighted inputs, then transmitted to the body, where the weighted inputs and bias are summed up. The information is then processed with the transfer function, and the output is obtained, as shown in Figure 5.6. The model described can be mathematically depicted using the below equation:

$$y(t) = F(b + \sum_{i=1}^{N} w_i(t)x_i(t)) \tag{5.21}$$

where *y(t)* is the output in discrete time t, F is the activation function, w_i is the weight of the input in discrete time t, *x(t)* is the input value in discrete time t, and b is the bias. The activation function f decides whether a neuron should be activated or not. In other words, an activation function decides whether the neuron's input to the network is essential or not in the prediction process using straightforward mathematical operations. The primary role of the activation function is to convert the summed weighted input from the node into an output value to be fed to the subsequent hidden layer or as the final output. There are many activation functions to choose from to build an ANN, such as binary step function, linear activation function, Sigmoid activation function, Tanh function (Hyperbolic Tangent), ReLU Function, and others.

Neural Networks are modeled as collections of neurons connected in an acyclic graph such that the outputs of some neurons can become inputs to other neurons. Neural Network models are often organized into distinct layers of neurons. The most common layer type for regular neural networks is the fully-connected layer in which neurons between two adjacent layers are fully pairwise connected, but neurons within a single layer share no connections.

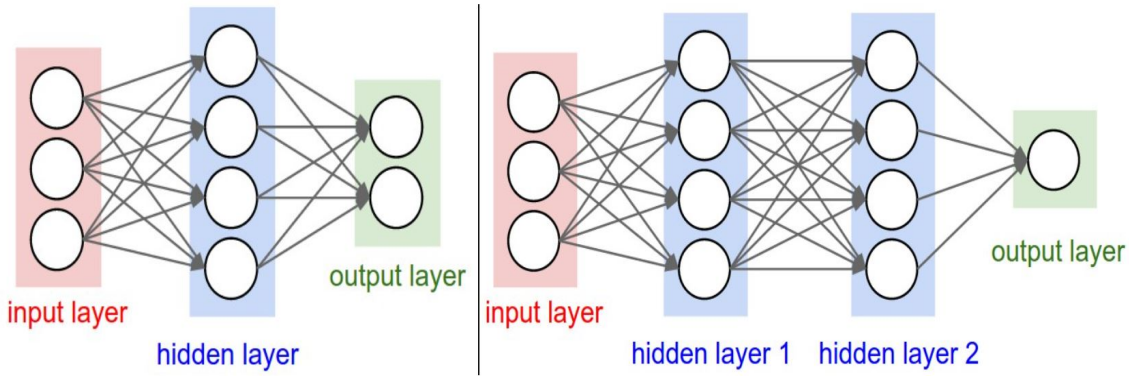

Figure 5.8 – Left: A 2-layer Neural Network (one hidden layer of 4 neurons and one output layer with 2 neurons), and three inputs. Right: A 3-layer neural network with three inputs, two hidden layers of 4 neurons each and one output layer.

Artificial Neural Networks possess the ability to solve complex real-life problems by processing information in their basic building blocks (artificial neurons) in a non-linear, distributed, parallel, and local way. How individual artificial neurons are interconnected is called topology, architecture, or graph of an artificial neural network [KBK11]. Two examples of Neural Network topologies are presented in Figure 5.8 that use a stack of fully-connected layers.

Scientists have come up with a variety of ANN topologies that are suitable for various kinds of applications selected based on the requirements of problem at hand [AAIO18]. For example, in Feed-Forward Neural Network FNN, the flow of information is in one direction from input nodes through the hidden nodes and then to output nodes without back loops [Zel94]. On the other hand, Recurrent Neural Networks RNN permits the

backflow of information with back loops. As such, they are multidirectional and can use their internal state (memory) to process sequences of inputs. These properties make them suitable for tasks such as unsegmented, connected handwriting recognition [GLF+08] or speech recognition [SSB14, LW15]. In addition, the RNNs have been modified for better suitability in exceptional cases of information backflow, such as; Hopfield, Elman, Jordan, bi-directional and other networks[SCD13].

A selected ANN topology is taken through a learning process in which the ANN is taught how to perform the required task effectively [Sim99]. Learning is done by updating the weights and bias levels at the network's input layer as it is simulated in a given data environment, and the actual result obtained is compared with the expected result [Sim99].

Supervised, unsupervised, and reinforcement learning are the various ANN learning paradigms available [NJT01] whose selection is dependent on the kind of problem.

— *Supervised learning* is also called learning with a teacher [Sim99]. It involves modification of the synaptic weights of a neural network by applying a set of labeled training samples or task examples (input-output pairs) and learning a function that maps from input to output [Sim99]. It is used for pattern recognition (classification), Sequential data (handwriting, speech, and gesture recognition), and Regression (Function approximation).

— In *unsupervised learning*, also called learning without a teacher, [Sim99] the network learns patterns in the input even when no explicit feedback or label is supplied. Once the network is tuned to the statistical regularities of the input data, the network develops the ability to form internal representations for encoding features of the input and, thereby, to create new classes automatically [KRSS01]. It is used for filtering, compression, and commonly for clustering tasks.

— In *reinforcement learning*, the learning of an input-output mapping is executed through continued interaction with the environment in order to minimize a scalar index of performance [JLDN19].

5.6.2 Deep Learning

The concept of deep learning is not novel. However, its hype has recently increased, and deep learning is getting more attention. Deep Learning is a subfield of machine learning motivated by *artificial neural networks* [NNFM14].

Deep Learning achieves exceptional power and flexibility by learning to describe the world as a nested hierarchy of concepts or abstractions. It simply takes in data connections between all artificial neurons and modifies them according to the data pattern. More neurons are needed with larger datasets. It automatically select features at multiple abstraction levels, allowing a system to learn complex functions mapping without depending on any specific algorithm. Now that the world has reached a technological level with fast enough computers and enough data to train large neural networks, deep learning is possible to achieve [NNFM14].

One particular benefit of using deep learning is related to scalability [NNFM14]. As we construct more extensive neural networks and train them with more and more data, their performance continues to increase. This is generally not possible from other machine learning techniques that reach a plateau in terms of performance as shown in Figure 5.9.

In addition to scalability, another often cited advantage of deep learning models is their capacity to perform automatic feature extraction from raw data, also called feature learning. Deep learning algorithms aim to manipulate the unknown structure in the input distribution in order to discover reasonable representations, often at multiple

levels, with higher-level learned features described in terms of lower-level features [Ben12].

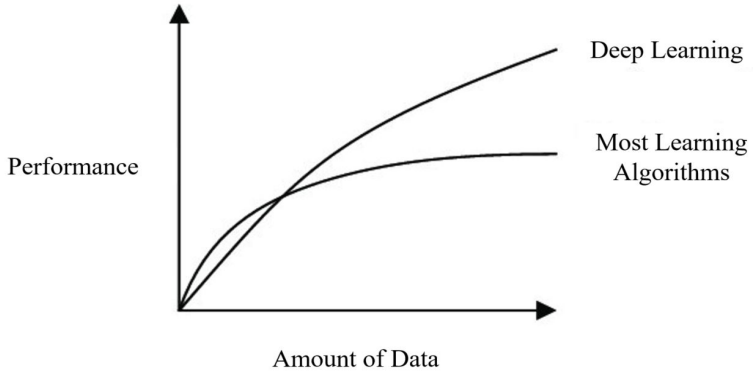

Figure 5.9 – A Comparison between the Performance of Deep Learning against that of Other Machine Learning Algorithms

5.6.3 Convolutional Neural Networks (CNN)

Convolutional layers are the primary building blocks used in convolutional neural networks [AMAZ17]. A *convolution* is a simple filter applied to an input that results in an *activation*. Repeated application of the same filter to an input results in a map of activations called a *feature*, indicating the locations and strength of a detected feature in an input, such as an image [AMAZ17].

The innovation of convolutional neural networks is the ability to automatically learn many filters in parallel specific to a training dataset under the constraints of a specific predictive modeling problem, such as image classification. The result is particular features detected anywhere on input images [AMAZ17].

5.6.3.1 Convolution in CNN

Convolutional Neural Network is a unique neural network model under deep learning as shown in the heirarchy diagram in Figure 5.10. CNN is designed for working with

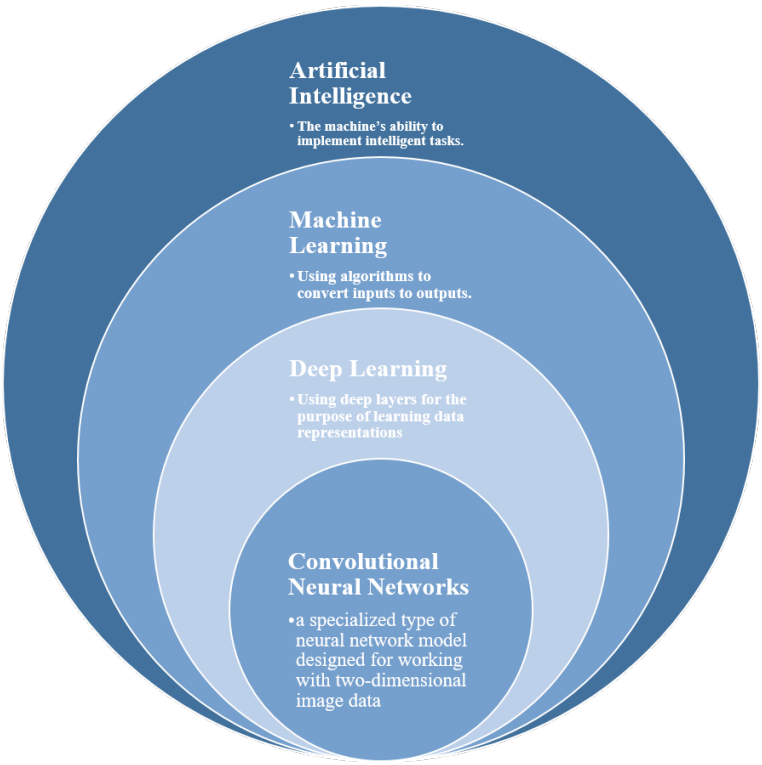

Figure 5.10 – Convolutional Neural Networks Belonging in Deep Learning Methods

two-dimensional input image data. However, it can be used with one-dimensional and three-dimensional data [LLY+21]. The convolutional layer in the network executes an operation referred to as a convolution. Under the context of a convolutional neural network, convolution is a linear operation that involves the multiplication of a set of weights with the input, the same as a traditional neural network. However, since the technique was designed for two-dimensional input, multiplication is applied between an array of input data and a two-dimensional array of weights, called a kernel or a filter [LLY+21]. The filter is much smaller than the input data, and the type of multiplication applied between a filter-sized patch of the input, and the filter is called dot product. A dot product is an element wise multiplication between the filter sized patch of the input and filter, which is then summed, always resulting in a single value. Because it results in a single value, the operation is usually named the "scalar product" [LLY+21].

Employing a filter smaller than the input is intended to allow the same filter to be

multiplied by the input array multiple times at different points on the input. Mainly, the filter is applied systematically to each overlapping part or filter-sized patch of the input data, left to right, top to bottom. This systematic application of the same filter across the entire image is a vital concept. For example, if the filter is developed to detect a precise type of feature in the input, then the application of that filter systematically across the input image allows the filter to identify that feature in the image. This ability is commonly referred to as translation invariance, where the general interest is whether the feature is present instead of where it was [LLY+21].

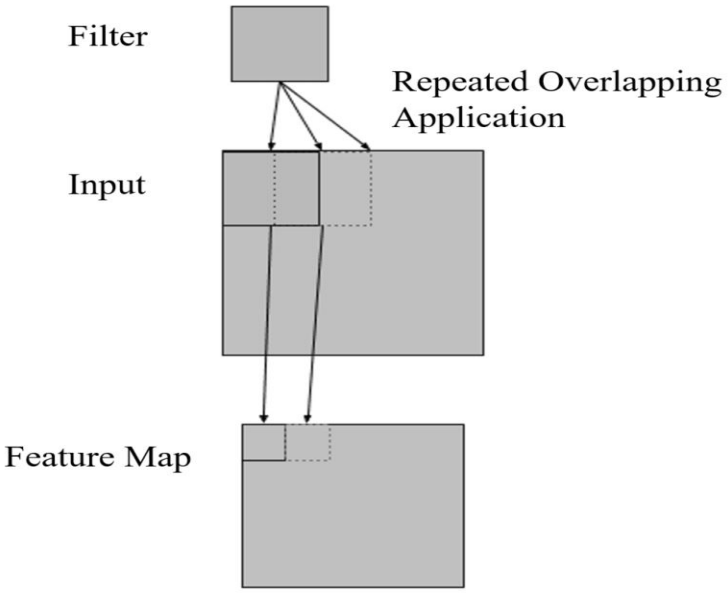

Figure 5.11 – How the Feature Map is Created

The output resulting from multiplying the filter with the input array one time is a single value. However, as the filter is applied multiple times to the input array, the result is a two-dimensional array of output values representing the filtered input called a feature map [Mur16]. To summarize the concept of convolution in convolutional neural networks, with an input image of pixel values and a filter (set of weights), the filter is systematically applied to the input image to create a feature map as shown in Figure 5.12.

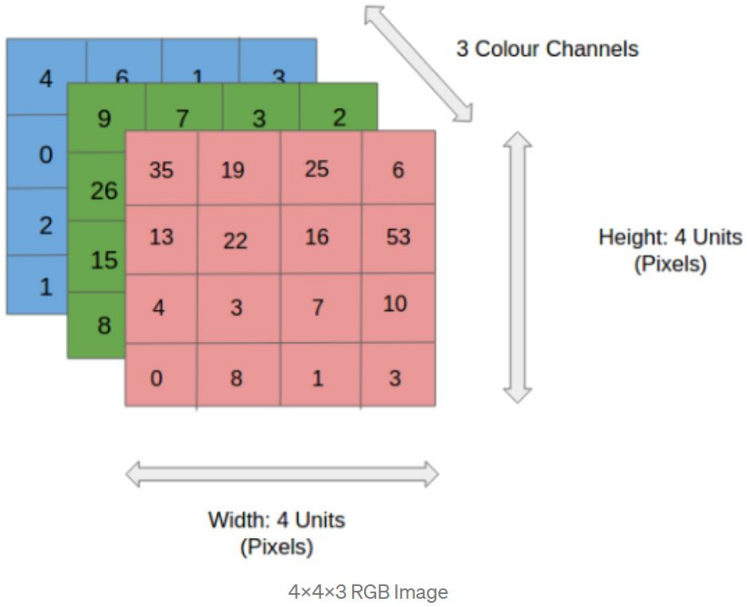

Figure 5.12 – A 4x4x3 RGB Image [Sah18]

In Figure 5.12, an RGB image is separated by its three color planes: Red, Green, and Blue. There are several color spaces in which images exist, such as Grayscale and RGB. The role of the CNN is to reduce the images into a form that is easier to process without losing features essential for obtaining a good prediction. Reducing image complexity is important when designing an architecture that is good at learning features and scalable to massive data-sets .

5.6.3.2 Setting Up a Convolutional Neural Network Architecture

Many decisions need to be made to build a CNN. These decisions will ultimately affect the performance of the model. For example, there are different types of layers, such as the convolutional layer, pooling layer, fully connected layer, softmax layer, and dropout layer. In addition, it's pretty common to have multiple layers of the same type. Further, most of the different types of layers can be customized, and you'll usually have to set the number of input and output nodes and other parameters. This section explains how to select an appropriate number for the different types of layers and set

reasonable parameter values.

— **Convolutional layers** perform convolutions, which, as explained earlier, are operations where a filter is moved over an input image, calculating the values in a resulting feature map. A convolutional layer is usually built of multiple filters, producing multiple feature maps. During the training of the CNN, the model will learn the weights to apply to the different feature maps and, hence, be able to identify which features to extract from the input images. An example is shown in Figure 5.13 [Mur16] where the green part represents a 5x5x1 input image. The element involved in the convolution operation in the first part of a Convolutional Layer is the Kernel/Filter, K, depicted using yellow color. In this example, K is a 3x3x1 matrix. The Kernel shifts 9 times since the stride length is 1. With every shift, a matrix multiplication operation occurs between K and the portion of the image over which the Kernel is hovering. The filter moves toward the right direction with a particular stride value till it parses the complete width [Sah18]. Then, moving on, it goes down to the far left side of the image with the same stride value and repeats this process until the entire image is crossed as shown in the below Figure 5.14 [Sah18].

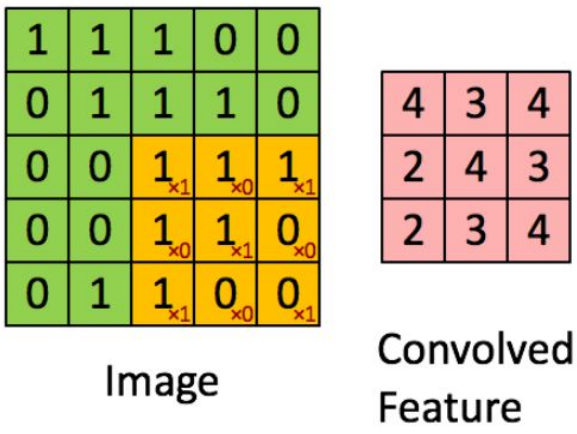

Figure 5.13 – Convolution of a 4x4x3 RGB Image [Sah18]

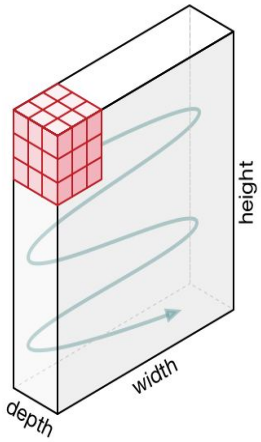

Figure 5.14 – A 4x4x3 RGB Image Convolution Filter Movement[Sah18]

In images with multiple channels, such as RGB, the Kernel has the same depth as the input image. Then, matrix multiplication is performed between Kn and the image stack, and all the results are summed with the bias to give us a squashed one-depth channel Convoluted Feature Output as shown in Figure 5.15 [Sah18]. The convolution operation's purpose is to extract high-level features, such as edges, from the input image. Therefore, CNNs should not be limited to only one convolutional layer. Conventionally, the first convolutional layer is responsible for capturing the low-level features such as edges, color, and gradient orientation. Then, with additional layers, the architecture adapts to the high-level features, giving us a network that has a generalized understanding of images in the dataset, similar to how we would [Sah18].

There are two kinds of results to the convolution operation. The first is Valid Padding, where the convolved feature is reduced in dimensionality compared to the input. The other result is the Same Padding in which the dimensionality is either increased or remains the same. An example of the Same Padding (Figure 5.16) is when a 5x5x1 image is augmented into a 6x6x1 image, and then the 3x3x1 kernel is applied over it; the convolved matrix turns out to be of dimensions 5x5x1. On the other hand, if the same operation is performed without padding,

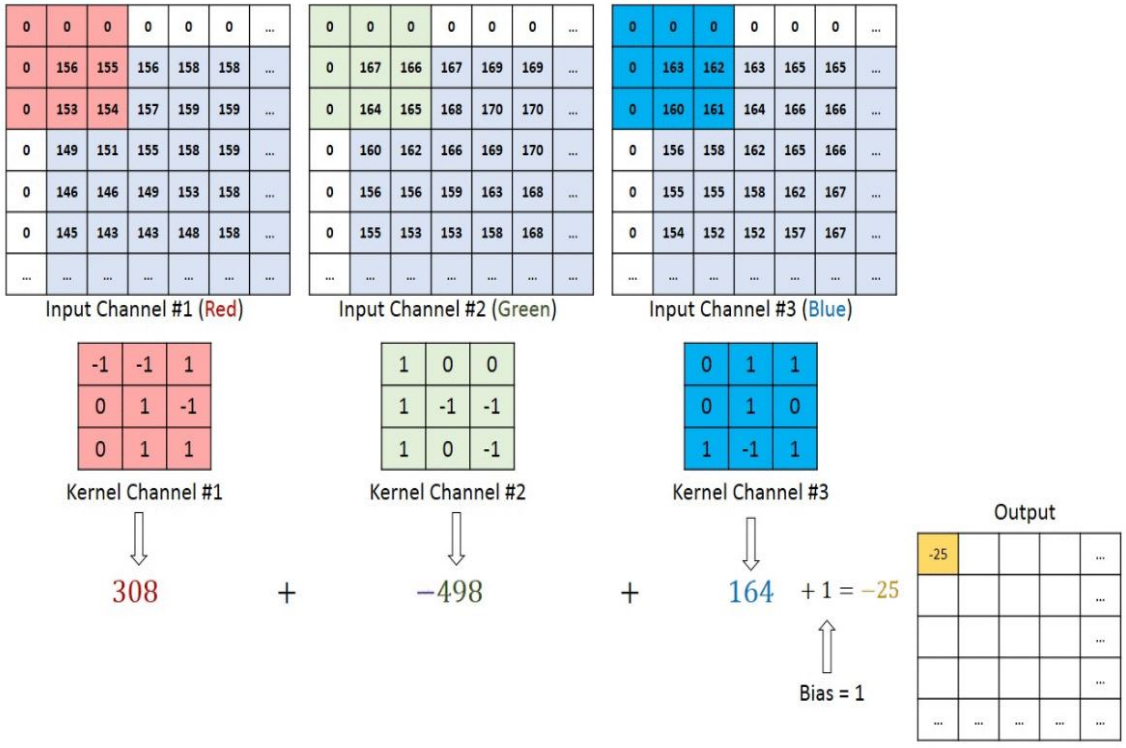

Figure 5.15 – A Convolution Operation on an MxNx3 Image Matrix with a 3x3x3 Kernel [Sah18]

the result is a matrix with dimensions of the kernel (3x3x1) itself. This is an example of Valid Padding [Sah18].

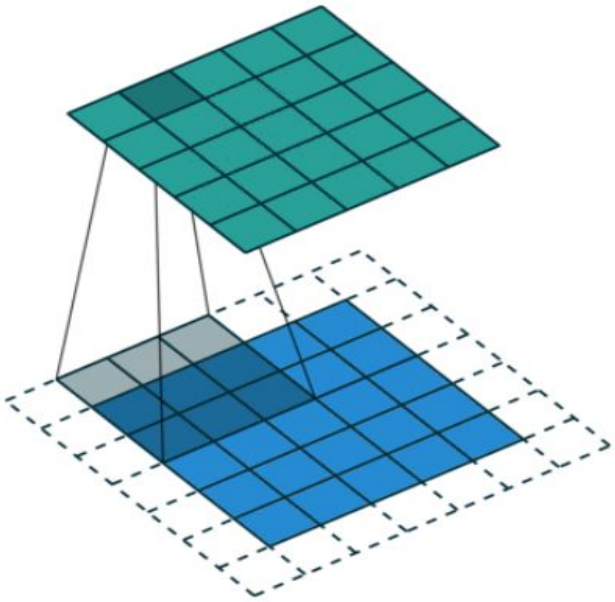

Figure 5.16 – An Example of Same Padding: 5x5x1 Image is Padded with 0s to Create a 6x6x1 Image Convolution Operation on an MxNx3 Image Matrix with a 3x3x3 Kernel [Sah18]

By increasing the number of convolutional layers in the CNN, the model can detect more complex features in an image. However, more layers will take more time to train the model and increase the likelihood of overfitting. So, two convolutional layers will usually be enough while setting up a relatively simple classification task. And then, the number of layers can be increased if the resulting accuracy is considered unsatisfactory.

The suitable number of nodes is also highly dependent on the complexity of the images and the task at hand. By altering the number of nodes and evaluating the resulting accuracy, the model can be run multiple times until an acceptable performance is achieved.

— **Pooling Layers** are layers that help reduce the computational cost of the model and aim to minimize overfitting by decreasing the dimensionality of its input.

There are two different types of pooling layers, listed below and shown also in Figure 5.17:

— Max pooling, which chooses the largest value in the matrix;

— Average pooling, which chooses the mean of the values in the matrix.

Max Pooling performs as a noise suppressant. It discards noisy activations and performs de-noising and dimensionality reduction. *Max pooling* helps extract prominent features such as edges. On the other hand, average pooling adds a small amount of translation invariance, which translates the image by a small amount that does not significantly affect the values of most pooled outputs. In addition, it extracts features more smoothly than *Max Pooling*. Therefore, Max Pooling is thought to perform much better than Average Pooling and is used more often in CNNs especially when edge detection is important in the classification task.

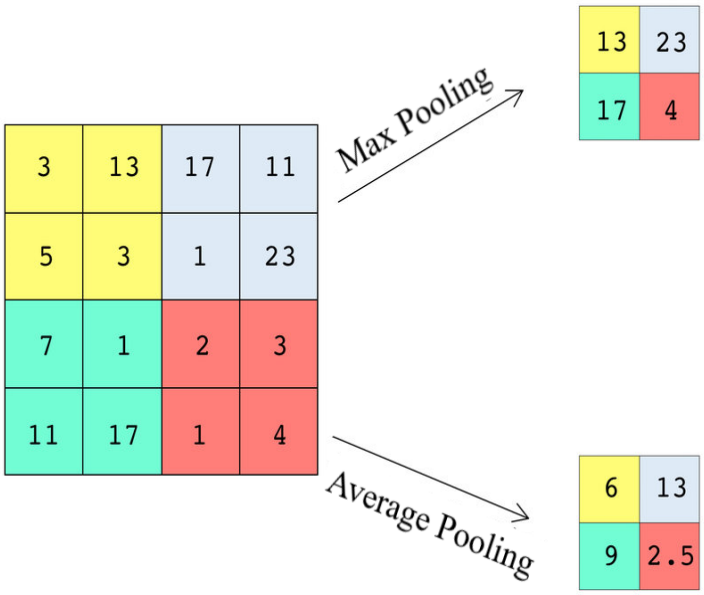

Figure 5.17 – The Two Types of Pooling Layers [Alj18].

In a pooling layer, a filter is passed to the entire image. The window size and stride will determine the output size and how the filter is moved over the input

matrix. The most standard is to choose a window size of 2x2. There is no straightforward answer for how often to include a pooling layer in a CNN, and iterations should be executed until an adequate performance is achieved. For example, the well-known computer vision model VGG-16 operates using two to three convolutional layers between the pooling layers, while VGG-19 operates using up to four [AEV20],

The Convolutional Layer and the Pooling Layer jointly constitute a CNN's i-th layer. Depending on the complexities in the images, the number of such layers may be increased for capturing low-level details more, but at the cost of more computational power. Now that the model understands the features, the final output will be flattened and fed to a regular ANN for classification purposes.

— **Fully Connected Layers** transform their inputs to the preferred output format. For example, a classification task normally contains converting a matrix of image features into a 1xC vector where C is the number of classes. There is not necessarily a correct answer to how many fully connected layers should be chosen in a CNN model. For most models, however, starting with one or two fully connected layers would be sufficient, later optimizing the number depending on the resulting performance.

— **Softmax Layers** are commonly used after the fully connected layers. A Softmax Layer takes a vector of size 1xC as an input, where C is the number of classes, and all numbers add up to 1. The softmax layer then utilizes this vector and builds a new vector where each input denotes a probability for the image to be of that particular class. A softmax is, therefore, mainly used in classification tasks. For most computer vision projects, one softmax layer will be sufficient [LWYY16].

— **Dropout Layers** involve turning off nodes randomly with a probability during training as shown in Figure 5.18. Such layers are especially helpful in fighting

overfitting in models with much complexity. Dropout layers can be suitable to apply to fully connected layers and convolutional layers [ZPST18].

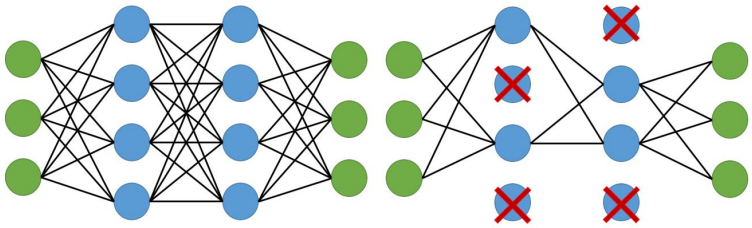

Figure 5.18 – An Example of a Dropout operation with drop rate 0.5 [Alj18]

The correct number of layers and nodes is usually determined by experimentation and trying a different number of layers and nodes. The step is to add the first convolutional layer. Next, is to add the Leaky ReLU activation function, which helps the network learn non-linear decision boundaries.

The ReLU activation function is utilized extensively in neural network architectures, especially in CNNs, where it has proven to be more effective than the widely used logistic sigmoid function.

As of 2017, this activation function is the most popular for deep neural networks. The ReLU function allows the activation to be thresholded at zero. However, during the training, ReLU units can "die." This is possible when a large gradient flows through a ReLU neuron: it can cause the weights to update so that the neuron will never activate on any data point again. If this happens, the gradient flowing through the unit will always be zero. Leaky ReLUs attempt to avoid this problem by setting a slight negative slope so the function will not be zero.

Next, the max-pooling layer is added, and so on. The fully connected layer's purpose is to flatten the high-level features learned by convolutional layers and combine all the features. Finally, it passes the flattened output to the output layer, where a softmax classifier or a sigmoid is used to predict the input class label.

The last layer is a Dense layer with a softmax activation function with two units, which is needed for this binary class classification problem.

After the model is created, it is compiled using the Adam optimizer, one of the most popular optimization algorithms. Adam is a machine learning optimization algorithm recently introduced in 2015 to be utilized rather than the classical stochastic gradient descent method to update network weights iteratively based on the training data. Adam is further explained in Chapter 6.

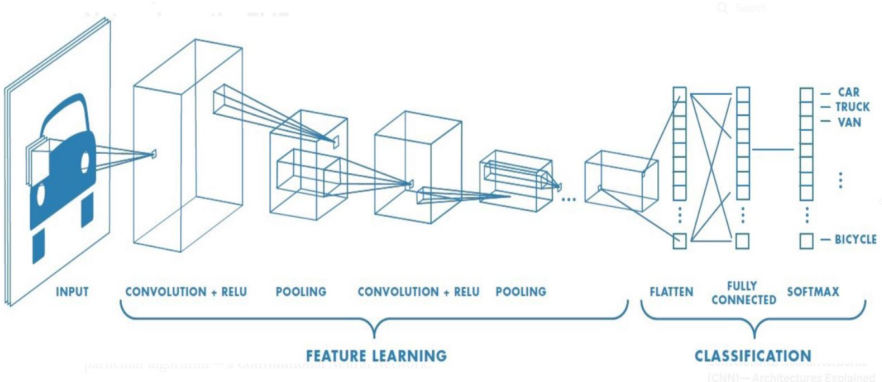

Figure 5.19 – A CNN Example of Classifying Images of Vehicles [Sah18]

Figure 5.19 shows an example of a CNN to classify images into different classes of vehicles. The first part of the CNN is dedicated to feature learning which is then followed by the classification. The layers: Convolutional layer + ReLu and Pooling layers are repeated until satisfactory performance of the model is reached.

Another example is the CNN shown in Figure 5.20 used to classify handwritten numbers. The input is a 28x28 image first fed to a convolutional layer with a 5x5 kernel and then to a 2x2 max pooling layer. Other convolutional and max pooling layers are used before feeding the data to a fully connected layer with ReLU activation. The data is flattened and then entered into another fully connected layer with dropout to finally reach the output.

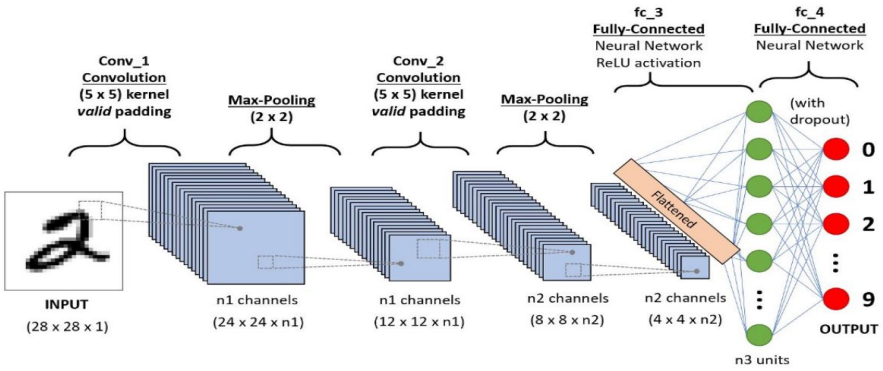

Figure 5.20 – A CNN Example of Classifying Handwritten Digits [Sah18]

5.7 Machine Learning Performance Metrics

Evaluating the performance of a Machine learning model is one of the essential steps while building an effective Machine learning model. Different metrics are used to evaluate the model's performance or quality, known as performance or evaluation metrics. These performance metrics help us understand how well the model has performed for the given data. This way, the model's performance can be improved by tuning the hyperparameters. In addition, each Machine learning model aims to generalize well on new data, and performance metrics help determine how well the model generalizes [JS15].

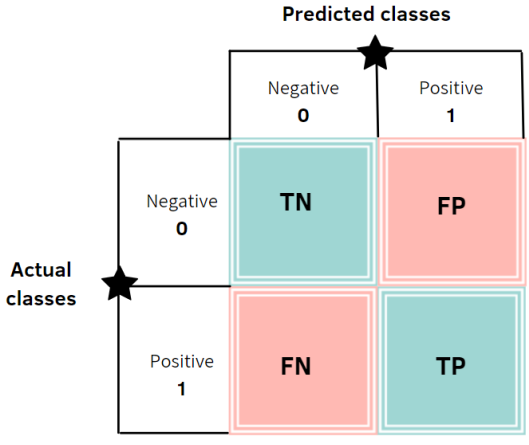

Figure 5.21 – A Confusion Matrix.

5.7.1 The Confusion Matrix

A confusion matrix is a table layout that allows visualization of the performance of a supervised machine learning algorithm. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa (as shown in Figure 5.21).The name stems from that it makes it easy to see whether the system is confusing or is mislabeling two classes [JS15]. It has two dimensions ("actual" and "predicted") and identical sets of "classes" in both dimensions. There are 4 cases in the confusion matrix:

— **True positive (TP)**: The prediction is positive, and the subject has a risk of falling. (correct prediction)

— **True negative (TN)**: Prediction is negative and the subject does not have a risk of falling. (correct prediction)

— **False positive (FP)**: Prediction is positive and the subject does not have a risk of falling. (bad prediction)

— **False negative (FN)**: Prediction is neagtive and the subject has a risk falling.

The below supervised machine learning parameters are extracted from the confusion matrix:

— **Accuracy** is the ratio of the correctly labeled subjects to the whole pool of subjects. Accuracy answers the following question: How many subjects were correctly labeled out of all the subjects?

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \tag{5.22}$$

  — Numerator: all correctly labeled subjects (All trues)

  — Denominator: all the subjects

— **Sensitivity** is the ratio of the correctly positive labeled by the algorithm to

all who are diabetic in reality. Sensitivity answers the following question: Of all the subjects who are with risk of falling, how many of those were correctly predicted?

$$Sensitivity = TP/(TP + FN) \tag{5.23}$$

- Numerator: positive labeled diabetic people.
- Denominator: all people who have the risk of falling (whether detected by our program or not)

— **Specificity** is the correctly negative labeled by the algorithm to all who do not have a risk of falling in reality. Specificity answers the following question: Of all the people who do not have a risk of falling, how many of were correctly predicted?

$$Specificity = TN/(TN + FP) \tag{5.24}$$

- Numerator: negative labeled people with no risk of falling.
- Denominator: all people who do not have a risk of falling in reality (whether positive or negative labeled)

— **Precision** is the ratio of the correctly positive labeled by the algorithm to all positive labeled. Precision answers the following: How many of those who we labeled as fallers are actually fallers?

$$Precision = TP/(TP + FP) \tag{5.25}$$

- Numerator: positive labeled people with risk of falling.
- Denominator: all positive labeled by our program (whether they have a risk of falling or not in reality).

5.8 Conclusion

Early and more accurate disease detection, better diagnosis, and preventable painful treatments could all be made possible by utilizing machine learning to detect future falls in the elderly. After performing feature extraction and selection, the machine learning algorithms used in this thesis were explained in this chapter. Proper optimization of the hyperparameters follows as they play a significant role in optimizing the performance of the classification models.

Chapter 6

Model Optimization Methods

L'optimisation de l'apprentissage automatique consiste à ajuster les hyperparamètres pour minimiser la fonction de coût (écart entre l'étiquette réelle et ce que le modèle a classifié). Ce chapitre explique deux techniques utilisées dans cette thèse pour l'optimisation de l'apprentissage automatique : la validation croisée "Grid Search" et la méthode d'optimisation Adam pour l'apprentissage en profondeur.

Machine learning optimization is adjusting hyperparameters to minimize the cost function. It is essential to minimize the cost function as it describes the discrepancy between the actual label and what the model has classified. This chapter explains two techniques used in this thesis for machine learning optimization: Grid Search Cross-Validation and the Adam Optimization Method for Deep Learning.

6.1 Introduction

This thesis used supervised machine learning to predict the risk of falling in elderly people. Supervised learning is considered similar to learning from mistakes. It uses an iterative approach to find the conditions to achieve the minimum prediction error of a machine learning model for a given input data. The error is usually expressed as the difference between the output predicted by the model and the actual/target or labeled output (included in the training data). This error is calculated using the so-called loss function. The mathematical procedure for minimizing error/loss function is known as the optimization method. Following this general outline, as shown in Figure 6.1, different machine learning algorithms can be implemented using different models, loss functions, and optimization methods.

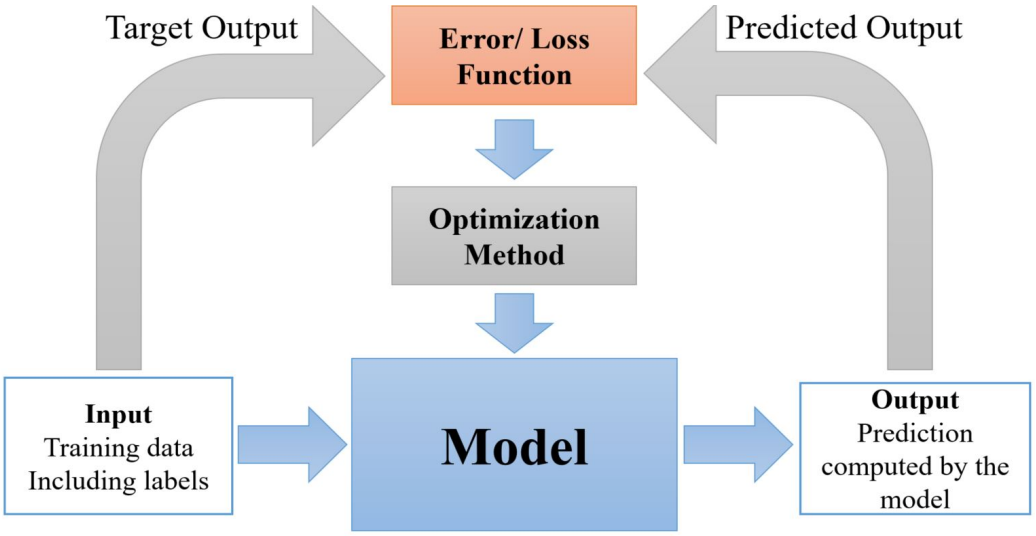

Figure 6.1 – Optimization method based on minimizing the Error/Loss Function

6.1.1 Parameters and Hyperparameters of the Model

It is essential not to be confused between model parameters and hyperparameters (Figure 6.2). The main difference is that hyperparameters must be set before training the model [YS20]. In contrast, the model's parameters are obtained during training. Some examples of model hyperparameters include the learning rate (α), the number of layers, and the number of nodes in each hidden layer of an artificial neural network. The hyperparameters describe the structure of the model [YS20]. The hyperparameters of each of the classification methods used in this thesis are listed in Table 6.1. On the other hand, the model's parameters are obtained while training the model. Some examples of model parameters are the weights and biases for artificial neural networks. These values are internal data to the model and are altered based on the inputs. Therefore, Hyperparameter optimization needs to be performed to tune the model. The error can be decreased by finding the optimal combination of their values [YS20] .

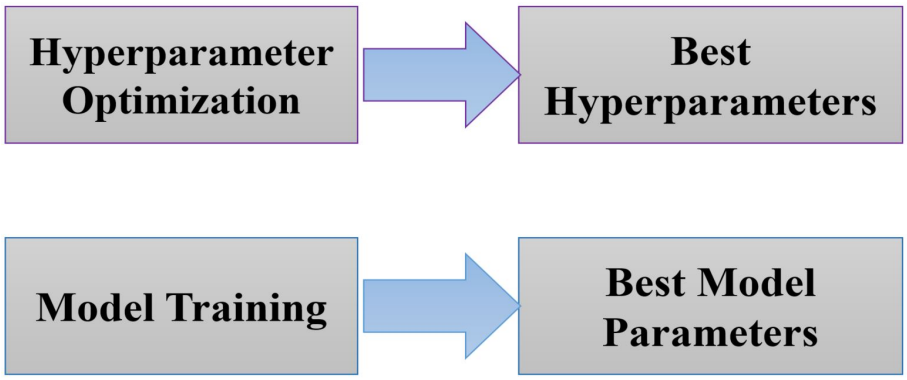

Figure 6.2 – The difference between model parameters and hyperparameters.

6.2 Hyperparameters Optimization Techniques

As mentioned previously, the hyperparameters are set before training. However, it is impossible to know in advance, for instance, which is the best learning rate to use

Table 6.1 – The Hyperparameters for each Classification Algorithm

Method	Hyperparameters
Logistic Regression	Regularization(C) Learning rate(α)
Support Vector Machine (SVM)	Regularization(C) Kernel Coefficient (gamma) Degree
k-Nearest Neighbors(KNN)	Weights:'uniform' or 'distance' Distance metric:'euclidean' or 'manhattan' Number of neighbors k
Decision Trees	Criterion: 'gini' or 'entropy' Maximum depth Minimum samples at a leaf node
Artificial Neural Networks (ANN)	learning rate (α) Number of neurons in the hidden layer Number of hidden layers the activation function

for specific scenarios or cases in the data-set. Therefore, a starting value needs to be set; then, hyperparameters should be optimized to improve the model's performance. Hyperparameter tuning is done by comparing the output with the expected results, assessing the accuracy, and adjusting the hyperparameters if necessary with every iteration. This repeated process can be done manually or by using optimization techniques, which can be helpful, especially when dealing with large datasets.

6.2.1 Hyperparameters Tuning Using Grid Search Cross-Validation

Grid search is the most straightforward algorithm for hyperparameter tuning. First, the domain of the hyperparameters is divided into a discrete grid, as shown in Figure 6.3. Then, every combination of values of this grid is run, calculating some performance

metrics using cross-validation [SD19].

Grid search is an exhaustive algorithm that spans all the combinat Also, it is essential to note that every point in the grid needs k-fold cross-validation, which requires k training steps. So, tuning the hyperparameters of a model in this way can be complex and expensive. However, in cases where accuracy is essential, such as diagnostic applications in the medical field, using the grid search to find the best combination, time is an acceptable price to pay for finding the best combination of values of the hyperparameters [SD19].

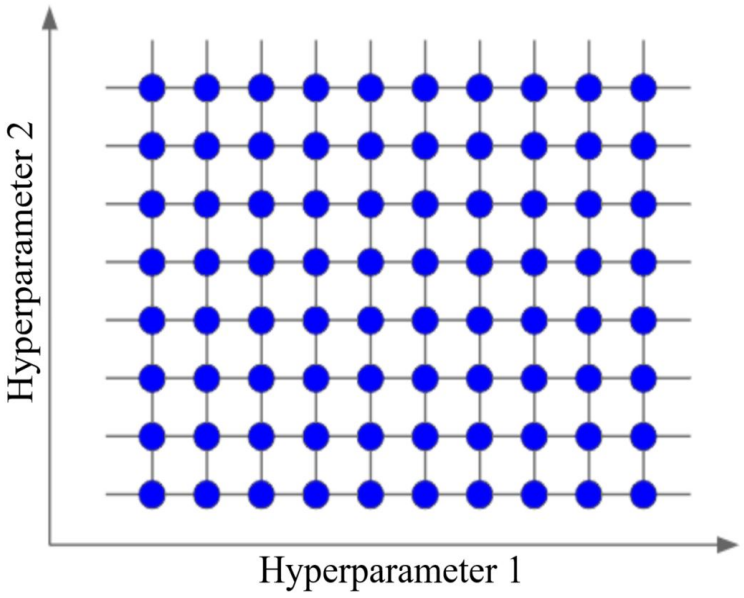

Figure 6.3 – An Example of a Grid Search

6.3 Adam Optimization Algorithm for Deep Learning

The choice of optimization algorithm for deep learning models can significantly impact the model's performance. The Adam optimization algorithm is an extension

to stochastic gradient descent that has recently seen wider adoption for deep learning applications in medical diagnosis, natural language processing, and computer vision [JIN19]. This section addresses the use of the Adam optimization algorithm in deep learning applications, how it works, how it can be configured, the commonly used configuration parameters, and its benefits in optimizing learning models [JIN19].

6.3.1 Gradient Descent Optimization Method

Gradient descent is the most popular optimization method. It is a first-order optimization algorithm that explicitly uses the first-order derivative of the objective target function [Rud16]. The first-order methods rely on gradient information to help direct the search for a minimum. The first-order derivative is the rate of change or slope of the target function at a specific point [Rud16].

If the target function holds multiple input variables, it is called a *multivariate function*, and the input variables can be considered a vector. In turn, the derivative of the multivariate target function can also be considered a vector and is generally called the *gradient* [Rud16].

The *gradient* is the first-order derivative for a multivariate objective function: the derivative or the gradient points toward the steepest ascent of the target function for specific input. *Gradient descent* is a minimization optimization algorithm that follows the negative gradient downhill of the target function to locate the minimum point of the function [Rud16].

The *gradient descent* algorithm demands a target function that is being optimized and the derivative function for the objective function. The target function *f()* gives a score for a given set of inputs, and the derivative function *f'()* returns the derivative of the target function for a given data set [Rud16].

The gradient descent algorithm needs a starting point x in the problem, such as a randomly selected point in a given set of inputs. The derivative is then computed, and a step is taken in the input space intended to lead to a downhill movement in the target function, assuming we minimize the target function [Rud16]. A downhill movement is made by first computing how far to move in the input space, estimated as the step size or *learning rate* (referred to as α) multiplied by the gradient. This is then subtracted from the current point, guaranteeing we move against the gradient or down the target function [Rud16].

$$x(t) = x(t-1) - (\alpha f' x(t-1)) \tag{6.1}$$

The steeper the objective function is at a given point, the larger the magnitude of the gradient and, in return, the larger the step is taken in the search space. The step size carried out is scaled using a step size hyperparameter. The Step Size (α) is a hyperparameter that controls how far to move in the search space against the gradient of each iteration performed by the algorithm [Rud16].

In case the step size is too small, the movement in the search space will be slight, and the search will take longer. On the other hand, if the step size is too large, the search may bounce around the search space and might risk skipping the optima [Rud16].

6.3.2 Adam Optimization Algorithm

Adaptive Movement Estimation algorithm, or Adam for short, is an extension to the gradient descent optimization algorithm. The algorithm was described in [KB14]. Adam is designed to accelerate the optimization process, e.g. decrease the number of function evaluations required to reach the optima, or to improve the capability of the optimization algorithm, e.g. result in a better final result. This is achieved by

calculating a step size for each input parameter that is being optimized. Importantly, each step size is automatically adapted throughput the search process based on the gradients (partial derivatives) encountered for each variable [KB14].

First, we must maintain a moment vector and exponentially weighted infinity norm for each parameter being optimized as part of the search, referred to as m and ν respectively. They are initialized to 0 at the start of the search [KB14].

The algorithm is executed iteratively over time t starting at $t = 1$, and each iteration involves calculating a new set of parameter values x, e.g. going from $x(t-1)$ to $x(t)$.

It is perhaps easy to understand the algorithm if we focus on updating one parameter, which generalizes to updating all parameters via vector operations [KB14].

First, the gradient (partial derivatives) are calculated for the current time step.

$$g(t) = f'(x(t-1)) \tag{6.2}$$

Next, the first moment is updated using the gradient and a hyperparameter β_1.

$$m(t) = \beta_1 m(t-1) + (1-\beta_1)g(t) \tag{6.3}$$

Then the second moment is updated using the squared gradient and a hyperparameter $\beta 2$.

$$\nu(t) = \beta_2 \nu(t-1) + (1-\beta_2)g(t)^2 \tag{6.4}$$

The first and second moments are biased because they are initialized with zero

values. These moving averages are initialized as vectors of null values, leading to moment estimates biased towards null values, especially during the initial time steps and when the decay rates are low (the betas are close to 1). Fortunately, this initialization bias can be easily counteracted, resulting in bias-corrected estimates. [KB14].

Thus, the first and second moments are bias-corrected, starting with the first moement:

$$\hat{m}(t) = m(t)/(1 - \beta_1(t)) \tag{6.5}$$

And then the second moment:

$$\hat{\nu}(t) = \nu(t)/(1 - \beta_2(t)) \tag{6.6}$$

Note that $\beta_1(t)$ and $\beta_2(t)$ refer to the β_1 and β_2 hyperparameters that are decayed on a schedule over the iterations of the algorithm. A static decay schedule can be used, although the paper recommend the following:

$$\beta_1(t) = \beta_1^t \tag{6.7}$$

$$\beta_2(t) = \beta_2^t \tag{6.8}$$

Finally, the moving averages are used to scale the learning rate individually for each parameter. The way it's done in Adam is the following:

$$x(t) = x(t-1) - \alpha\hat{m}(t)/(\sqrt{\hat{\nu}(t)} + \epsilon) \tag{6.9}$$

Where α is the step size hyperparameter, ϵ is a small value such as 1e-8 that ensures we do not encounter a divide by zero error.

A more efficient reordering of the update rule is mentioned in the paper and can be used:

$$\alpha(t) = \alpha\sqrt{1 - \beta_2(t)}/(1 - \beta_1(t)) \tag{6.10}$$

$$x(t) = x(t-1) - \alpha(t)\hat{m}(t)/(\sqrt{\hat{\nu}(t)} + \epsilon) \tag{6.11}$$

It should be noted that there are three hyperparameters for the algorithm, they are:

α: Initial step size (learning rate), a typical value is 0.001.

β_1: Decay factor for first momentum, a typical value is 0.9.

β_2: Decay factor for infinity norm, a typical value is 0.999.

6.4 Conclusion

Hyperparameters are essential because they directly control the training algorithm's behavior and significantly impact the performance of the model being trained. Hyperparameters are the knobs or settings that can be tuned before running a training assignment to control the behavior of a machine learning algorithm. As a result, they can significantly impact model training regarding training time, infrastructure resource requirements, computational cost, model convergence, and model accuracy. The following chapter demonstrates the results obtained throughout this thesis.

Chapter 7

Results and Discussion

La prévention des chutes chez les personnes âgées est considérée comme l'un des sujets de santé publique les plus critiques au vu de la population importante de personnes agées. L'identification du risque de chute chez les personnes âgées est considérée comme la première étape de la prévention. Dans cette thèse, nous avons exploré plusieurs indicateurs pour voir si ils sont indicatifs de futures chutes chez les personnes âgées. Nous nous sommes spécifiquement concentrés sur le degré de cyclostationnarité et sur une méthode alternative de représentation de la cyclostationnarité du signal sous forme d'images de carte thermique. Nous avons également exploré différentes méthodes de sélection de caractéristiques, des algorithmes de classification et des méthodes d'optimisation d'hyperparamètres. Ce chapitre présente les résultats obtenus grâce à cette thèse et une interprétation de ces résultats.

Falls prevention among the elderly community is regarded as one of the most critical public health topics in today's aging society. Identifying the risk of falling in elderly individuals is considered the first step in prevention. In this thesis, we explored several features to study if they are indicative of future falls in the elderly. We specifically focused on the degree of cyclostationary and an alternative method of representing signal cyclostationarity as heat-map images. We also explored different feature selection methods, classification algorithms, and hyperparameter optimization methods. This chapter presents the results obtained through this thesis and an interpretation of these results.

7.1 The Degree of Cyclostationarity as an Important Feature for Predicting Eledrly Falls

As a first step in working with the unbalanced data-set and exploring features relevant to falling risk in the elderly and methods for classification, we dealt with the data-set as balanced, meaning having the same number of fallers(54) and non-fallers(54). The Cyclostationarity properties used in the analysis of walking signals with light shed on the degree of cyclostationarity, using ANOVA for selecting the ten best classical features for prediction of fallers and non-fallers, and the five different machine learning models used for classification. The results of the ANOVA tests and the classification models are presented and discussed.

7.1.1 ANOVA Test for Statistically Significantly Features

Two sets of features are compared in this section of this chapter. The features are extracted from the innersole pressure signals of the 108 subjects divided equally between fallers and non-fallers. The features were extracted separately from the signals of each foot (left and right) The first set includes 10 already studied classical features (10 CF), while the second set includes a single feature, which is the degree of cyclostationarity (DC).

The classical features extracted from the pressure signals were: mean, rise time, fall time, pulse width, overshoot, undershoot, duty cycle, slew rate, midcross, autocorrelation, standard deviation, band power, median, root mean square, range, Pwelch, skewness, interquartile range, and kurtosis. The statistical significance of these features for elderly falling risk were tested using a one-way repeated measures ANOVA [Gad98].

The results showed that *10 features out of those had statistical significant differences* between fallers and non-fallers, in at least 1 type of walking condition. These were: *pulsewidth (right foot), undershoot (right and left feet), duty cycle (left foot), slew rate (right and left feet), range (right and left feet), and skewness (right and left feet).*

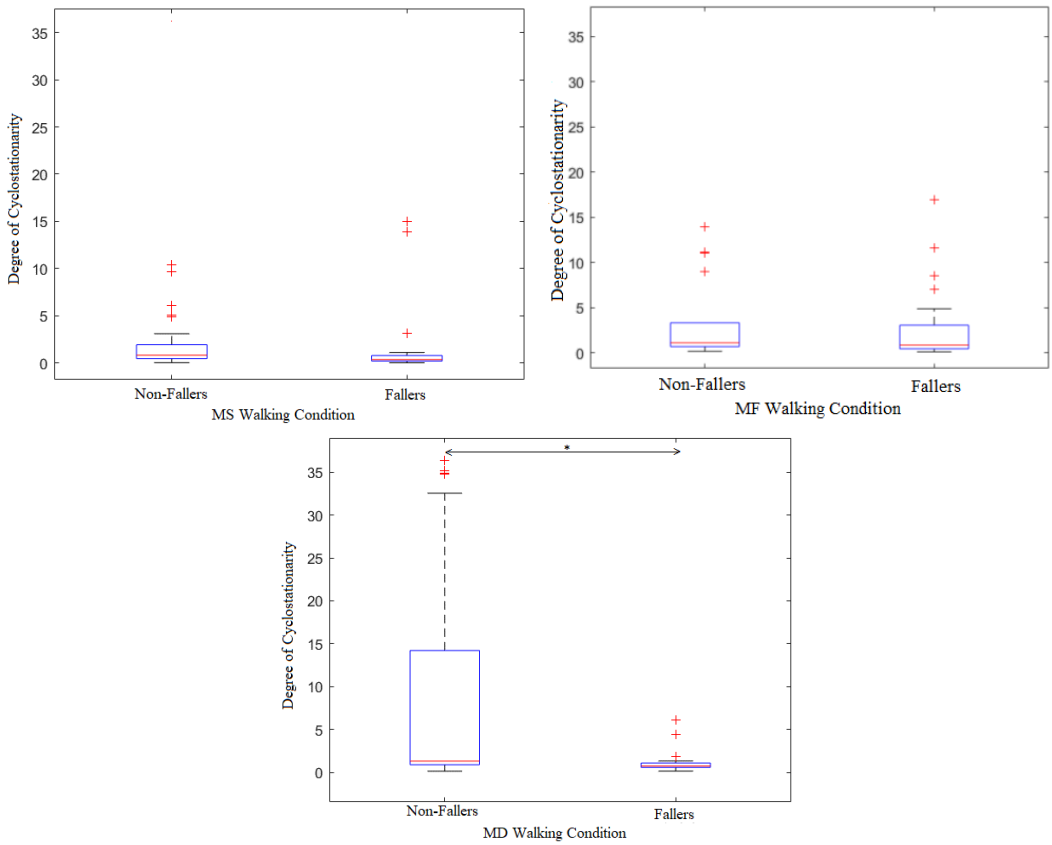

Figure 7.1 – The boxplot of average degree of cyclostationarity between fallers and non-fallers

A one-way repeated measures ANOVA [Gad98] was also conducted to determine whether there was a statistically significant difference in the DC between fallers and non-fallers in the three different cases of walking. There were relatively few outliers (the red Xs on the figure) in the three cases. No statistical significant differences were found between faller and non-fallers (Figure 7.1) in the cases of MS and MF walking conditions. Whereas in the case of MD case of de-counting while walking, the average DC was statistically significantly different between fallers and non-fallers (p<0.05).

7.1.2 Results of the Classification1 Models with 10 Classical Features versus the Degree of Cyclostationarity as a Single Feature

Five classifier models were used with default hyperparameters set: KNN, SVM with polynomial kernels of degree 3, ANN with 10 nodes in a single hidden layer, Decision Trees, and Logistic Regression.

The results of a 100 times 10 folds cross validation was compared to the 10 times 10 folds cross validation and found to have no statistical significant differences using the t-student test. Therefore, a 10 times 10 folds cross validation process was chosen to be used in all classification models. This is to prevent over-fitting by partitioning the data-set into folds and evaluating their performance at each fold.

Model evaluation parameters included accuracy, sensitivity, specificity, and precision with their standard deviation (SD). Table 7.1 shows the results of the different supervised classification models with 3 different walking conditions, 2 different feature sets, and 5 different classification methods. The best performance with the ten classical features as inputs was 61.85% accuracy, 53.52% sensitivity, 70.19% specificity, and 64.24% precision using K-nearest neighbors as a classifier. The use of the averaged degree of cyclostationarity as a single feature instead of the ten classical features improved model performance with 68.43% accuracy, 54.26% sensitivity, 82.59% specificity, and 75.83% precision using K-nearest neighbors.

The statistical t-test for pairwise comparison was computed and it was confirmed that there is statistical significant differences between the KNN model of highest accuracy and the other models listed in the table.

Table 7.1 – Results of the Classification Models using 10 Classical Features and the Degree of Cyclostationarity as a Single Feature

Walking Condition	Feature Set	Classification Model	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
MS	10 CF		52.04% ± 2.54%	39.07% ± 13.04%	65% ± 14.14%	53.24% ± 4.90%
	DC		63.06% ± 1.27%	55.56% ± 2.14%	70.56% ± 2.54%	65.40% ± 1.75%
MF	10 CF	**KNN**	48.06% ± 2.20 %	29.07% ± 20.62%	60.56% ± 9.24%	38.61 %±18.23 %
	DC		48.52% ± 5.37%	36.48% ± 15.02%	67.04% ± 22.88%	46.58% ± 8.07%
MD	10 CF		61.85% ± 2.17%	53.52% ± 3.32%	70.19% ± 2.95%	64.24% ± 2.68%
	DC		**68.43**% ± **1.66**%	54.26% ± 0.89%	82.59% ± 3.05%	75.83% ± 3.38%
MS	10 CF		58.15% ± 1.62%	77.41%± 7.19%	38.89%± 7.15%	55.95% ±1.31 %
	DC		62.78% ± 1.62%	80.37% ± 1.79%	45.19% ± 1.79%	59.45% ± 1.20%
MF	10 CF	SVM	50.56% ± 5.51%	52.22%± 5.08%	48.89% ± 7.16%	50.65% ± 5.28%
	DC		56.85% ± 2.70%	59.07% ± 20.22%	54.63% ± 16.22%	56.93% ± 1.90%
MD	10 CF		59.91% ± 0.62%	70.19% ± 4.32%	28.70% ± 2.66%	56.10% ± 0.33%
	DC		**63.33**% ± **1.81**%	90.63% ± 3.24%	56.48% ±6.25%	61.87% ± 2.23%
MS	10 CF		60.00%± 5.55%	63.15% ± 16.76%	56.85% ± 12.92%	59.37% ± 6.05%
	DC		62.69% ± 6.83%	72.22% ± 26.99%	53.15% ± 24.65%	60.39% ± 5.81%
MF	10 CF	ANN	57.31% ± 7.20%	54.07% ± 19.96%	60.56% ± 18.06%	58.36%± 6.72%
	DC		56.48% ± 6.16%	72.96% ± 12.35%	40.00% ± 13.92%	55.09% ± 5.44%
MD	10 CF		59.91% ± 0.62%	60.93% ±26.12%	70.93% ± 11.84%	69.73% ± 12.45%
	DC		**65.93**% ± **8.83**%	91.11% ± 3.24%	28.70% ± 2.66%	56.10% ± 0.33%
MS	10 CF		54.44% ± 2.86%	55.00% ± 5.02%	53.89% ± 3.54%	54.36% ± 2.67%
	DC		**64.63**% ± **2.95**%	62.41% ± 5.52%	66.85% ± 2.82%	65.26% ± 2.52%
MF	10 CF	Decision Trees	51.85% ± 4.39v	50.74% ± 7.67%	52.96% ± 5.03%	51.75% ± 4.37%
	DC		44.44% ± 5.38%	44.81% ± 7.45%	44.07% ± 5.84%	44.35% ± 5.62%
MD	10 CF		56.76% ± 4.56%	56.11% ± 6.18%	57.41% ± 6.05%	56.88% ± 4.45%
	DC		58.33% ± 1.85%	54.81% ±3.83%	61.85% ± 2.50v	58.95%± 1.82%
MS	10 CF		56.11% ± 3.09%	62.22% ± 3.29%	50.00% ± 4.09%	55.47% ± 2.85%
	DC	Logistic	50.28% ± 3.15%	75.19%± 5.32%	25.37% ± 4.46%	50.17% ± 20.8%
MF	10 CF	Regression	55.00% ± 4.46%	57.96%± 4.79%	52.04% ± 6.79%	54.85% ± 4.36%
	DC		49.72% ± 2.55%	73.33% ± 5.87%	26.11%± 4.66%	49.79% ± 1.69%
MD	10 CF		59.81% ± 0.48%	62.96% ± 2.31%	30.74% ± 0.62%	56.21% ± 0.34%
	DC		**60.56**% ± **2.06**%	88.89% ± 0.36%	58.15% ± 3.40%	60.11% ± 2.06%

From the above results, one can conclude that utilizing the degree of cyclostationarity improved model predictive performance while reducing its complexity. Therefore, we advocated its inclusion for elderly fall-risk prediction. In addition, the MD walking condition (de-counting as a dual task) improved model prediction accuracy in the KNN, SVM, ANN, and Logistic Regression classifiers. KNN achieved the highest accuracy using a single cyclostationary feature during the MD walking condition. A drawback that needs to be noted, is that KNN compared to other classification methods requires a large real time computation as it needs the complete data for every classification. This opened the door to look into further improvement and optimization.

Results of this section was published in [BNB+21] and won Best Paper Award.

7.2 Relief-F Feature Selection with Different Walking Conditions

In the previous section, we showed using both statistical and machine learning methods that the features selected by ANOVA can be used to detect fallers in elderly people. After conducting further statistical analysis on the data-set, we were able to identify two new categories of features: gender and stride time. Stride time is defined as the time elapsed between the first contact of two consecutive footsteps of the same foot and is expressed in milliseconds. In particular, we found the mean and standard deviation of the stride time are significant features to detect fallers and non-fallers. Therefore, we included these three features in our model.

Then, we constructed a second set of features using Relief-F method, which ranks the features according to their importance. The number of ranked best predictors were selected based on the number that yielded the highest classification accuracy in each

set of features associated with each type of walking condition. The features used in this section are shown in Table 7.2.

Table 7.2 – Features for Each Type of Walking Condition: MS, MF, and MD

Feature Reference Number	Feature Abbreviation	Feature Description
1	PW_R	Pulse Width of the Right Foot
2	US_R	Undershoot of the Right Foot
3	US_L	Undershoot of the Left Foot
4	DTC_L	Duty Cycle of the Left Foot
5	SR_R	Slew Rate of the Right Foot
6	SR_L	Slew Rate of the Left Foot
7	Range_R	Range of the Right Foot
8	Range_L	Range of the Left Foot
9	Skw_R	Skewness of the Right Foot
10	Skw_L	Skewness of the Left Foot
11	Gender	Male or Female
12	M_ST	Mean of the Stride Time
13	STD_ST	Standard Deviation of the Stride Time
14	DC	Degree of Cyclostationarity

In this section, we used the same five classification methods as the section before also with the same default settings. A 100 times 10 folds cross-validation was applied on all classifier models.

The results obtained are shown in Table 7.3 for the MS normal walking condition, Table 7.4 for the MF condition, and Table 7.5 for the MD walking condition. Finally, Table 7.6 shows the results for all the features in all types of walking conditions. The first feature set includes all the 14 features mentioned in Table 1. The second feature set were selected using Relief-F. The number of best predictors used in the models was chosen based on the combination, leading to the highest classification accuracy, as shown in figures7.2, 7.3, 7.4, and 7.5.

Table 7.3 shows the results in the MS walking condition. The best performance with

Table 7.3 – Results of the Classification Models for the MS Walking Condition

Feature Set	Classification Model	Accuracy% ±SD%	Sensitivity% ±SD%	Specificity%±SD%	Precision% ±SD%
All Features (14)	**KNN**	66.72 % ± 2.55%	62.96% ± 1.23%	70.37% ± 7.96%	68.00% ± 3.85%
Selected Features (12,3,13,6,9)		70.43% ± 1.65%	68.52% ± 4.69%	72.22% ± 8.74%	71.15% ± 3.56%
All Features (14)	**SVM**	69.40 % ± 5.55%	74.07% ± 7.21%	64.81% ± 4.14%	67.80% ± 5.73%
Selected Features (12,3,13,6,9)		71.32% ± 2.53%	77.78% ± 5.74%	64.81% ± 6.39%	68.85% ± 8.06%
All Features (14)	**ANN**	75.92 % ± 5.41%	76.92% ± 1.35%	75.00% ± 4.21%	74.07% ± 6.42%
Selected Features (12,3,13,6,9)		**78.70**% ± **4.57**%	**78.18**% ± **4.29**%	**79.25**% ± **6.35**%	**79.63**% ± **7.58**%
All Features (14)	**Decision Tree**	63.12% ± 4.37%	68.52% ± 5.84%	57.41% ± 7.53%	61.67% ± 4.47%
Selected Features (12,3,13,6,9)		62.35% ± 4.02%	59.26% ± 3.78%	64.81% ± 7.58%	62.75% ± 5.71%
All Features (14)	**Logistic Regression**	68.54% ± 2.12%	74.07% ± 5.78%	62.96% ± 7.14%	66.67% ± 5.98%
Selected Features (12,3,13,6,9)		69.44% ± 5.01%	75.93% ± 2.30%	62.96% ± 7.84%	67.21% ± 6.33%

Table 7.4 – Results of the Classification Models for the MF Walking Condition

Feature Set	Classification Model	Accuracy% ±SD%	Sensitivity% ±SD%	Specificity%±SD%	Precision% ±SD%
All Features (14)	**KNN**	68.72 % ± 3.46%	60.04% ± 5.63%	76.17% ± 5.79%	71.32% ± 7.80%
Selected Features (12,13,14,3,6, 11,4,9)		72.00% ± 2.28%	65.39% ± 4.87%	78.02% ± 3.67%	75.24% ± 5.35%
All Features (14)	**SVM**	67.92% ± 5.23%	69.02% ± 6.16%	67.91% ± 5.13%	67.56% ± 8.49%
Selected Features (12,13,14,3,6, 11,4,9)		72.37% ± 4.05%	80.55% ± 2.57%	63.83% ± 6.00%	69.68% ± 7.08%
All Features (14)	**ANN**	71.86 % ± 2.77%	75.63% ± 4.69%	70.06% ± 5.84%	64.96% ± 5.63%
Selected Features (12,13,14,3,6, 11,4,9)		**75.37**% ± **6.41**%	**74.37**% ± **6.72**%	**76.54**% ± **7.08**%	**76.95**% ± **6.57**%
All Features (14)	**Decision Tree**	63.21% ± 5.32%	58.62% ± 7.63%	66.06% ± 7.65%	63.05% ± 6.84%
Selected Features (12,13,14,3,6, 11,4,9)		68.64% ± 6.12%	76.37% ± 5.64%	63.07% ± 5.32%	67.33% ± 6.97%
All Features (14)	**Logistic Regression**	56.58% ± 3.04%	60.32% ± 2.63%	55.56% ± 8.32%	60.51% ± 3.19%
Selected Features (12,13,14,3,6, 11,4,9)		68.37% ± 3.64%	74.98% ± 5.63%	62.33% ± 4.97%	66.36% ± 7.63%

Table 7.5 – Results of the Classification Models for the MD Walking Condition

Feature Set	Classification Model	Accuracy% ±SD%	Sensitivity% ±SD%	Specificity%±SD%	Precision% ±SD%
All Features (14)	**KNN**	69.90 % ± 6.19%	60.07% ± 7.01%	79.94% ± 6.88%	75.00% ± 7.24%
Selected Features (3,14,12,11,5, 13)		70.90% ± 4.22%	63.52% ± 4.68%	78.81% ± 5.64%	74.01% ± 4.46%
All Features (14)	**SVM**	67.98% ± 5.00%	67.24% ± 2.36%	69.52% ± 4.59%	68.39% ± 5.63%
Selected Features (3,14,12,11,5, 13)		73.57% ± 6.18%	72.94% ± 6.49%	73.03% ± 4.81%	72.84% ± 5.39%
All Features (14)	**ANN**	75.06 % ± 5.61%	72.67% ± 6.74%	73.12% ± 2.94%	75.99% ± 3.78%
Selected Features (3,14,12,11,5, 13)		**79.93**% ± **2.56**%	**77.81**% ± **5.72**%	**77.78**% ± **5.28**%	**78.43**% ± **5.34**%
All Features (14)	**Decision Tree**	70.26% ± 2.37%	73.11% ± 3.61%	65.64% ± 4.88%	68.20% ± 4.67%
Selected Features (3,14,12,11,5, 13)		71.31% ± 5.41%	71.33% ± 4.85%	71.36% ± 5.23%	72.66% ± 5.23%
All Features (14)	**Logistic Regression**	67.17% ± 3.21%	67.81% ± 5.74%	67.24% ± 5.99%	67.83% ± 6.41%
Selected Features (3,14,12,11,5, 13)		68.63% ± 6.27%	70.35% ± 2.91%	69.83% ± 6.31%	69.11% ± 5.44%

Table 7.6 – Results of the Classification Models for All the Walking Conditions

Feature Set	Classification Model	Accuracy% ±SD%	Sensitivity% ±SD%	Specificity%±SD%	Precision% ±SD%
All Features (40)	**KNN**	63.45 % ± 5.12%	55.23% ± 5.61%	75.62% ± 7.59%	70.56% ± 6.14%
Selected Features (12_MS, 14_MD, 11. 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		68.53% ± 1.92%	60.32% ± 4.58%	70.55% ± 6.33%	74.54% ± 6.81%
All Features (14)	**SVM**	64.82% ± 3.67%	68.73% ± 4.38%	62.75% ± 5.64%	64.26% ± 5.32%
Selected Features (12_MS, 14_MD, 11. 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		66.59% ± 3.88%	68.96% ± 2.37%	62.40% ± 5.07%	64.78% ± 5.48%
All Features (14)	**ANN**	75.51 % ± 2.74%	70.62% ± 2.99%	73.83% ± 5.20%	76.82% ± 4.17%
Selected Features (12_MS, 14_MD, 11. 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		**81.16**% ± **2.87**%	**79.63**% ± **5.63**%	**78.23**% ± **5.11**%	**79.81**% ± **4.82**%
All Features (14)	**Decision Tree**	63.47% ± 4.27%	63.98% ± 3.75%	65.32% ± 5.84%	63.71% ± 4.27%
Selected Features (12_MS, 14_MD, 11. 13_MS, 13_MD, 3_MS, 13_MF, 14_MF)		72.69% ± 2.86%	71.41% ± 3.78%	73.47% ± 5.21%	73.20% ± 3.68%
All Features (14)	**Logistic Regression**	64.50% ± 4.87%	68.87% ± 4.34%	59.89% ± 3.33%	63.11% ± 3.96%
Selected Features (12_MS, 14_MD, 11. 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		70.71% ± 3.04%	73.28% ± 3.85%	68.21% ± 4.25%	68.57% ± 5.42%

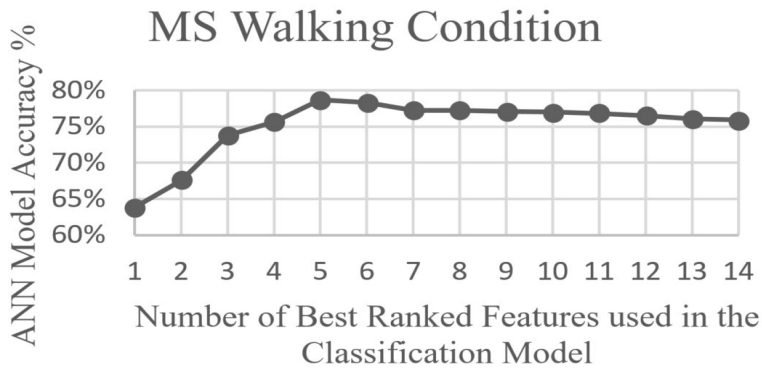

Figure 7.2 – Number of Best-Ranked Features in MS Walking Condition

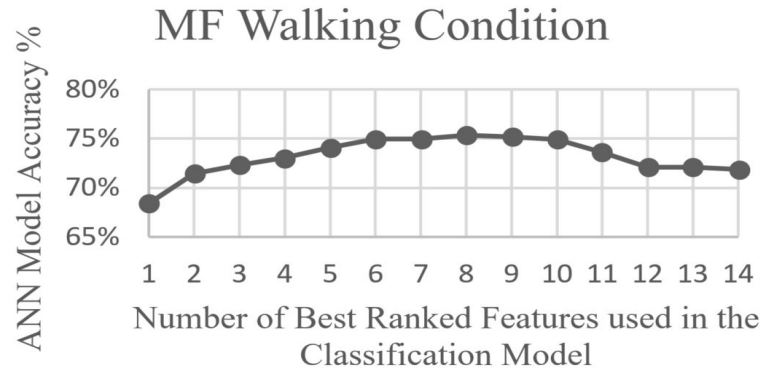

Figure 7.3 – Number of Best-Ranked Features in MF Walking Condition

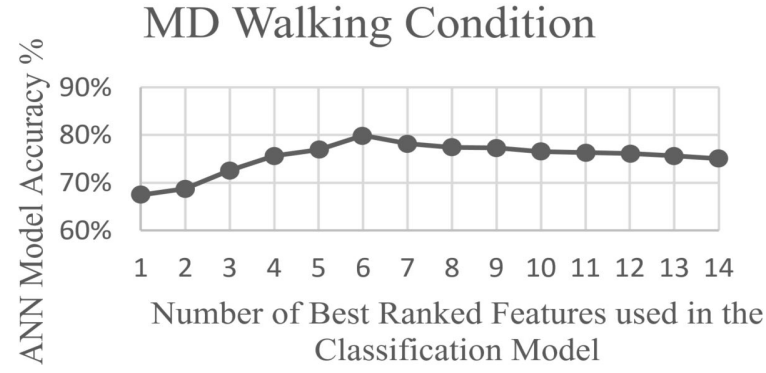

Figure 7.4 – Number of Best-Ranked Features in MD Walking Condition

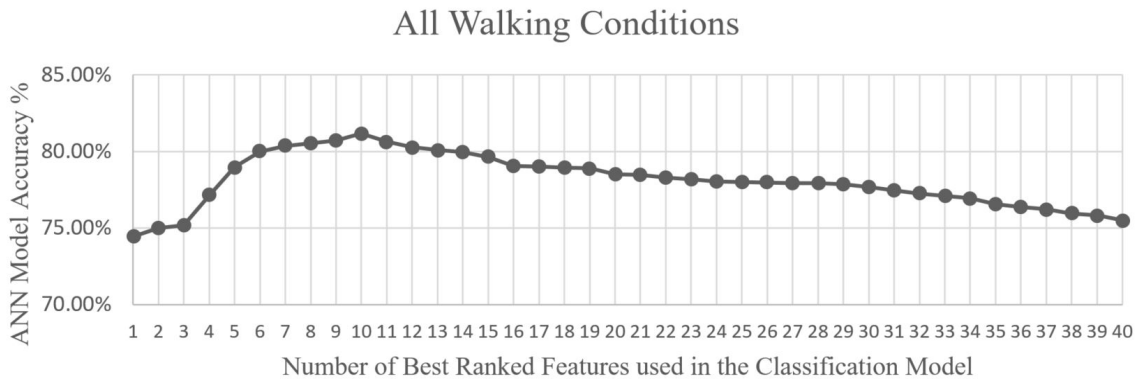

Figure 7.5 – Number of Best-Ranked Features in All Walking Conditions

all the 14 features as inputs was 75.9%accuracy, 76.92% sensitivity, 75.00% specificity, and 74.07% precision using ANN as a classifier. This performance was improved using Relief-F where, as shown in figure 7.2, performed best with the first 5 best predictors for this model (12, 3, 13, 6, 9). With feature selection, the ANN model performance improved to 78.7% accuracy, 78.18% sensitivity, 79.25%specificity, and 79.63% precision.

Table 7.4 shows the results in the MF walking condition. The best performance with all the 14 features as inputs was 71.86% accuracy, 75.63% sensitivity, 70.06% specificity, and 64.96% precision also using ANN as a classifier. This performance was improved using the best 8 predictors for this model, as shown in figure 7.3 (which were: 12, 13, 14, 3, 6, 11, 14, 9). With feature selection, the ANN model performance improved to 75.37% accuracy, 74.37% sensitivity, 76.54% specificity, and 76.95%precision.

Table 7.5 shows the results in the MD walking condition. The best performance with all the 14 features as inputs was 75.05% accuracy, 72.67% sensitivity, 73.12% specificity, and 75.99% precision, also using ANN as a classifier. This performance was improved using Relief-F, which selected the best 6 predictors for this model, as shown in figure 7.4. (These features were: 3, 14, 12, 11, 5, 6). With feature selection, the ANN model performance improved to 79.93% accuracy, 77.81% sensitivity, 77.78% specificity, and 78.43% precision.

The results shown in Table 7.6 show the performance of the model using all the predictors from the 3 walking conditions combined. As shown in figure 7.5, the model accuracy scored the highest with the first 10 best-ranked features: 12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF. The best performance with all the 40 features as inputs was 75.51 % accuracy, 70.62% sensitivity, 73.83% specificity, and 76.82% precision, also using ANN as a classifier. The ANN model performance improved to 81.16% accuracy, 79.63% sensitivity, 78.23% specificity, and 79.81% precision with the feature reduction done using Relief –F.

The results show that the ANN classification model performed the best out of the classification models explored in our study using the features chosen as inputs for the models. This performance was generally improved using the Relief-F feature reduction method. Furthermore, the MD walking condition helped achieving the highest accuracies compared to the MS and MF conditions. This points out that in the MD walking condition, which is walking while de-counting from 50, there are higher differences that can be captured between faller and non-fallers compared to the baseline and the secondary task of naming animals.

It is noticeable that the common most important features in prediction of elderly people at risk of falling were: The mean stride time (M_ST), standard deviation stride time (STD_ST), and Undershoot of left foot (US_L) in all the three walking conditions. The Degree of Cyclostationarity (DC) and gender were important features in the MF and MD walking conditions only.

Using all the features from all walking conditions improved system accuracy after applying Relief-F and achieved the best performance in terms of accuracy with the ANN machine-learning algorithm. 10 features were selected out of the 40 features: one of them is gender, and 3 features from each type of walking condition. This points out that combining the 3 types of walking conditions in one classification model helped in

achieving better performance than considering each type of walk condition by itself.

The statistical paired t-test for pairwise comparison was computed and it was confirmed that there were statistical significant differences between the ANN model in each type of walking condition and the other classification models with the same type of walking condition. In addition, the statistical independent t-test was applied to confirm statistical significant differences between the ANN models in the MS, MF, MD, and the case involving all walking conditions.

To summarize, in this section, we included three additional features: Mean of the stride time, Standard Deviation of the stride time, and Gender. In addition, we used Relief-F method for feature selection. We compared the results of different classification models in terms of feature, type of walking condition, and methods of classification. The use of the Relief-F method reduced the features and selected the best out of them. This step generally improved and optimized the classification model. Results show that some of the best features for prediction of elderly people at risk of falling were the stride time in all walking conditions, and the degree of cyclostationariy and gender in the MF and MD walking condition. Combining the features of all types of walking conditions in the ANN classification model along with using Relief-F feature selection method lead to the best performance in terms of accuracy. The results of of this section was publish in a journal [BND+21].

7.3 Hyperparameter Tuning and Feature Selection to Improve Model Performance

This section is an extension of the previous sections. Combining the features of all types of walking conditions mentioned in Table 7.7 in an Artificial Neural Network

(ANN) classification model and using the Relief-F feature selection method led to improved performance in terms of accuracy that reached 81.16% [BND+21]. In this section we introduce a new feature, which is the average difference in pressure between toes and heels, use two feature selection methods, and use grid search cross-validation as a hyperparameter tuning method to improve and optimize the classification model.

A 100 times 10 folds cross-validation was applied on all classifier models. For each classification algorithm 3 different feature sets from all walking conditions were used: all features, features selected by Relief-F, and features selected by SBS. The Hyperparameters chosen for each classification model are shown in Table 7.8.

Table 7.7 – Features for Each Type of Walking Condition: MS, MF, and MD

Abbreviation	**Description**
PW_R	Pulse Width of the Right Foot
US_R	Undershoot of the Right Foot
US_L	Undershoot of the Left Foot
DTC_L	Duty Cycle of the Left Foot
SR_R	Slew Rate of the Right Foot
SR_L	Slew Rate of the Left Foot
Range_R	Range of the Right Foot
Range_L	Range of the Left Foot
Skw_R	Skewness of the Right Foot
Skw_L	Skewness of the Left Foot
Gender	Male or Female
M_ST	Mean of the Stride Time
STD_ST	Standard Deviation of the Stride Time
Diff_P	Difference in Pressure in Toes and Heels
DC	Degree of Cyclostationarity

The features selected by each feature selection technique are shown in Table 7.9. It is noticeable that the common most important features include: the difference in pressure between toes and heel, mean stride time, standard deviation of stride time, and undershoot of left foot, degree of Cyclostationarity and gender. The results also show that the features extracted during the MD walking condition were more than that of MS

and MF among the important selected features. This suggests that in the MD walking condition (de-counting from 50), higher differences can be captured between fallers and non-fallers compared to the baseline and the secondary task of naming animals.

Table 7.8 – Hyperparameters Tuned Using Grid Search CV

Algorithm	Hyperparameters Chosen
KNN	Leaf size= 5, K=7, Euclidean distance
SVM	Gaussian RBF, Kernel, C(regularization)= 10, Gamma=0.1
ANN	number of neurons=30, activation function= Sigmoid, optimizer, learning rate=0.07, batch size=9, and epochs =50
Decision Trees	criterion= gini, splitter=best, Min samples required to split= 2, min samples required to be at leaf node = 1
Logistic Regression	C=0.3 and Alpha=0.2

The results shown in Table 7.10 exhibit the performance of the classification models using the different feature sets. When using all the 43 features, ANN achieved the best performance with all the 43 features as inputs with 75.93% accuracy, 77.78% sensitivity, 74.07% specificity, and 75.00% precision.

The ANN model performance improved to 84.11% accuracy, 79.63% sensitivity, 88.68% specificity, and 87.76% precision with the feature reduction done using Relief-F. Furthermore, the SBS improved the performance to 89.81% accuracy, 90.74% sensitivity, 88.89% specificity, and 89.09% precision.

The results show that the ANN classification model performed the best out of the classification models explored in our study. This performance was improved by around 10% using the SBS feature selection technique.

Table 7.9 – Feature Sets Selected by Relief-F and SBS

Feature Selection	**ML Algorithm**	**Feature Sets**
Relief-F	All	M_ST (MS), DC (MD),Diff_P (MD), Gender, STD_ST (MS), M_ST (MF), Diff_P (MF) M_ST (MD),STD_ST (MD)
SBS	KNN	M_ST (MS), DC (MD), Gender, SR_R (MF), STD_ST (MD), M_ST (MD), Diff_P (MD), US_L (MD)
SBS	SVM	Diff_P (MD), DC (MD), Gender, Diff_P (MF), M_ST (MS), SR_L (MD) STD_ST(MD), DTC_L (MD), PW_R (MD)
SBS	ANN	DC (MD), Gender, Diff_P (MD), US_L (MD), SR_R (MD), M_ST (MS) Diff_P (MF), M_ST(MF), STD_ST(MD)
SBS	Decision Trees	DC (MD), Gender, Diff_P (MD), M_ST (MD), STD_ST (MD), US_L (MF), Skx_R (MS), Diff_P (MF), Range_L (M), M_ST (MS)
SBS	Logistic Regression	M_ST (MS), Gender, PW_R (MD), M_ST(MD), STD_ST(MD), Diff_P (MD), Diff_P (MF), DC (MD), Skw_L (MF), DTC_L (MS), Range_R (MF)

Table 7.10 – Results of the Classification Models

Feature Set	Classification Model	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
All	KNN	63.89% ± 0.98%	61.11% ± 1.27%	66.67% ± 0.26%	64.71% ± 0.99%
Selected by Relief-F		70.37% ± 2.78%	66.60% ± 3.74%	74.07% ± 2.27%	72.00% ± 1.88%
Selected by SBS		84.266% ± 2.73 %	90.74% ± 2.95%	77.78% ± 2.08%	80.33% ± 3.11 %
All	SVM	65.74% ± 0.47%	61.11% ± 0.52%	70.37% ± 0.85%	67.35% ±0.37 %
Selected by Relief-F		73.15% ± 1.35%	66.67% ± 2.14%	79.63% ± 3.07%	76.605% ± 2.72%
Selected by SBS		85.19% ± 0.84%	83.33% ± 0.68%	87.04% ± 1.24%	86.54% ± 0.86%
All	**ANN**	75.93% ± 1.34%	77.78% ± 2.47%	74.07% ± 1.39%	75.00% ± 1.52%
Selected by Relief-F		84.11% ± 3.95%	79.63% ± 3.01%	88.68% ± 2.41%	87.76% ± 2.56%
Selected by SBS		**89.81**% ± **2.78**%	**90.74**% ± **2.14**%	**88.89**% ± **3.06**%	**89.09**% ± **2.67**%
All	Decision Tree	65.74% ± 0.88%	61.11% ± 1.24%	70.37% ± 0.52%	67.35% ± 1.81%
Selected by Relief-F		73.15% ± 3.27%	64.8% ±2.11%	81.48% ± 3.58%	77.78% ± 2.49%
Selected by SBS		83.33% ± 1.48%	81.48% ± 2.13%	85.19% ± 1.11%	84.62% ± 2.06%
All	Logistic Regression	64.81% ± 0.46%	61.11% ± 0.67%	68.52% ± 0.99%	66.00% ± 1.02%
Selected by Relief-F		72.22% ± 3.13%	65.81% ±3.38%	79.63% ± 2.47%	76.09% ± 2.64%
Selected by SBS		80.56% ± 2.32%	83.33% ± 2.14%	77.78% ± 2.68%	78.95% ± 2.04%

The statistical paired t-test for pairwise comparison was computed to test the differences between the different algorithms, and it was confirmed that there were statistically significant differences between the ANN model and the other classification models with 95% confidence.

Results in this study show that the Degree of Cyclostationarity, gender, the mean and standard deviation of stride time of pressure walking signals are significant predictors for elderly fallers. We also showed that using feature selection methods improved the performance of the model and implementing hyperparameter tuning using grid search on each of the five classification models to obtain an optimized architecture for each classification model. Our analysis showed that the proposed model outperforms existing models by at least 10%. It is worth noting also that all performance measures of the selected model are above 88%, which is not the case for most existing models that increases one measure while reducing the others. The results obtained in this section were presented in two conferences Congrès National de la Recherche des IUT (CNRIUT) 2022 and GRETSI 2022.

7.4 Heat Map Representation of Spectral Correlation as Inputs to CNN

In this section we present an alternative method of representing signal cyclostationarity as heat-map images and using convolutional neural network (CNN) with the ADAM optimization method to predict the risk of falling in 411 subjects over the age of 65.

Table 7.11 shows the performance of the CNN in 3 different walking conditions using the default hyperparameters, whereas Table 7.12 displays the improvement in

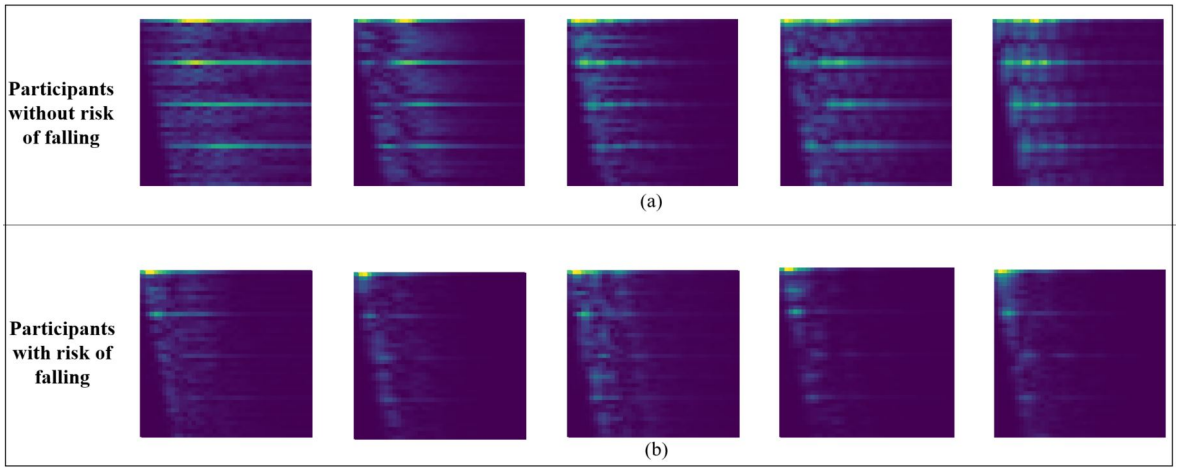

Figure 7.6 – 5 Examples of each of the two classes: (a) without risk and (b) with risk of falling during MD walking condition

CNN performance when using ADAM Optimization. In both tables, the MD walking condition achieves the best performance out of the three walking modes. In our previous work, using the same dataset but using conventional machine learning methods, the best model performance was able to achieve an accuracy of 81.16%, sensitivity of 79.63%, specificity of 78.23%, and precision of 79.81% using an artificial neural network and including cyclostationary, time and frequency domain features from all walking conditions. The suggested CNN model was able to surpass the latter's performance with the use of only one type of walking conditions. The proposed CNN model might make the data collection of future patients simpler and faster with only one walking mode rather than having three types of walking conditions.

Table 7.11 – Performance of the CNN Models without using ADAM Optimization

CNN Performance Without ADAM Optimization				
Walking Condition	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
MS	68.89% ± 0.98%	67.11% ±1.27%	68.67% ± 0.26%	68.71% ± 0.99 %
MF	70.37% ± 2.78%	66.60% ±3.74%	74.07% ± 2.27%	72.00% ± 1.88 %
MD	79.93% ± 2.56%	77.81% ±5.72%	77.78% ± 5.28%	78.43% ± 5.34 %

The statistical t-test for pairwise comparison was calculated and a statistically significant difference was confirmed between the CNN model during MD walking

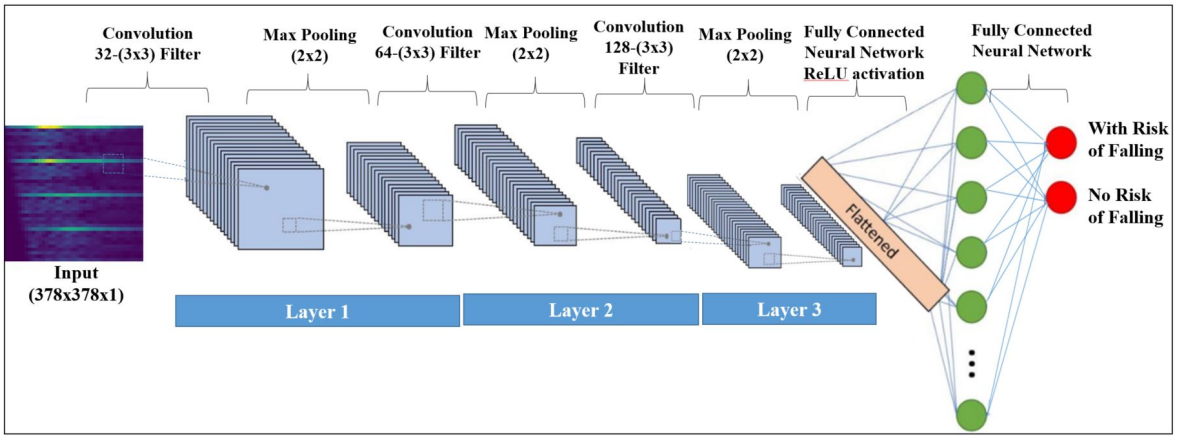

Figure 7.7 – CNN architecture with 3 layers

Table 7.12 – Performance of the CNN Models using ADAM Optimization

CNN Performancde With ADAM Optimization				
Walking Condition	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
MS	83.33% ± 1.23%	87.04% ±1.47%	79.63% ± 1.55%	81.03% ± 1.73 %
MF	86.11% ± 2.04%	88.89% ±2.14%	83.33% ± 2.12%	84.21% ± 2.31 %
MD	87.96% ± 0.14%	85.19% ±0.27%	90.74% ± 0.21%	90.20% ± 0.47 %

condition and the other models with other walking conditions mentioned n the table.

Previous results in this thesis advocated the inclusion of cyclostationary properties of insole pressure walking signals in predicting the risk of falling within the elderly community. The above mentioned results suggested a novel method of representing cyclostationary properties of walking signals as images to be used as inputs to a CNN model. The suggested model outperformed the previous conventional machine learning models, reduced the complexity with fewer needed inputs, and saved time using one single walking condition (walking while de-counting from 50) instead of three. The results obtained in this section is accepted as a conference paper IECBES 2022 that will be presented in December 2022.

7.5 The Performance of the CNN with Different Types of Input Images: Combination Walking Modes, Grey Scale and Log-Transformation

Table 7.13 – Performance of the CNN Models with Different Input Images and without using Adam Optimization

CNN Performance Without ADAM Optimization					
Images	Walking Modes	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
Colored	MS	68.89% ± 0.98%	67.11% ±1.27%	68.67% ± 0.26%	68.71% ± 0.99 %
	MF	70.37% ± 2.78%	66.60% ±3.74%	74.07% ± 2.27%	72.00% ± 1.88 %
	MD	79.93% ± 2.56%	77.81% ±5.72%	77.78% ± 5.28%	78.43% ± 5.34 %
	All Walking Modes	80.74% ± 3.12%	80.89% ± 3.35%	80.59% ± 4.08%	80.31% ± 3.26%
Grey Scale	MS	70.74% ± 0.47%	61.11 % ± 0.52%	70.37% ± 0.85%	67.35% ± 0.37%
	MF	73.15% ± 1.35%	66.67% ± 2.14%	79.63% ± 3.07%	76.60% ± 2.72%
	MD	81.19% ± 0.84%	79.33% ± 0.68%	80.04% ± 1.24%	80.54% ± 0.86%
	All Walking Modes	82.52% ± 2.44%	82.59% ± 2.56%	82.44% ± 2.77%	82.34% ± 2.89%
Log-T Colored	MS	73.57% ± 6.18%	72.94% ± 6.49%	73.03% ± 4.81%	72.84% ± 5.39%
	MF	75.05% ± 5.61%	72.67% ± 6.74%	73.12% ± 2.94%	75.99% ± 3.78%
	MD	83.26% ± 2.73%	82.74% ± 2.95%	84.78% ± 2.08%	84.33% ± 3.11%
	All Walking Modes	85.59% ± 1.71%	84.74%± 1.84%	85.44% ± 2.07%	85.23% ± 2.39%
Log-T Grey Scale	MS	75.93% ± 1.34%	77.78% ± 2.47%	74.07% ± 1.39 %	75.00 % ± 1.52%
	MF	84.11% ± 3.95%	83.63% ± 3.01%	84.68% ± 2.41%	84.76% ± 2.56%
	MD	85.81% ± 2.78%	85.74% ± 2.14%	85.89% ± 3.06%	85.09% ± 2.67%
	All Walking Modes	**86.44**% ± **2.35**%	**85.59**% ± **2.64**%	**86.3**% ± **2.72**%	**86.15**% ± **2.51**%

Table 7.14 – Performance of the CNN Models with Different Input Images and using Adam Optimization

CNN Performancde With ADAM Optimization					
Images	Walking Modes	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
Colored	MS	83.33% ± 1.23%	87.04% ±1.47%	79.63% ± 1.55%	81.03% ± 1.73 %
	MF	86.11% ± 2.04%	88.89% ±2.14%	83.33% ± 2.12%	84.21% ± 2.31 %
	MD	87.96% ± 0.14%	85.19% ±0.27%	90.74% ± 0.21%	90.20% ± 0.47 %
	All Walking Modes	90.74% ± 0.11%	88.89% ± 0.45%	92.59% ± 0.32%	92.31% ± 0.64%
Grey Scale	MS	85.19% ± 2.44%	87.04 % ± 2.17%	83.33% ± 2.36%	83.93% ± 2.78%
	MF	88.89% ± 1.58%	87.04% ± 1.34%	90.74% ± 1.72%	90.38% ± 1.04%
	MD	92.59% ± 1.07%	96.3% ± 1.42%	88.89% ± 1.06%	89.66% ± 1.12%
	All Walking Conditions	93.52% ± 1.28%	92.59% ± 1.63%	94.44% ± 1.04%	94.34% ± 1.44%
Log-T Colored	MS	84.26% ± 3.10%	83.33% ± 3.14%	85.19% ± 3.67%	84.91% ± 3.44%
	MF	87.04% ± 2.04%	85.19% ± 2.11%	88.89% ± 2.35%	88.46% ± 2.71%
	MD	91.67% ± 3.65%	92.59% ± 3.29%	90.74% ± 3.22%	90.91% ± 3.48%
	All Walking Modes	92.59% ± 1.44%	90.74% ± 1.26%	94.44% ± 1.37%	94.23% ± 1.69%
Log-T Grey Scale	MS	86.11% ± 2.77%	85.19% ± 2.54%	87.04% ± 2.31%	86.79% ± 2.19%
	MF	89.81% ± 2.59%	88.89% ± 2.37%	90.74% ± 2.25%	90.57% ± 2.67%
	MD	**93.98**% ± **1.07**%	**92.59**% ± **1.33**%	**94.44**% ± **1.41**%	**94.34**% ± **1.27**%
	All Walking Modes	**94.44**% ± **3.46**%	**92.59**% ± **3.28**%	**96.30**% ± **3.61**%	**96.15**% ± **3.08**%

The results in the previous section showed that using heat map images of the

average spectral correlation as input to a CNN model can be employed to predict future falls of elderly people. In addition, incorporating the Adam technique to optimize the hyperparameters also helped improve performance, which can also be seen in Tables 7.13 and 7.14.

In this section, we present the results of the same CNN model in the previous section but with different types of images. These images differed in their treatment, whether they were kept colored, used in greyscale, or with log translation. Also, these images were different in the walking mode, whether the image involved a single mode of walking (MS, MF, or MD) or the three walking modes in one image, which contained the MS, MF, and MD images attached next to each other, as explained in Chapter 4.

The best obtained performances appear to be in the cases of the log transformed greyscale images and involving the All Walking Conditions modes (bold font). The statistical t-test for pairwise comparison was calculated in both tables between the MD and All Walking Modes in the Log Transformed Greyscale images and a statistically significant difference was confirmed in Table 7.13 where Adam was not used. Whereas, no statistically significant difference was obtained in Table 7.14 where Adam was used.

The results obtained in this section will be submitted to the Journal of Biomedical and Health Informatics (JBHI).

Conclusion and Perspectives

General Conclusion

French Version

Les chutes chez les personnes âgées constituent un grave problème de santé qui entraîne des conséquences malheureuses et dévastatrices. Cependant, certaines chutes peuvent être évitées grâce à des interventions, une thérapie physique, une gestion appropriée et des précautions. Par conséquent, identifier le risque de chutes à l'avance peut minimiser le risque de leur occurrence et la gravité des blessures résultant de ces chutes. De même, l'identification des patients à risque fait partie intégrante de la prise en charge, car l'application de mesures préventives à cette population vulnérable peut affecter profondément la santé publique. Dans cette thèse, différents indicateurs ont été étudiées en mettant l'accent sur la cyclostationnarité pour identifier les personnes âgées à risque de chute. De plus, différentes méthodes d'apprentissage automatique ont été appliquées et comparées à cette fin.

Dans toute les publications sur la prédiction des chuteurs âgés, l'utilisation d'indicateurs classiques extraits de capteurs portables dans des algorithmes d'apprentissage automatique ont permis d'atteindre au mieux une précision d'environ 70% ; atteindre une précision supérieure à cela semblait difficile. Les travaux menés dans cette thèse portent sur la prévision des risques de chute chez les personnes âgées à l'aide de signaux de semelles de pression. Ces signaux détiennent des propriétés de cyclostationnarité ainsi que d'autres caractéristiques classiques précédemment explorées dans la littérature, ensemble semblent atteindre des performances de classification

améliorées.

Une étude bibliographique approfondie a été présentée dans cette thèse, suivie d'un traitement du signal et de l'extraction de diverses caractéristiques mettant en évidence celles qui sont liées à l'analyse cyclostationnaire. Les performances des modèles de classification se sont améliorées avec l'ajout de techniques de sélection de caractéristiques (Relief-F et SBS) et l'optimisation des hyperparamètres.

Il a également été noté que les données de la condition de marche MD (marche en décomptant à partir de 50) conduisaient à de meilleures performances de classification du modèle que la MS (marche sans tâches secondaires) et la MF (marche en appelant les noms des animaux). Cela peut être interprété comme le fait que les tâches cognitives, plus spécifiquement la mémoire de travail, aident à faire ressortir les différences dans les habitudes de marche des personnes âgées avec et sans risque de chute.

Une nouvelle méthode de représentation des propriétés cyclostationnaires des signaux de marche sous forme d'images à utiliser comme entrées d'un modèle CNN a été étudiée. Le modèle suggéré a surpassé les modèles d'apprentissage automatique conventionnels précédents avec l'optimisation Adam, a réduit la complexité avec moins d'entrées nécessaires et a permis de gagner du temps en utilisant une seule condition de marche (MD) au lieu de trois. De plus, de simples transformations logarithmiques et en échelle de gris ont entraîné une amélioration supplémentaire des performances de classification du modèle.

English Version

Elderly falls are a severe health problem that leads to unfortunate and devastating consequences. However, some falls are preventable through interventions, physical therapy, proper management, and precaution. Therefore, identifying the risk of falls ahead of time can minimize the risk of their occurrence and the severity of injuries

resulting from these falls. Similarly, identifying at-risk patients is an integral part of management, as applying preventive measures in this vulnerable population can profoundly affect public health. In this thesis, different features were studied with a focus on cyclostationarity to identify elderly people with fall risk. In addition, different machine learning methods were applied and compared for this purpose.

Throughout the literature on the prediction of elderly fallers, using classical features extracted from wearable sensors in machine learning algorithms were able to achieve at best accuracies of around 70%; achieving accuracies higher than that seemed challenging. The work carried out in this dissertation concerns the prediction of prospective risks of falling in the elderly community using pressure insole signals. These signals hold properties of cyclostationarity along with other classical features previously explored in the literature, together appear to achieve improved classification performances.

A thorough systematic literature review was presented in this thesis, followed by signal treatment and extraction of various features highlighting those that are related to cyclostationary analysis. The performance of the classification models improved with the addition of feature selection techniques (Relief-F and SBS) and hyperparameter optimization.

It was also noticed that data from the MD walking condition (walking while decounting from 50) led to better model classification performance than the MS (walking without secondary tasks) and MF (walking while calling out names of animals). This can be interpreted as that cognitive tasks, more specifically working memory, does help bring out differences in walking patterns of elderly people with and without the risk of falling.

A novel method of representing cyclostationary properties of walking signals as images to be used as inputs to a CNN model was investigated. The suggested model

outperformed the previous conventional machine learning models with Adam Optimization, reduced the complexity with fewer needed inputs, and saved time using one single walking condition (MD) instead of three. In addition, simple log and grey scale transformations caused further improvement in the model's classification performance.

Perspectives et travaux futurs

- Explorer des caractéristiques supplémentaires qui indiquent des risques de chutes et les incorporer dans les modèles de classification.
- Étudier d'autres caractéristiques liées à la cyclostationnarité qui sont liées au risque de chute chez les personnes âgées.
- Analyse d'autres méthodes pour la sélection des fonctionnalités et l'optimisation des hyperparamètres.
- Étudier d'autres méthodes d'optimisation des hyperparamètres de CNN.
- Recherche d'autres images représentatives de cyclostationnaire à insérer dans le CNN.
- Explorer d'autres méthodes de traitement des données déséquilibrées.
- Explorer des méthodes d'apprentissage automatique non supervisé pour utiliser les sujets non étiquetés dans l'ensemble de données.
- Obtenir le point de vue du domaine médical sur les performances obtenues du modèle et obtenir leur point de vue sur la détermination de la validité temporelle de ce diagnostic ou sur les moyens de le déterminer.
- Proposer les images de la carte thermique à utiliser dans une application médicale différente pour mesurer les progrès des patients après une thérapie physique ou simplement mesurer la détérioration de leur mode de marche avec l'âge pour évaluer les cas nécessitant une intervention médicale.

Perspectives and Future Work

— Exploring additional features that are indicative of risks of falls and incorporating them in the classification models.

— Investigating other features related to cyclostationarity that are related to falling risk in the elderly.

— Analyzing other methods for feature selection and hyperparameter optimization.

— Studying other methods for hyperparameter optimization of CNN.

— Researching other images representative of cyclostationarty to input into the CNN such as the synchronous average pattern of the signal produced by insole pressure sensors while walking.

— Exploring other methods of dealing with unbalanced data.

— Exploring Unsupervised Machine Learning methods to use the unlabeled subjects in the data-set.

— Getting the medical field's perspective on the achieved performance of the model and getting their viewpoint on determining the time validity of this diagnosis or on ways to determine that.

— Proposing the heat map images to be used in a different medical application of measuring patients' progress after physical therapy or simply measuring the deterioration of their walking pattern with age to assess cases that require medical intervention.

Bibliography

[AAIO18] E Atoyebi, S Adedayo, Nosa Idusuyi, and A Olorunnisola.
A review of artificial neural networks for biomedical applications: Trends and prospects.
2018.

[AAJM+20] Mohamed Alloghani, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J Aljaaf.
A systematic review on supervised and unsupervised machine learning algorithms for data science.
Supervised and unsupervised learning for data science, pages 3–21, 2020.

[AB21] Michael Appeadu and Bruno Bordoni.
Falls and fall prevention in the elderly.
In StatPearls [Internet]. StatPearls Publishing, 2021.

[ABRB04] J. Antoni, F. Bonnardot, A. Raad, and M. El Badaoui.
Cyclostationary modelling of rotating machine vibration signals.
Mechanical Systems and Signal Processing, 18(6):1285–1314, 2004.

[AD11] Mahmood Amiri and Katayoun Derakhshandeh.
Applied artificial neural networks: from associative memories to biomedical applications.

Edited by Kenji Suzuki, page 93, 2011.

[AEV20] Enes Ayan, Hasan Erbay, and Fatih Varçın.
Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks.
Computers and Electronics in Agriculture, 179:105809, 2020.

[AKB00] S Agatonovic-Kustrin and Rosemary Beresford.
Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research.
Journal of pharmaceutical and biomedical analysis, 22(5):717–727, 2000.

[Alj18] Nura Aljaafari.
Ichthyoplankton classification tool using Generative Adversarial Networks and transfer learning.
PhD thesis, 2018.

[AMAZ17] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi.
Understanding of a convolutional neural network.
In 2017 international conference on engineering and technology (ICET), pages 1–6. Ieee, 2017.

[ASS18] Shilpa Amarya, Kalyani Singh, and Manisha Sabharwal.
Ageing process and physiological changes.
In Gerontology. IntechOpen, 2018.

[AXH17] Jérôme Antoni, Ge Xin, and Nacer Hamzaoui.
Fast computation of the spectral correlation.
Mechanical Systems and Signal Processing, 92:248–277, 2017.

[BEB10] Frédéric Bonnardot and Mohamed El Badaoui.
Etude de la fatigue d'un coureur, de l'instrumentation à l'analyse vibratoire.
In 10ème Congrès Français d'Acoustique, 2010.

[BEBR+05] Frédéric Bonnardot, Mohamed El Badaoui, RB Randall, J Daniere, and François Guillet.
Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation).
Mechanical Systems and Signal Processing, 19(4):766–785, 2005.

[Ben58] W. R. Bennett.
Statistics of regenerative digital transmission.
Bell System Technical Journal, 37(6):1501–1542, 1958.

[Ben12] Yoshua Bengio.
Deep learning of representations for unsupervised and transfer learning.
In Proceedings of ICML workshop on unsupervised and transfer learning, pages 17–36. JMLR Workshop and Conference Proceedings, 2012.

[BG83] R. Boyles and W. Gardner.
Cycloergodic properties of discrete- parameter nonstationary stochastic processes.
IEEE Transactions on Information Theory, 29(1):105–114, 1983.

[BJVKM11] Ivan Bautmans, Bart Jansen, Bart Van Keymolen, and Tony Mets.
Reliability and clinical correlates of 3d-accelerometry based gait analysis outcomes according to age and fall-risk.
Gait & posture, 33(3):366–372, 2011.

[BK59] Richard Bellman and Robert Kalaba.
A mathematical theory of adaptive control processes.
Proceedings of the National Academy of Sciences, 45(8):1288–1290, 1959.

[BNB+21] Reem Brome, Jad Nasreddine, Frédéric Bonnardot, Mohamad O Diab, and Mohamed El Badaoui.
Prediction of elderly falls using the degree of cyclostationarity of walk pressure signals.

In 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), pages 477–482. IEEE, 2021.

[BND+21] Reem Brome, Jad Nasreddine, Mohamad O Diab, Mohamed El Badaoui, et al.

Identifying elderly patients at risk of falling using time-domain and cyclostationarity related features.

International Journal of Integrated Engineering, 13(5):57–66, 2021.

[Boh06] Richard W Bohannon.

Reference values for the timed up and go test: a descriptive meta-analysis.

Journal of geriatric physical therapy, 29(2):64–68, 2006.

[CKdSH+11] Benoit Caby, Suzanne Kieffer, Marie de Saint Hubert, Gerald Cremer, and Benoit Macq.

Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry.

Biomedical engineering online, 10(1):1–19, 2011.

[CL90] M. Charbit and J. Lavergnat.

Eléments de théorie du signal: signaux aléatoires, volume 91.

Ellipses Paris, 1990.

[CRS+14] Paolo Colagiorgio, Fausto Romano, Francesca Sardi, Marco Moraschini, Arianna Sozzi, Maurizio Bejor, Giovanni Ricevuti, Angelo Buizza, and Stefano Ramat.

Affordable, automatic quantitative fall risk assessment based on clinical balance scales and kinect data.

In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 3500–3503. IEEE, 2014.

[CSL17] A. Choklati, K. Sabri, and M. Lahlimi.

Cyclic analysis of phonocardiogram signals.

International Journal of Image, Graphics & Signal Processing, 9(10):1–11, 2017.

[CTPC13] Yu-Hsiu Chu, Pei-Fang Tang, Ya-Chi Peng, and Hui-Ya Chen.
Meta-analysis of type and complexity of a secondary task during walking on the prediction of elderly falls.
Geriatrics & gerontology international, 13(2):289–297, 2013.

[DLB+10] Silvia Deandrea, Ersilia Lucenteforte, Francesca Bravi, Roberto Foschi, Carlo La Vecchia, and Eva Negri.
Risk factors for falls in community-dwelling older people:" a systematic review and meta-analysis".
Epidemiology, pages 658–668, 2010.

[DWF+13] Emer P Doheny, Cathal Walsh, Timothy Foran, Barry R Greene, Chie Wei Fan, Clodagh Cunningham, and Rose Anne Kenny.
Falls classification using tri-axial accelerometers during the five-times-sit-to-stand test.
Gait & posture, 38(4):1021–1025, 2013.

[EA13] OS Eluyode and Dipo Theophilus Akomolafe.
Comparative study of biological and artificial neural networks.
European Journal of Applied Engineering and Scientific Research, 2(1):36–46, 2013.

[FJ91] C. J. Finelli and J. M. Jenkins.
A cyclostationary least mean squares algorithm for discrimination of ventricular tachycardia from sinus rhythm.
In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 13, pages 740–741, 1991.

[FL01] Stephen E Fienberg and Nicole Lazar.

William sealy gosset.

In Statisticians of the Centuries, pages 312–317. Springer, 2001.

[Fra69] L. E. Franks.

Signal theory.

Prentice-Hall, Englewood Cliffs, NJ, 1969.

[Gad98] Monica L Gaddis.

Statistical methodology: Iv. analysis of variance, analysis of co variance, and multivariate analysis of variance.

Academic emergency medicine, 5(3):258–265, 1998.

[Gar86] W. A. Gardner.

Introduction to Random Processes.

Macmillan USA, 1986.

[Gar88] W. A. Gardner.

Statistical Spectral Analysis: A Non-Probabilistic Theory.

Prentice Hall, 1988.

[Gar91] W. A. Gardner.

Two alternative philosophies for estimation of the parameters of time-series.

IEEE Transactions on Information Theory, 37(1):216–218, 1991.

[Gar94] W. A. Gardner.

Cyclostationarity in communications and signal processing.

Technical report, Statistical Signal Processing INC Yountville CA, 1994.

[GDW+12] Barry R Greene, Emer P Doheny, Cathal Walsh, Clodagh Cunningham, Lisa Crosby, and Rose A Kenny.

Evaluation of falls risk in community-dwelling older adults using body-worn sensors.

Gerontology, 58(5):472–480, 2012.

[GF75] W. Gardner and L. Franks.
Characterization of cyclostationary random signal processes.
IEEE Transactions on Information Theory, 21(1):4–14, 1975.

[GFG+14] Matthias Gietzelt, Florian Feldwieser, Mehmet Gövercin, Elisabeth Steinhagen-Thiessen, and Michael Marschollek.
A prospective field study for sensor-based identification of fall risk in older people with dementia.
Informatics for health and social care, 39(3-4):249–261, 2014.

[Gla61] E. G. Gladyshev.
Periodically correlated random sequences.
In Doklady Akademii Nauk, volume 137, pages 1026–1029. Russian Academy of Sciences, 1961.

[Gla63] E. G. Gladyshev.
Periodically and almost-periodically correlated random processes with a continuous time parameter.
Theory of Probability & Its Applications, 8(2):173–177, 1963.

[GLF+08] Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber.
A novel connectionist system for unconstrained handwriting recognition.
IEEE transactions on pattern analysis and machine intelligence, 31(5):855–868, 2008.

[GMC14] Barry R Greene, Denise McGrath, and Brian Caulfield.
A comparison of cross-sectional and prospective algorithms for falls risk assessment.
In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 4527–4530. IEEE, 2014.

[GMW+12] Barry R Greene, Denise McGrath, Lorcan Walsh, Emer P Doheny, David

McKeown, Chiara Garattini, Clodagh Cunningham, Lisa Crosby, Brian Caulfield, and Rose A Kenny.

Quantitative falls risk estimation through multi-sensor assessment of standing balance.

Physiological measurement, 33(12):2049, 2012.

[GNP06] W. A. Gardner, A. Napolitano, and L. Paura.

Cyclostationarity: Half a century of research.

Signal Processing, 86(4):639–697, 2006.

[GRC16] Barry R Greene, Stephen J Redmond, and Brian Caulfield.

Fall risk assessment through automatic combination of clinical fall risk factors and body-worn sensor data.

IEEE journal of biomedical and health informatics, 21(3):725–731, 2016.

[GRG+12] Lesley D Gillespie, M Clare Robertson, William J Gillespie, Catherine Sherrington, Simon Gates, Lindy Clemson, and Sarah E Lamb.

Interventions for preventing falls in older people living in the community.

Cochrane database of systematic reviews, (9), 2012.

[Gud59] L. I. Gudzenko.

On periodic nonstationary processes.

Radiotekhnika i elektronika, 4(6):1062–1064, 1959.

[GV12] Dhiren Ghosh and Andrew Vogt.

Outliers: An evaluation of methodologies.

In Joint statistical meetings, volume 2012, 2012.

[GVRP10] N Ganesan, K Venkatesh, MA Rama, and A Malathi Palani.

Application of neural networks in diagnosing cancer disease using demographic data.

International Journal of Computer Applications, 1(26):76–85, 2010.

[HJLS13] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant.

Applied logistic regression, volume 398.

John Wiley & Sons, 2013.

[HKL13] Jennifer Howcroft, Jonathan Kofman, and Edward D Lemaire.

Review of fall risk assessment in geriatric populations using inertial sensors.

Journal of neuroengineering and rehabilitation, 10(1):1–12, 2013.

[HLK16] Jennifer Howcroft, Edward D Lemaire, and Jonathan Kofman.

Wearable-sensor-based classification models of faller status in older adults.

PLoS one, 11(4):e0153240, 2016.

[HLK18] Jennifer Howcroft, Edward D Lemaire, and Jonathan Kofman.

Prospective elderly fall prediction by older-adult fall-risk modeling with feature selection.

Biomedical Signal Processing and Control, 43:320–328, 2018.

[HLKM17] Jennifer Howcroft, Edward D Lemaire, Jonathan Kofman, and William E McIlroy.

Elderly fall risk prediction using static posturography.

PLoS one, 12(2):e0172398, 2017.

[HOT+13] Soichiro Hirata, Rei Ono, Kota Tsutsumimoto, Shogo Misu, Hiroshi Ando, et al.

The harmonic ratio of trunk acceleration predicts falling among older people: results of a 1-year prospective study.

Journal of neuroengineering and rehabilitation, 10(1):1–6, 2013.

[Hua11] Jui-Chen Huang.

Telecare adoption model based on artificial neural networks.

Edited by Kenji Suzuki, page 343, 2011.

[Hur70] H. L. Hurd.

An investigation of periodically correlated stochastic processes.

PhD thesis, Duke University, 1970.

[JIN19] Imran Khan Mohd Jais, Amelia Ritahani Ismail, and Syed Qamrun Nisa.
Adam optimization algorithm for wide and deep neural network.
Knowledge Engineering and Data Science, 2(1):41–46, 2019.

[JLDN19] Yesmina Jaafra, Jean Luc Laurent, Aline Deruyver, and Mohamed Saber Naceur.
Reinforcement learning for neural architecture search: A review.
Image and Vision Computing, 89:57–66, 2019.

[JS15] Nathalie Japkowicz and Mohak Shah.
Performance evaluation in machine learning.
In Machine Learning in Radiation Oncology, pages 41–56. Springer, 2015.

[JSM+15] K Lyders Johansen, R Derby Stistrup, J Madsen, C Skibdal Schjøtt, and A Vinther.
The timed up and go test and 30 second chair-stand test are reliable for hospitalized patients with stroke.
Physiotherapy, 101:e918, 2015.

[JTB+11] Bart Jansen, Maxine Tan, Ivan Bautmans, Bart Van Keymolen, Tony Mets, and Rudi Deklerck.
Accelerometer based gait analysis.
In Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, pages 138–143, 2011.

[Kad19] Ammar Ismael Kadhim.
Survey on supervised machine learning techniques for automatic text classification.
Artificial Intelligence Review, 52(1):273–292, 2019.

[KAM13] Khadijeh Karamzadeh, Hadis Askarifar, and Hamid Moharrami.
Biomedical engineering via computational intelligence.

International Journal of Advancements in Research and Technology, 2(7):47, 2013.

[Kas09] Susan Kasseroler.
the stratify prediction tool has limited accuracy for predicting falls in hospital and geriatric rehabilitation inpatientscommentary.
Evidence-Based Nursing, 12(3):91–91, 2009.

[KB14] Diederik P Kingma and Jimmy Ba.
Adam: A method for stochastic optimization.
arXiv preprint arXiv:1412.6980, 2014.

[KBK11] Andrej Krenker, Janez Bešter, and Andrej Kos.
Introduction to the artificial neural networks.
Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, pages 1–18, 2011.

[KCP15] Boon-Chong Kwok, Ross A Clark, and Yong-Hao Pua.
Novel use of the wii balance board to prospectively predict falls in community-dwelling older adults.
Clinical biomechanics, 30(5):481–484, 2015.

[KMS+14] B Amir H Kargar, Ali Mollahosseini, Taylor Struemph, Wilson Pace, Rodney D Nielsen, and Mohammad H Mahoor.
Automatic measurement of physical mobility in get-up-and-go test using kinect sensor.
In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 3492–3495. IEEE, 2014.

[KR92a] Kenji Kira and Larry A Rendell.
A practical approach to feature selection.
In Machine learning proceedings 1992, pages 249–256. Elsevier, 1992.

[KR+92b] Kenji Kira, Larry A Rendell, et al.

The feature selection problem: Traditional methods and a new algorithm. In Aaai, volume 2, pages 129–134, 1992.

[KRSP96] Igor Kononenko, Marko Robnik-Sikonja, and Uros Pompe.
Relieff for estimation and discretization of attributes in classification, regression, and ilp problems.
Artificial intelligence: methodology, systems, applications, pages 31–40, 1996.

[KRSS01] Magdalena Klapper-Rybicka, Nicol N Schraudolph, and Jürgen Schmidhuber.
Unsupervised learning in recurrent neural networks.
In Proceedings of the International Conference on Artificial Neural Networks (ICANN 2001), 2001.

[KS08] Carl Kingsford and Steven L Salzberg.
What are decision trees?
Nature biotechnology, 26(9):1011–1013, 2008.

[LH07] Hyun Ja Lim and Raymond G Hoffmann.
Study design: the basics.
Topics in Biostatistics, pages 1–17, 2007.

[LHBS95] R. Lund, H. Hurd, P. Bloomfield, and R. Smith.
Climatological time series with periodic correlation.
Journal of Climate, 8(11):2787–2809, 1995.

[LLY+21] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou.
A survey of convolutional neural networks: analysis, applications, and prospects.
IEEE transactions on neural networks and learning systems, 2021.

[LRNL11] Ying Liu, Stephen J Redmond, Michael R Narayanan, and Nigel H Lovell.

Classification between non-multiple fallers and multiple fallers using a triaxial accelerometry-based system.

In 2011 annual international conference of the IEEE engineering in medicine and biology society, pages 1499–1502. IEEE, 2011.

[LRNP18] CA Lima, NA Ricci, EC Nogueira, and Mônica Rodriguez Perracini.

The berg balance scale as a clinical screening tool to predict fall risk in older adults: a systematic review.

Physiotherapy, 104(4):383–394, 2018.

[LW15] Xiangang Li and Xihong Wu.

Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition.

In 2015 ieee international conference on acoustics, speech and signal processing (icassp), pages 4520–4524. IEEE, 2015.

[LWYY16] Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang.

Large-margin softmax loss for convolutional neural networks.

arXiv preprint arXiv:1612.02295, 2016.

[MHB+12] Anat Mirelman, Talia Herman, Marina Brozgol, Moran Dorfman, Elliot Sprecher, Avraham Schweiger, Nir Giladi, and Jeffrey M Hausdorff.

Executive function and falls in older adults: new findings from a five-year prospective study link fall risk to cognition.

PloS one, 7(6):e40297, 2012.

[MM17] Lainie Van Voast Moncada and L Glen Mire.

Preventing falls in older persons.

American family physician, 96(4):240–247, 2017.

[MRW+11] Michael Marschollek, A Rehwald, KH Wolf, M Gietzelt, G Nemitz, H Meyer Zu Schwabedissen, and R Haux.

Sensor-based fall risk assessment–an expert 'to go'.

Methods of information in medicine, 50(05):420–426, 2011.

[MSC+15] David Moher, Larissa Shamseer, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, and Lesley A Stewart.
Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement.
Systematic reviews, 4(1):1–9, 2015.

[Mur16] John Murphy.
An overview of convolutional neural network architectures for deep learning.
Microway Inc, pages 1–22, 2016.

[Ned63] J. Nedoma.
Uber die ergodizitat and r-ergozitat stationarer wahrscheinlichkeitsmasse.
Zeitschrift fur Wahrscheinlichkeitstheorie, 2(1):90–97, 1963.

[NJT01] Richi Nayak, LC Jain, and BKH Ting.
Artificial neural networks in biomedical engineering: a review.
Computational Mechanics–New Frontiers for the New Millennium, pages 887–892, 2001.

[NMMR18] NH Diyana Nordin, Asan GA Muthalif, and M Khusyaie M Razali.
Control of transtibial prosthetic limb with magnetorheological fluid damper by using a fuzzy pid controller.
Journal of Low Frequency Noise, Vibration and Active Control, 37(4):1067–1078, 2018.

[NNFM14] Andrew Ng, Jiquan Ngiam, Chuan Yu Foo, and Y Mai.
Deep learning.
CS229 Lecture Notes, pages 1–30, 2014.

[Nob06] William S Noble.
What is a support vector machine?

Nature biotechnology, 24(12):1565–1567, 2006.

[NPC+04] Sudip Nandy, Suzanne Parsons, Colin Cryer, Martin Underwood, Elaine Rashbrook, Yvonne Carter, Sandra Eldridge, Jacqueline Close, Dawn Skelton, Stephanie Taylor, et al.
Development and preliminary examination of the predictive validity of the falls risk assessment tool (frat) for use in primary care.
Journal of public health, 26(2):138–143, 2004.

[NYG+21] Byungjoo Noh, Changhong Youm, Eunkyoung Goh, Myeounggon Lee, Hwayoung Park, Hyojeong Jeon, and Oh Yoen Kim.
Xgboost based machine learning approach to predict the risk of fall in older adults using gait outcomes.
Scientific reports, 11(1):1–9, 2021.

[NYU+13] Shu Nishiguchi, Minoru Yamada, Kazuki Uemura, Tetsuya Matsumura, Masaki Takahashi, Toshiki Moriguchi, and Tomoki Aoyama.
A novel infrared laser device that measures multilateral parameters of stepping performance for assessment of all risk in elderly individuals.
Aging clinical and experimental research, 25(3):311–316, 2013.

[OAU08] World Health Organization, World Health Organization. Ageing, and Life Course Unit.
WHO global report on falls prevention in older age.
World Health Organization, 2008.

[OBS+97] David Oliver, M Britton, P Seed, FC Martin, and AH Hopper.
Development and evaluation of evidence based risk assessment tool (stratify) to predict which elderly inpatients will fall: case-control and cohort studies.
Bmj, 315(7115):1049–1053, 1997.

[OPG+08] David Oliver, Alexandra Papaioannou, Lora Giangregorio, Lehana Thabane, Katerina Reizgys, and Gary Foster.
A systematic review and meta-analysis of studies using the stratify tool for prediction of falls in hospital patients: how well does it work?
Age and ageing, 37(6):621–627, 2008.

[Pet09] Leif E Peterson.
K-nearest neighbor.
Scholarpedia, 4(2):1883, 2009.

[PFKLO22] Nicholas Pudjihartono, Tayaza Fadason, Andreas W Kempa-Liehr, and Justin M O'Sullivan.
A review of feature selection methods for machine learning-based disease risk prediction.
Frontiers in Bioinformatics, 2:927312, 2022.

[PHL11] Kade Paterson, Keith Hill, and Noel Lythgo.
Stride dynamics, gait variability and prospective falls risk in active community dwelling older women.
Gait & posture, 33(2):251–255, 2011.

[PNG+01] Karen L Perell, Audrey Nelson, Ronald L Goldman, Stephen L Luther, Nicole Prieto-Lewis, and Laurence Z Rubenstein.
Fall risk assessment measures: an analytic review.
The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 56(12):M761–M766, 2001.

[PP79] E. Parzen and M. Pagano.
An approach to modeling seasonally stationary time series.
Journal of Econometrics, 9(1-2):137–153, 1979.

[PR91] Diane Podsiadlo and Sandra Richardson.

The timed “up & go”: a test of basic functional mobility for frail elderly persons.

Journal of the American geriatrics Society, 39(2):142–148, 1991.

[PS20] Derek A Pisner and David M Schnyer.

Support vector machine.

In Machine learning, pages 101–121. Elsevier, 2020.

[RHPC00] Michel Raîche, Réjean Hébert, François Prince, and Hélène Corriveau.

Screening older adults at risk of falling with the tinetti balance scale.

The Lancet, 356(9234):1001–1002, 2000.

[RJBB21] Venous Roshdibenam, Gerald J Jogerst, Nicholas R Butler, and Stephen Baek.

Machine learning prediction of fall risk in older adults using timed up and go test kinematics.

Sensors, 21(10):3481, 2021.

[RLJ17] Ramesh Rajagopalan, Irene Litvan, and Tzyy-Ping Jung.

Fall prediction and prevention systems: recent trends, challenges, and future research directions.

Sensors, 17(11):2509, 2017.

[RLS13] Gabriele Rescio, Alessandro Leone, and Pietro Siciliano.

Supervised expert system for wearable mems accelerometer-based fall detector.

Journal of Sensors, 2013, 2013.

[RTP+13] F Riva, MJP Toebes, MAGM Pijnappels, R Stagni, and JH Van Dieën.

Estimating fall risk with inertial sensors using gait stability measures that do not require step detection.

Gait & posture, 38(2):170–174, 2013.

[Rub06] Laurence Z Rubenstein.

Falls in older people: epidemiology, risk factors and strategies for prevention.

Age and ageing, 35(suppl_2):ii37–ii41, 2006.

[Rud16] Sebastian Ruder.

An overview of gradient descent optimization algorithms.

arXiv preprint arXiv:1609.04747, 2016.

[Rux06] Graeme D Ruxton.

The unequal variance t-test is an underused alternative to student's t-test and the mann–whitney u test.

Behavioral Ecology, 17(4):688–690, 2006.

[Sah18] Sumit Saha.

A comprehensive guide to convolutional neural networks the eli5 way, 2018.

[SCBW00] Anne Shumway-Cook, Sandy Brauer, and Marjorie Woollacott.

Predicting the probability for falls in community-dwelling older adults using the timed up & go test.

Physical therapy, 80(9):896–903, 2000.

[SCD13] Jatinderpal Sharma, Shikha Chawla, and Sunil Dalhotra.

A research agenda on artificial neural network topologies & data mining in neural network.

International Journal of Data & Network Security, 1:41–47, 2013.

[SD19] BH Shekar and Guesh Dagnew.

Grid search-based hyperparameter tuning and classification of microarray cancer data.

In 2019 second international conference on advanced computational and communication paradigms (ICACCP), pages 1–8. IEEE, 2019.

[SdBUM10] Jaap Swanenburg, Eling D de Bruin, Daniel Uebelhart, and Theo Mulder.

Falls prediction in elderly people: a 1-year prospective study.
Gait & posture, 31(3):317–321, 2010.

[SDH+13] George M Savva, Orna A Donoghue, Frances Horgan, Claire O'Regan, Hilary Cronin, and Rose Anne Kenny.
Using timed up-and-go to identify frail members of the older population.
Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 68(4):441–446, 2013.

[SEBG+10] Khalid Sabri, Mohamed El Badaoui, François Guillet, Alain Belli, Guillaume Millet, and Jean Benoit Morin.
Cyclostationary modeling of ground reaction force signals.
Signal Processing, 90(4):1146–1152, 2010.

[SG64] Abraham Savitzky and Marcel JE Golay.
Smoothing and differentiation of data by simplified least squares procedures.
Analytical chemistry, 36(8):1627–1639, 1964.

[SG94] Chad M Spooner and William A Gardner.
The cumulant theory of cyclostationary time-series. ii. development and applications.
IEEE Transactions on Signal Processing, 42(12):3409–3429, 1994.

[SH12] Allen M Spiegel and Meredith Hawkins.
'personalized medicine'to identify genetic risks for type 2 diabetes and focus prevention: can it fulfill its promise?
Health Affairs, 31(1):43–49, 2012.

[Sim99] Haykin Simon.
Neural networks: a comprehensive foundation.
Prentice hall, 1999.

[SKS+21] LT Southerland, AD Kloos, L Slattery, Y Tan, G Young, J Rosenthal, and DA Kegelmeyer.
Accuracy of the 4-stage balance test and sensor-based trunk sway as fall risk assessment tools in the emergency department.
Journal of Acute Care Physical Therapy, 12(2):79–87, 2021.

[SMT+17] Joana Silva, João Madureira, Cláudia Tonelo, Daniela Baltazar, Catarina Silva, Anabela Martins, Carlos Alcobia, and Inês Sousa.
Comparing machine learning approaches for fall risk assessment.
In International Conference on Bio-inspired Systems and Signal Processing, volume 5, pages 223–230. SCITEPRESS, 2017.

[Spr14] Vincent Spruyt.
The curse of dimensionality in classification.
Computer vision for dummies, 21(3):35–40, 2014.

[SPSG05] E. Serpedin, F. Panduru, I. Sarı, and G. B. Giannakis.
Bibliography on cyclostationarity.
Signal Processing, 85(12):2233–2303, 2005.

[SRA+] Jared W Skinner, Ryan Roemmich, Shinichi Amano, Elizabeth Stegemoller, Lori Altmann, and Chris J Hass.
Frequency domain analysis of ground reaction force does not differentiate between hyperkinetic and hypokinetic movement disorders.
jskinner1@ ufl. edu. University of Florida Gainesville FL USA.

[SRNL11] Tal Shany, Stephen J Redmond, Michael R Narayanan, and Nigel H Lovell.
Sensors-based wearable systems for monitoring of human movement and falls.
IEEE Sensors journal, 12(3):658–670, 2011.

[SSB14] Hasim Sak, Andrew W Senior, and Françoise Beaufays.

Long short-term memory recurrent neural network architectures for large scale acoustic modeling.

2014.

[SSG+12] R Senden, HHCM Savelberg, B Grimm, IC Heyligers, and K Meijer.

Accelerometry-based gait analysis, an additional objective approach to screen subjects at risk for falling.

Gait & posture, 36(2):296–300, 2012.

[Suz11] Kenji Suzuki.

Pixel-based artificial neural networks in computer-aided diagnosis.

Artificial Neural Networks-Methodological Advances and Biomedical Applications, pages 71–92, 2011.

[SV21] Patrick Schober and Thomas R Vetter.

Logistic regression in medical research.

Anesthesia and analgesia, 132(2):365, 2021.

[SW+89] Lars St, Svante Wold, et al.

Analysis of variance (anova).

Chemometrics and intelligent laboratory systems, 6(4):259–272, 1989.

[SWL+15] Tal Shany, Kejia Wang, Ying Liu, Nigel H Lovell, and Stephen J Redmond.

Are we stumbling in our quest to find the best predictor? over-optimism in sensor-based models for predicting falls in older adults.

Healthcare technology letters, 2(4):79–88, 2015.

[TEBS15] CV Toulouse, M El Badaoui, and C Serviere.

Gait analysis and falls estimation using the slope's variation of synchronous statistics.

In 2015 International Conference on Advances in Biomedical Engineering (ICABME), pages 293–296. IEEE, 2015.

[TLS10] Anne Tiedemann, Stephen R Lord, and Catherine Sherrington.

The development and validation of a brief performance-based fall risk assessment tool for use in primary care.

Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences, 65(8):896–903, 2010.

[TPS+01] AM Tromp, SMF Pluijm, JH Smit, DJH Deeg, LM Bouter, and PTAM Lips.

Fall-risk screening test: a prospective study on predictors for falls in community-dwelling elderly.

Journal of clinical epidemiology, 54(8):837–844, 2001.

[Tuk93] John W Tukey.

Exploratory data analysis: past, present and future.

Technical report, PRINCETON UNIV NJ DEPT OF STATISTICS, 1993.

[UMG11] Craig A Umscheid, David J Margolis, and Craig E Grossman.

Key concepts of clinical trials: a narrative review.

Postgraduate medicine, 123(5):194–204, 2011.

[Vau09] Christopher Leonard Kit Vaughan.

The biomechanics of human locomotion.

2009.

[VSD15] K Vembandasamy, R Sasipriya, and E Deepa.

Heart diseases detection using naive bayes algorithm.

International Journal of Innovative Science, Engineering & Technology, 2(9):441–444, 2015.

[WHP+11] Alexander Weiss, Talia Herman, Meir Plotnik, Marina Brozgol, N Giladi, and JM Hausdorff.

An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers.

Physiological measurement, 32(12):2003, 2011.

[Wil14] RL Williams.
Engineering biomechanics of human motion.
NotesBook Supplement for ME 4670/BME, 5670, 2014.

[YAN+11] Minoru Yamada, Tomoki Aoyama, Masatoshi Nakamura, Buichi Tanaka, Koutatsu Nagai, Noriatsu Tatematsu, Kazuki Uemura, Takashi Nakamura, Tadao Tsuboyama, and Noriaki Ichihashi.
The reliability and preliminary validity of game-based fall risk assessment in community-dwelling older adults.
Geriatric Nursing, 32(3):188–194, 2011.

[YS20] Li Yang and Abdallah Shami.
On hyperparameter optimization of machine learning algorithms: Theory and practice.
Neurocomputing, 415:295–316, 2020.

[ZEBM+13] FA Zakaria, M El-Badaoui, S Maiz, F Guillet, M Khalil, K Khalil, and M Halimi.
Walking analysis: Empirical relation between kurtosis and degree of cyclostationarity.
In 2013 2nd International Conference on Advances in Biomedical Engineering, pages 93–96. IEEE, 2013.

[Zel94] Andreas Zell.
Simulation neuronaler netze, volume 1.
Addison-Wesley Bonn, 1994.

[ŽG91] Goran D Živanović and William A Gardner.
Degrees of cyclostationarity and their application to signal detection and estimation.
Signal processing, 22(3):287–297, 1991.

[Zho21] Zhi-Hua Zhou.

Machine learning.

Springer Nature, 2021.

[ZPST18] Yu-Dong Zhang, Chichun Pan, Junding Sun, and Chaosheng Tang.

Multiple sclerosis identification by convolutional neural network with dropout and parametric relu.

Journal of computational science, 28:1–10, 2018.

[ZTEB+14] Firas A Zakaria, Claude-Vivien Toulouse, Mohamed El Badaoui, Christine Serviere, and Mohamad Khalil.

Contribution of the cyclic correlation in gait analysis: Variation between fallers and non-fallers.

In 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom), pages 176–181. IEEE, 2014.

Caractérisation et analyse des signaux de pression cyclostationnaire générés lors de la marche : prédiction des chutes chez les personnes âgées

Résumé : Il existe un intérêt croissant pour le développement de modèles de prédiction du risque de chute chez les personnes âgées qui peuvent être utilisés comme approche préventive pour prédire le risque futur de chute dans la communauté des personnes âgées.

Les principaux objectifs de ce travail sont, premièrement, d'étudier les aspects cyclostationnaires des signaux de pression de la semelle intérieure des personnes âgées. Deuxièmement, pour extraire les caractéristiques essentielles indicatives du risque de chutes futures. Troisièmement, mettre en œuvre et comparer différentes méthodes d'apprentissage automatique supervisé pour classer les sujets âgés en sujets avec ou sans risque de chute dans le futur.

L'ensemble de données se compose de signaux de pression collectés à partir des semelles intérieures de 519 personnes âgées qui ont indiqué si elles avaient déjà fait des chutes. La première étape de cette étude a consisté en une revue approfondie de la littérature des travaux antérieurs partageant des objectifs principaux similaires. La section suivante explique l'aspect de la cyclostationnarité et sa caractérisation dans les signaux de pression de semelle étudiés. Après cela, l'ensemble de données est analysé statistiquement pour extraire des caractéristiques utiles. Ensuite, notre étude propose les caractéristiques indicatives des chutes futures, les modèles d'apprentissage automatique et les méthodes d'optimisation pour développer l'évaluation du risque de chute dans la communauté des personnes âgées.

Enfin, notre étude propose une nouvelle méthode pour représenter les signaux de la semelle intérieure de pression cylcostationnaire et les utiliser dans un modèle

d'apprentissage en profondeur pour prédire les chutes potentielles dans la communauté des personnes âgées.

Mots-clés : parcimonie, cyclostationnarité, déconvolution, approximation parcimonieuse, diagnostic, défaut de roulement, phonocardiogramme.

Characterization and Analysis of the Cyclostationary Pressure Signals Generated during Walking: Predicting Falls for the Elderly

Abstract : There is an increasing interest in developing older adult fall-risk prediction models that can be used as a preventive approach to predicting the future risk of falling in the elderly community.

The primary objectives of this thesis are, first, to study the cyclostationary aspects of the pressure insole signals of older adults. Second, to extract essential features indicative of the risk of future falls. Third, to implement and compare different supervised machine-learning methods to classify elderly subjects as subjects with or without risk of falling in the future.

The data set consists of pressure signals collected from the innersoles of 519 elderly people who reported whether they had experienced previous falls. The first stage of this study consisted of a thorough literature review of prior work sharing similar main objectives. The following section explains the aspect of cyclostationarity and its characterization in the studied pressure insole signals. After that, the data set is analyzed statistically to extract useful features. Then our study proposes the features indicative of future falls, the machine learning models, and optimization methods to develop fall risk assessment in the elderly community.

Finally, our study proposes a novel method for representing cylcostationary pressure insole signals and using them in a deep learning model to predict prospective falls in the elderly community.

Keywords : sparsity, cyclostationarity, deconvolution, sparse approximation, diagnostic, bearing failure, phonocardiogram.