Universidade do Minho
Escola de Engenharia

José António Fernandes Braga

Machine learning for differential diagnosis of white matter lesions in Fabry Disease patients based on gait and cardiac characteristics

Master's Thesis

Integrated Master's in Industrial Electronics and Computers Engineering

Work elaborated under supervision of:

Prof. Dr. Estela Bicho (Supervisor)

Dr. Flora Ferreira (Supervisor)

June 2021

Acknowledgments

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my master's degree!

The Algoritmi Center at Universidade do Minho, especially to my supervisors Estela Bicho Erlhagen and Flora Ferreira, whose insight and knowledge into the subject matter steered me through this research. And special thanks to Carlos Fernandes, whose support as part of his research allowed my studies to go the extra mile. I would also like to thank Doctor Miguel Gago and Doctor Olga Azevedo for all the help regarding the medical issues that came through the development of this dissertation.

And my biggest thanks to my parents and friends for all the support you have shown me through this research, the culmination of five years of learning, especially to my friend Coelho who has helped me across the years. And, last but not least, for my girlfriend Claudia, thanks for all your support, without which I would have stopped these studies a long time ago. I can only say thank you to everyone who has supported me and had to put up with my stresses and moans for the past five years of study and sorry for being even grumpier than normal whilst I wrote this thesis!

STATEMENT OF INTEGRITY

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

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

Resumo

Machine learning para o diagnóstico de lesões na massa branca em pacientes com Doença de Fabry baseado na marcha e em características cardíacas

As manifestações cerebrais da doença de Fabry (FD) incluem lesões na matéria branca (WMLs). O objetivo deste estudo é identificar quais as caracteristicas da marcha e cardíacas que permitem diferenciar os pacientes com FD com WMLS de pacientes com FD sem WMLs.

Para o estudo da marcha, foram avaliados 76 sujeitos através de sensores vestíveis. Os valores da série temporal das 16 variáveis da marcha foram normalizados usando modelos de regressão múltipla. Usando as 32 medidas de marcha (média e variabilidade), foi aplicado um algoritmo de feature selection seguido de cinco classificadores diferentes (LR, SVM Linear e kernel RBF, RF e KNN). Os algoritmos CNN e LSTM foram implementados utilizando como input os conjuntos de séries temporais da marcha. Para pacientes com FD e com WMLs vs controlos, a maior exatidão de 71,50% foi obtida usando RF. Para pacientes com FD e sem WMLs vs controlos, o melhor desempenho foi observado usando KNN com uma exatidão de 86,67%. Para pacientes com FD com vs sem WMLs, os melhores modelos foram obtidos usando o algoritmo CNN e usando LR com base com uma exatidão de 81,43% e 80,76%, respetivamente.

Em relação aos dados cardíacos, foram utilizados os dados de dois exames: o eletrocardiograma (ECG) e o ecocardiograma. Um total de 114 pacientes com FD (61 deste com WMLs) foram avaliados com o exame de ECG. Para pacientes com FD com vs sem WMLs, a maior exatidão foi de 79,72%. Com o uso simultâneo de marcha e ECG, dois modelos foram avaliados com um grupo de teste de nove pacientes. O melhor resultado foi a exatidão de 80% com o algoritmo LR. Finalmente, análises de regressão logística também foram realizadas nas 22 características do ecocardiograma de 93 pacientes com FD (49 com WMLs). Os resultados confirmaram que a idade está significativamente associada à presença de WMLs.

Esses resultados demonstram o potencial das técnicas de machine learning baseadas na marcha e nas características cardíacas para entender o papel dos WMLs em pacientes com FD.

Palavras-chave: Doença de Fabry, Machine Learning, Lesões da matéria Branca, Marcha, Exames cardíacos.

Abstract

Machine learning for differential diagnosis of white matter lesions in Fabry Disease patients based on gait and cardiac characteristics

Brain manifestations in FD include progressive white matter lesions (WMLs).This research aims to identify a set of gait and cardiac characteristics to discriminate FD patients with WMLs from FD patients without WMLs.

For the gait study, 76 subjects walked through a predefined circuit using wearable sensors that continuously acquired different stride features. All strides of 16 gait variables were normalized using multiple regression models. The mean and the variability of each gait time series were calculated, resulting in 32 gait measures. Using the 32 gait measures, a feature selection algorithm were applied. Then, five different classifiers (LR, SVM Linear and RBF kernel, RF, and KNN) based on different selected set features were evaluated. CNN and LSTM algorithms were implemented using as input the gait time series. For FD patients with WMLs vs controls the highest accuracy of 71.50% was obtained using RF. For FD patients without WMLs vs controls, the best performance was observed using KNN with an accuracy of 86.67%. For FD patients with vs without WMLs the best models were obtained using the CNN algorithm and using LR algorithm with an accuracy of 81.43% and 80.76%, respectively.

Regarding the cardiac data, was used the data from two exams: an electrocardiogram (ECG) and an echocardiogram. A total of 114 FD patients were evaluated with the ECG (61 have WMLs). For FD patients with vs without WMLs the highest accuracy of 79.72%. With the simultaneously use of both gait and ECG features, two models were evaluated on a test group of nine patients. The best result was 80% accuracy with LR. Finally, logistic regression analyses were also performed on the 22 features of the echocardiogram of ninety three FD patients (forty-nine with WMLs). The results confirmed that age is significantly associated with the presence of WMLs.

These findings are the first step to demonstrate the potential of machine learning techniques based on gait and cardiac characteristics as a complementary tool to understand the role of WMLs in FD patients.

Keywords: Fabry Disease, Machine Learning, White Matter Lesions, Gait, Cardiac exams.

Contents

Acknowledgments iii

Resumo v

Abstract vi

List of Figures **x**

List of Tables **xiii**

List of acronyms xvi

I : Introduction and State of the art **20**

1 Introduction **21**

1.1 Contextualization and Motivation 21

1.2 Objectives 22

1.3 Overview of the Thesis 23

1.4 Contributions of this thesis 24

2 State of Art **26**

2.1 Fabry Disease 26

2.2 Machine Learning for the diagnosis in healthcare 29

II : Materials and methodology **31**

3 Materials **32**

3.1 Gait assessment 32

3.2 Electrocardiogram 34

3.3 Echocardiogram 35

4 Methodology **38**

4.1 Multiple Regression Normalization 38

4.2 Statistical tests 39

4.3 Machine Learning 40

4.4 Feature selection 42

4.5 Models theoretical basis 43

4.6 Machine Learning metrics 56

III : Results, Discussion, Conclusion, and Future Work **57**

5 Gait assessment **58**

5.1 Participants 58

5.2 Gait Normalization 59

5.3 FD patients with WMLs vs controls 62

5.4 FD patients without WMLs vs controls 68

5.5 FD patients with vs without WMLs 75

5.6 Discussion 82

6 Electrocardiogram **85**

6.1 Participants and Holter data 85

6.2 Feature selection with RFE 90

6.3 Electrocardiogram classification 93

6.4 Gait + Electrocardiogram analysis 94

6.5 Discussion 96

7 Echocardiogram **98**

7.1 Participants 98

7.2 Discussion 103

8 Conclusion, limitations, and future work **104**

8.1 General conclusions 104

8.2 Limitations 105

8.3 Future work 106

IV : Appendix **107**

List of Figures

1 Overview of the research steps of this thesis. 24

2 Frequency of erroneous diagnoses in patients with FD enrolled in FOS – the Fabry Outcome Survey (adapted from (Atul Mehta, Michael Beck & Sunder-Plassmann, 2006)). . . 28

3 Traditional approach to data analysis project (adapted from (Géron, 2019)). 40

4 Machine learning approach to data analysis project (adapted from (Géron, 2019)). . . . 41

5 CNN and LSTM flowchart. 55

6 Example of the mean and variability (SD) for patient N167 and his gait variables: cycle duration, cadence, stance, swing, loading, foot flat, pushing and double support. 60

7 Example of a gait time series for patient N167 and his gait variables: cycle duration, cadence, stance, swing, loading, foot flat, pushing and double support. 61

8 Comparison between the mean value of gait features in FD patients with WMLs vs controls. Data are shown for the raw gait data and the MR normalized gait. Max.: maximum; Var.: variability; Min.: minimum; H.C. : heel clearance; T.C. : toe clearance 63

9 Gait correlation heat map in FD patients with WMLs vs controls. Max.: maximum; Var.: variability; Min.: minimum. 64

10 F1 score vs number of features for selection of optimal numbers of gait characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients with WMLs vs controls. Recursive feature elimination was used through the 5-fold cross-validation (RFECV). Max.: maximum; Var.: variability; Min.: minimum. 66

11 Top 3 and Top 4 after performing RFE on gait features for FD patients with WMLs vs controls. 67

12 Comparison between the mean value of gait features in FD patients without WMLs vs controls. Data are shown for the raw gait data and the MR normalized gait. Max.: maximum; Var.: variability; Min.: minimum; H.C. : heel clearance; T.C. : toe clearance . 70

13 Gait correlation heat map in FD patients without WMLs vs controls. Max.: maximum; Var.: variability; Min.: minimum. 71

14 F1 score vs number of features for selection of optimal numbers of gait characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients without WMLs vs controls. Recursive feature elimination was used through the 5-fold cross-validation (RFECV). Max.: maximum; Var.: variability; Min.: minimum. 72

15 Top 3, Top 5, and Top 3 (LR & SVM) after performing RFE on gait features for FD patients without WMLs vs controls. 73

16 Comparison between the mean value of gait features in FD patients with vs without WMLs. Data are shown for the raw gait data and the MR normalized gait. Max.: maximum; Var.: variability; Min.: minimum; H.C. : heel clearance; T.C. : toe clearance . . . 77

17 Gait correlation heat map in FD patients with vs without WMLs. Max.: maximum; Var.: variability; Min.: minimum. 78

18 F1 score vs number of features for selection of optimal numbers of gait characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients with vs without WMLs. Recursive features elimination was used through the 5-fold cross-validation (RFECV). Max.: maximum; Var.: variability; Min.: minimum. 79

19 Top 3 and Top 5 after performing RFE on gait features for FD patients with vs without WMLs. 80

20 Boxplot of age stratified for the presence of WMLs in FD patients. 86

21 Boxplot of age stratified for the presence of WMLs in FD patients with ages between 40 and 59 years old, inclusive. 88

22 ECG correlation heat map in FD patients. 91

23 F1 score vs number of features for selection of optimal numbers of ECG characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients with vs without WMLs. Recursive features elimination was used through the 5-fold cross-validation (RFECV). Max.: maximum; Min.: minimum. 92

24 Box plot of FD patients with vs without WMLs according to age. 99

25 Box plot of patients with vs without WMLs in the interval of ages between 40 and 59 years, inclusive. 102

List of Tables

1 Gait variables. 33

2 ECG features. 34

3 ECG normal reference range of values (Rijnbeek et al., 2014; Umetani et al., 1998). . . 35

4 Echocardiogram features. 36

5 Echocardiogram normal reference range of values (Caballero et al., 2015; El Missiri et al., 2016). 37

6 Demographic variables for FD patients with and without WMLs and Controls. 59

7 Variance inflation factor for physical characteristics (age, weight, height, and sex), speed, and stride length. 59

8 Multiple linear regression models using only significative independent variables for each gait variable on the right foot. 60

9 Demographic variables for FD patients with WMLs and Controls. 62

10 Results for each of 3 machine learning algorithms (Random Forest, SVM with linear kernel, and Logistic Regression) with recursive feature elimination for FD patients with WMLs vs controls. 65

11 Classification accuracy on training and validation data for top common gait characteristics in FD patients with WMLs vs controls with LR, SVM (linear and RBF kernel), RF, and KNN. 67

12 CNN and LSTM results on selected gait features from the gait time series for FD patients with WMLs vs controls. 68

13 Demographic variables for FD patients without WMLs and Controls. 68

14 Results for each of 3 machine learning algorithms (LR, SVM with linear kernel, and RF) with optimal gait characteristics' from recursive feature elimination for FD patients without WML vs controls. 73
15 Classification accuracy on training and validation data for top common gait characteristics in FD patients without WMLs vs controls with LR, SVM (linear and RBF kernel), RF, and KNN. 74
16 CNN and LSTM results on selected gait features from the gait time series for FD patients without WMLs vs controls. 75
17 Demographic variables for FD patients with and without WMLs. 75
18 Results for each of 3 machine learning algorithms (Random Forest, SVM with linear kernel, and Logistic Regression) with recursive feature elimination for FD patients with vs without WMLs. 80
19 Classification accuracy on training and validation data for top common gait characteristics in FD patients with vs without WMLs with LR, SVM (linear and RBF kernel), RF, and KNN. 81
20 CNN and LSTM results on selected gait features from the gait time series for FD patients with vs without WMLs. 82

21 Demographic characteristics for 114 FD patients with and without WMLs. 85
22 ECG features for 114 FD patients with and without WMLs. 86
23 Univariate and multivariate (adjusted for age) logistic regression model to predict the presence of WMLs. 87
24 Number of FD patients with and without WMLs per age classes 88
25 Demographic characteristics for patients between 40 and 59 years old, inclusive. 88
26 Univariate and multivariate (adjusted for age or sex) logistic regression model to predict the presence of WMLs for patients between 40 and 59 year old, inclusive. 89
27 Association of heart rate variability SDANN5 and SDNN with WMLs in FD patients age ranged from 40 to 59 years. 90

28 Results for each of 3 machine learning algorithms (LR, SVM with linear kernel and RF) with optimal ECG characteristics' from recursive feature elimination for FD patients with vs without WMLs. 91
29 Classification accuracy on training and validation data for top common ECG characteristics and SDANN 5 in FD patients with vs without WMLs with LR, SVM (linear and RBF kernel), RF and KNN. 94
30 Classification accuracy on training and validation data plus testing accuracy for group of 9 patients that were assessed on both gait and ECG with gait's best model (Top LR), ECG's best model (Top RF), Gait ECG best model combined with LR and RF. 95
31 Gait and ECG plus gait + ECG best models prediction for 9 patients selected for test. . . 95

32 Demographic characteristics for 93 FD patients with and without WMLs. 98
33 Echocardiogram features for 93 FD patients with and without WMLs. 100
34 Univariate and multivariate (adjusted for age) logistic regression model to predict the presence of WMLs. 101
35 Division of patients according to respective ages plus age significance value for each age interval performed with Mann Whitney U Test. 101
36 Echocardiogram features presented as mean $\pm$ standard deviation, statistically significant after performing univariate logistic regression and adjusted logistic regression for age to predict the presence of WMLs for patients between 40 and 59 years, inclusive. . 102

37 Value of gait assessment of 17 gait variables for 39 FD patients with vs without WMLs, 25 FD patients with WMLs vs controls and 14 FD patients without WMLs vs controls. . . 108
38 Multiple linear regression models using only significative independent variables for each gait variable on the left foot. 109
39 Classification accuracy on training and validation data plus testing accuracy for a group of 9 patients that were assessed on both gait and ECG with Gait ECG best model combined with SVM Linear Kernel, SVM RBF Kernel, and KNN. 109

List of acronyms

- **α-gal A** - α-galactosidase A
- **AIC** - Akaike's Information Criterion
- **CI** - Confidence Interval
- **CNN** - Convolutionary Neural Networks
- **ERT** - Enzyme replacement therapy
- **ECG** - Electrocardiogram
- **FD** - Fabry Disease
- **Gb3** - Globotiaosylceramide
- **GNB** - Gaussian Naive Bayes
- **KNN** - K-Nearest Neighbor
- **LR** - Logistic Regression
- **LSTM** - Long-Short Term Memory
- **ML** - Machine Learning
- **MR** - Multiple Regression
- **NRR** - Artificial Neural Network activated with ReLU function
- **RF** - Random Forest
- **RFE** - Recursive Feature Elimination
- **SD** - Standard deviation

- **SVM** - Support Vector Machine
- **VIF** - Variance Inflation Factor
- **WMLs** - White Matter Lesions

Part I :

Introduction and State of the art

Chapter 1

Introduction

1.1 Contextualization and Motivation

Fabry Disease (FD) is a rare disease that greatly affects the quality of life and may lead to premature death. This disease is an X-linked lysosomal storage disorder caused by the deficiency or absent activity of the enzyme α-Galactosidase A (α-Gal A). This disease also affects several organs, including the kidney, heart, and brain. Brain manifestations in FD include progressive white matter lesions (WMLs) (Buechner et al., 2008; Körver et al., 2018a). White matter is the brain region responsible for the transmission of nerve signals and for communication between different parts of the brain. Brain WMLs were an early manifestation, affecting 11.1% of males and 26.9% of females under 30 years of age, even without cerebrovascular risk factor outside FD (Azevedo et al., 2020). In (Körver et al., 2018a), WMLs were found in 46% of 1276 patients which tend to occur earlier in males and their prevalence revealed to increase with patients' age. WMLs have been associated with gait impairment (Starr, 2003) and the risk of falls (Snir et al., 2019; Zheng et al., 2012). Gait abnormalities, such as slower gait and postural instability, have been reported in FD (Löhle et al., 2015). Furthermore, gait assessment has been recently revealed as a good complementary clinical tool to discriminate FD patients from healthy adults, however, not much is known regarding the impact of WMLs on gait performance in patients with FD. Specific location and distribution of WMLs suggest a specific underlying disease (Körver et al., 2018a) and may reflate in a

different gait profile. Furthermore, heart abnormalities have been associated with the presence of WMLs (Forte et al., 2019). However, it remains unclear whether cardiac biomarkers are associated with WMLs in FD patients. It is then of extreme importance to explore different biomarkers that can assist in the differential diagnosis of WMLs in FD patients.

Machine Learning (ML) is coming into its own, with a growing recognition that it can play a key role in a wide range of critical applications. ML provides potential solutions in several domains and is set to be a pillar of our future civilization. ML has been used in different industries, one of these being the healthcare industry. Machine learning is a growing field that has been extremely used because of the excellent results it can achieve when applied to different classification and pattern recognition tasks in the medical sector (see e.g. (Maity & Das, 2017)).

The potential of ML methods based on gait, electrocardiogram, and echocardiogram has not yet been explored for the task of assisting in the differential diagnosis of WMLs in FD patients. This may be due to the fact that FD is a rare disease and data to develop different ML methods is not easily available. It happens that in the district of Guimarães, Portugal, there is an unusually high number of FD patients. This leads to the possibility of collecting a substantially high quantity of data about this disease and puts us in a unique position to study different patterns and biomarkers for the differential diagnosis of WMLs in FD patients using ML.

1.2 Objectives

This dissertation aims to investigate different machine learning algorithms based on different datasets (gait, electrocardiogram, and echocardiogram) to develop a clinical decision support tool that enhances the differential diagnosis of WMLs in FD patients. More precisely, the objectives of this thesis are presented below:

- To evaluate the effectiveness of machine learning methods to discriminate:
  - FD patients with WMLs from FD patients without WMLs, and these two groups from healthy controls based on gait data;

  - FD patients with WMLs from FD patients without WMLs based on electrocardiogram data;
  - FD patients with WMLs from FD patients without WMLs based on echocardiogram.

- Investigate how the integration of more than one dataset (gait, electrocardiogram, or echocardiogram) can improve the predictive system for the differential diagnosis of WMLs in FD patients.

1.3 Overview of the Thesis

Figure 1 presents an overview of the steps taken to achieve the stated objectives. Four different datasets were used: clinical and physical data, gait data, electrocardiogram data, and echocardiogram data. The first step consisted of analyzing each dataset to understand the influence of each feature: their significance (statistical tests) and their correlations. The effects of inter-subject variations in each dataset due to physical characteristics were analyzed. To reduce these effects normalization of gait data was performed, and electrocardiogram and echocardiogram datasets were subdivided. Then, different feature selection techniques were used to select the best subset of features of each dataset. Finally, with the subsets created by the feature selection from one or more datasets, different classification machine learning algorithms were evaluated. The results of the performance evaluation of the models provide prior knowledge to develop a supervised predictive system to work as a support for the early diagnostic of the presence or absence of WMLs in FD patients.

This dissertation is composed of four parts and a total of eight chapters. The first part contains two chapters, introduction, and state of the art. Introduction talks about the general motivation and contextualization, the objectives, an overview of this thesis, and the scientific papers developed in the scope of this thesis. The last chapter of Part I is the state of art and it introduces the theoretical fundamentals of the concepts addressed in the dissertation and the work developed in the areas related to the dissertation. Part II describes the materials and methodology. The first chapter of Part II relates how the data gets collected and the inside of each dataset to be used further in the dissertation. Second chapter reports the methodology used across the development of this dissertation. Following is Part III with results, discussion, conclusion, and future work. The initial chapter of Part III contains the results of the data analysis on the

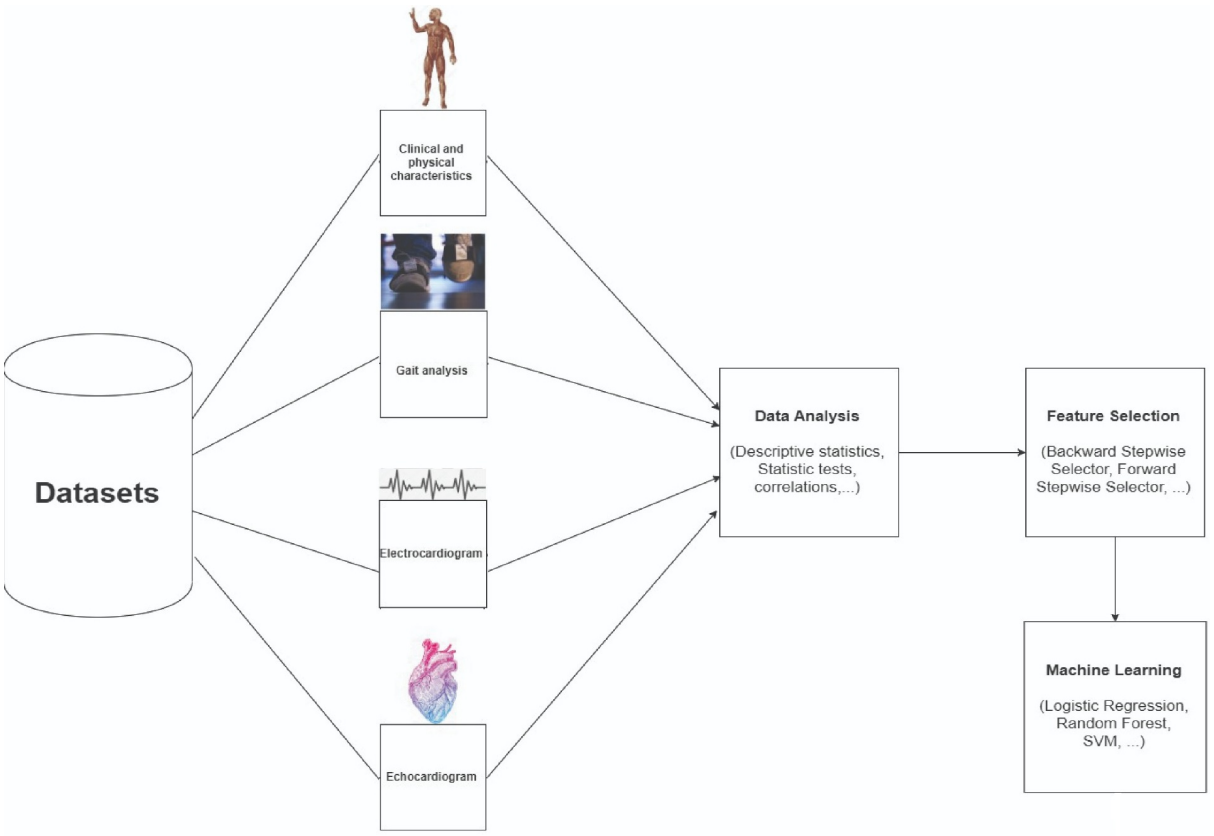

Figure 1: Overview of the research steps of this thesis.

information provided from the information in the third chapter plus a discussion for each of the different types of dataset. The next chapter concludes this thesis with a scope for future work. Finally, Part IV is the appendix, where is some extra information regarding the results.

1.4 Contributions of this thesis

The following publications have been made based on the work developed in this dissertation:

- **José Braga**, Flora Ferreira, Carlos Fernandes, Miguel F. Gago, Olga Azevedo, Nuno Sousa, Wolfram Erlhagen, Estela Bicho . "Gait characteristics and their discriminative ability in patients with fabry disease with and without white-matter lesions". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

Bioinformatics). Vol. 12251 LNCS. Springer Science and Business Media Deutschland GmbH, July 2020, pp 415-428. ISBN: 978303058876. DOI: 10.1007/978-3-030-58808-330.

- Flora Ferreira, Carlos Fernandes, **José Braga**, Estela Bicho, Wolfram Erlhagen, Miguel F. Gago."The effect of levodopa medication on stride time variability in patients with Parkinsonism". In: *Journal of Statistics on Health Decision* 2.2 (Oct. 2020), pp 63-64. DOI: 10.34624/jshd.v2i2.20109

Chapter 2

State of Art

2.1 Fabry Disease

Fabry Disease (FD) is a rare genetic X-linked lysosomal disorder caused by the deficiency or absent activity of the enzyme α-galactosidase A ($alpha$-gal A), which is responsible for destroying a type of fat called globotiaosylceramide (Gb3 or GL-3). When these fat molecules start accumulating due to the lack of $alpha$-gal A, they start causing damage to the cells. FD has a wide variety of signs and symptoms, that goes from heart attacks and strokes to kidney diseases (Giugliani et al., 2016).

With the proper care and treatment, patients with this disease can live longer and have a good life quality (Desnick et al., 2003).

2.1.1 Signs and Symptoms of FD

As said before, this disease can be related to many signs and symptoms, that can vary from person to person. Several organs, including the kidney, heart, and brain may be affected (Giugliani et al., 2016).

Brain manifestations in FD include progressive white matter lesions (WMLs) (Buechner et al., 2008; Körver et al., 2018a). In (Körver et al., 2018a), WMLs were found in 46% of 1276 patients which tend to occur earlier in males and their prevalence revealed to increase with patients' age. White matter is the

brain region responsible for the transmission of nerve signals and for communication between different parts of the brain. WMLs have been associated with gait impairment (Starr, 2003) and the risk of falls (Snir et al., 2019; Zheng et al., 2012). Gait abnormalities, such as slower gait and postural instability, have been reported in FD (Löhle et al., 2015). However, not much is known regarding the impact of WMLs on gait performance in patients with FD. Specific location and distribution of WMLs suggest a specific underlying disease (Körver et al., 2018a) and may reflect in a different gait profile.

2.1.2 FD diagnosis

"The timely diagnosis of FD is difficult" (Grünfeld, 2003).

It is easy to understand that this disease has a hard diagnose, although since it is a genetic disease, analysing the family history can be very helpful. FD results in the accumulation of a fat molecule in cells, with the years the symptoms have a higher percentage to start appearing, although symptoms have been reported in children with two years old (Ramaswami et al., 2006). In the absence of a family member who has already received a diagnosis of the disorder, many cases are not diagnosed until adulthood (average age, 29 years) (Morgan et al., 1987), when the pathology of the disorder may already be advanced. Pain, skin rashes, heat, intolerance, stomach upsets, fatigue, lack of energy, and the inability to exercise are generally the first signs and symptoms to appear, but because these can be associated with other conditions, it may take many years for a diagnosis of Fabry disease to be made. In fact, up to 25% of patients are misdiagnosed (Atul Mehta, Michael Beck & Sunder-Plassmann, 2006). Kidney, heart, and brain problems tend to become noticeable between the ages of 30 to 45 and it is at this point that many individuals with FD are first diagnosed (MacDermot, KD and Holmes & Miners, 2001).

What is particularly concerning about FD is that there is an average delay between the onset of symptoms and diagnosis of 12 years. This is the same for both sexes, although the onset of symptoms tends to occur about six years later in females than males (MacDermot, KD and Holmes & Miners, 2001).

Since this disease has many symptoms equals to symptoms related to other most commons diseases, the diagnosis becomes harder. In figure 2, we can see the frequency of erroneous diagnoses in patients with FD enrolled in FOS – the Fabry Outcome Survey.

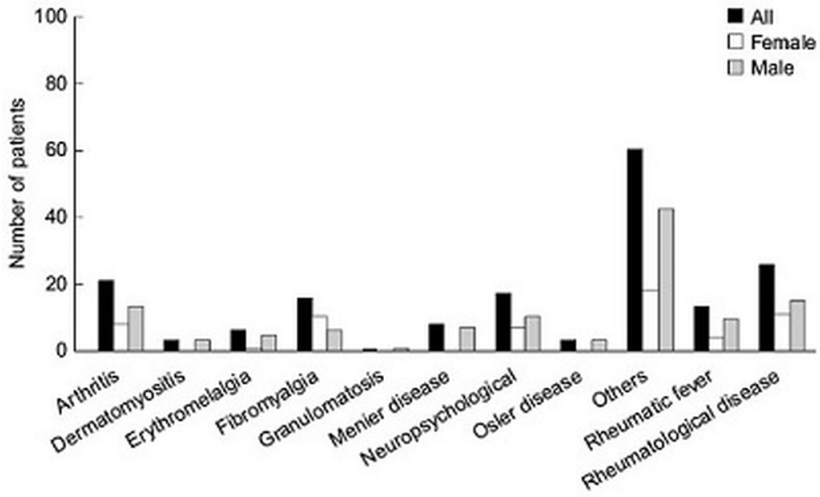

Figure 2: Frequency of erroneous diagnoses in patients with FD enrolled in FOS – the Fabry Outcome Survey (adapted from (Atul Mehta, Michael Beck & Sunder-Plassmann, 2006)).

Due to the complexity of the disorder, the FD diagnosis is often delayed. This is a reflection of the wide range of specialists required to correctly identify this disease. To reduce this diagnose some measures can be helpful like screening programs (Sunder-Plassmann & Födinger, 2006).

Another important measure in the FD diagnosis is to perform early diagnosis in family members of FD patients. It has been shown that the lifespan of a man with FD is reduced by approximately 20 years and by approximately 15 years for a woman (MacDermot, KD and Holmes & Miners, 2001).

2.1.3 FD treatment

If the diagnosis of FD goes positive, it is necessary to start a treatment. This treatment can go some different ways like - the most common - intravenous enzyme replacement therapy (ERT) (this treatment consists of replacing an enzyme which is either absent or deficient in patients, in this case, α-galactosidase A, Oral Chaperone Therapy, conventional medical treatment, therapy, and prophylactic medications (Mauer & Kopp, 2018)).

Available therapies for FD include enzyme replacement therapy (ERT) and the chaperone migalastat. There are still issues unsolved about ERT despite being heavily researched. Both FD specific therapies present limitations and the attempt to correct the enzymatic deficiency, are not able to fully revert FD

pathology and clinical manifestations. Therefore, several new therapies are under research, including new forms of ERT, substrate reduction therapy, mRNA therapy, and gene therapy (Azevedo et al., 2020).

Focusing on ERT, the commercially approved to treat FD may have some variances according to the area where it is being administrated: may variate from agalsidase beta to agalsidase alfa (Mauer & Kopp, 2018). The producers write on their website "The lowering of GL-3 suggests that Fabrazyme may improve how FD affects your body; however a relationship of lower GL-3 to specific signs and symptoms of FD has not been proven" (Desnick et al., 2003).

With the availability of enzyme replacement therapy, prompt diagnosis and treatment of FD have assumed new importance (Desnick et al., 2003).

2.2 Machine Learning for the diagnosis in healthcare

With the usage of all the data generated by medical exams, diagnoses, and treatments and focusing on the study of different biomarkers, it is possible to use machine learning algorithms to assist in the perception and diagnose of multiple diseases such as Rheumatoid Arthritis, Cancer, Lung Diseases, Heart Diseases, Diabetic Retinopathy, Hepatitis, Alzheimer, Liver, Dengue, Parkinson, etc (Joseph, 2017; Myszczynska et al., 2020).

Gait evaluation can be useful to differentiate different pathologies even in the presence of highly overlapping phenotypes, such as the differences found between two types of Parkinsonism (Vascular Parkinsonism vs Idiopathic Parkinson's Disease) in (Fernandes et al., 2021). Furthermore, gait assessment has been revealing as a good complementary clinical tool to discriminate adults with and without a pathology such as Parkinson's disease (Fernandes et al., 2019; Ferreira et al., 2019; Kubota et al., 2016), Huntington's disease (Mannini et al., 2016) and recently FD (Fernandes et al., 2020). Gait is usually described by its spatio-temporal and foot clearance characteristics such as speed, stride length, stride time, minimum toe clearance (mean gait characteristics), and their respective variability (given by the coefficient of variation or the standard deviation) (Ferreira et al., 2019; Kubota et al., 2016; Rehman et al., 2019). Different machine learning (ML) techniques have been used to select the best combination of relevant gait

characteristics for gait classification (Caramia et al., 2018; Fernandes et al., 2019; Rehman et al., 2019). Recent work (Rehman et al., 2019) shows that a subset of gait characteristics selected using random forest with information gain and recursive features elimination (RFE) technique with Support Vector Machine (SVM) and Logistic Regression improves the classification accuracy of Parkinson's Disease. Widely used machine learning models for gait classification are SVM (Aich et al., 2018; Fernandes et al., 2020; Mannini et al., 2016; Pradhan et al., 2015; Wahid et al., 2015), Random Forest (RF) (Aich et al., 2018; Fernandes et al., 2020; Wahid et al., 2015) and K-Nearest Neighbor (KNN) (Pradhan et al., 2015; Wahid et al., 2015). The outcomes of these studies show good performance accuracy in the classification of pathological gait. In particular, the outcomes of previous work (Fernandes et al., 2020) show promising results in the use of gait characteristics to discriminate FD patients and healthy adults. However, the implication of the presence or absence of WMLs in gait performance has not yet been investigated.

Functional measurements of the heart obtained by cardiac exams such as the electrocardiogram (ECG) and/or echocardiogram has been used for diagnosis and monitoring of patients with heart diseases (G. Jignesh Chowdary, Suganya. G, 2020; Galluzzi et al., 2009; Hijazi et al., 2016; Tsai et al., 2019), including FD patients (Satriano et al., 2020). The application of different ML techniques have been shown to improve risk prediction in various cardiovascular diseases. As an example, the model proposed recently in (G. Jignesh Chowdary, Suganya. G, 2020) which uses the voting of Logistic Regression(LR), Random Forest(RF), Artificial Neural Network activated with ReLU function(NNR), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes(GNB) based on ECG data reveals good performance (an accuracy of 89%) to predict the possibility of heart disease. In fact, ML techniques can learn the hypotheses available within the ECGs, which are then used to provide predictions regarding patients' health (Hijazi et al., 2016). In particular, some functional measurements observed in an ECG exam were proved through regression analysis to affect the presence or absence of WMLs (Galluzzi et al., 2009).

Extracting knowledge from medical datasets is quite difficult and time-consuming when performed manually since these datasets generally contain a large number of records. The use of ML techniques in these datasets is therefore recommendable for quick and accurate extraction of useful information that could support doctors in their work (Faust & Ng, 2016; Prakash et al., 2018).

Part II :

Materials and methodology

Chapter 3

Materials

The data to be used in the further studies are the outcome of gait assessment in healthy subjects (controls) and FD patients, and, electrocardiogram and echocardiogram performed in FD patients.

3.1 Gait assessment

Gait analysis is the study of walking exercises. Two Physilog® sensors (Gait Up®, Switzerland) positioned on both feet were used to measure different gait variables of each stride. This study consists of walking for 30m straight meters, turning around, and walking another 30m straight meters while the sensors are acquiring the data. This data consists of the arithmetic mean calculated for all subjects' stride time series for 13 spatial-temporal variables and 4 foot clearance variables in a total of 17 gait variables that contain the full step data, represented in table 1.

To perform the gait assessment the participants were asked to walking a 60-meter continuous course (30 meters corridor with one turn) in a self-selected walking while two Physilog® sensors (Gait Up®, Switzerland) positioned on both feet were used to measure different gait variables of each stride (also known as gait cycle). The Physilog® sensor is equipped with a high-quality 3D accelerometer, a 3D gyroscope, a barometric pressure sensor, and the capability of recording the data on an SD card.

This data consists of 11 spatial-temporal variables and 6 foot clearance variables in a total of 17 gait

variables described in Table 1.

Table 1: Gait variables.

Spatial-temporal variables	
Speed	Velocity of one cycle
Cycle duration	Duration of one gait cycle
Cadence	Number of gait cycles in a minute
Stride length	Distance between successive initial ground contact of the same foot
Stance	The time during which the foot is in the ground
Swing	The time during which the foot is in the air
Loading	Percent of stance between the heel strike and the foot is fully on the ground
Foot flat	Percent of stance where the foot is fully at on the ground
Pushing	Percent of stance between the foot being fully on the ground and the toe leaving the ground
Double support	Percent of gait cycle where both feet touch the ground
Peak swing	Maximum angular velocity during swing
Foot clearance variables	
Strike angle	Angle between the foot and the ground when the heel hits the ground
Lift-off angle	Angle between the foot and the ground when the toes are leaving the ground
Maximum heel clearance	Maximum height above the ground reached by the heel
Maximum toe clearance 1	Maximum height above the ground reached by the toes after heel max clearance
Minimum toe clearance	Minimum height of the toes during swing phase
Maximum toe clearance 2	Maximum height above the ground reached by the toes just before heel strike

Since the point here is to evaluate accurate forward gait performances and, due to the fact that the turn causes severe variations that can affect the final outcome, the turn will be discarded for further analysis.

3.2 Electrocardiogram

The data provided by the electrocardiogram (ECG) consists of 15 features that analyzed heart rate data and variability, QT (total time for the ventricles of the heart to depolarize and repolarize (contract and relax)) episodes, and supraventricular ectopy. All features are described in Table 2.

Table 2: ECG features.

Heart rate data	
HR Min	Minimum heart rate
HR Mean	Mean heart rate
HR Max	Maximum heart rate
Heart rate variability	
ASDNN 5	Average standard deviation of all 5 mins normal R-R (heartbeat) intervals
SDANN 5	Standard deviation of sequential 5 mins of normal R-R interval means
SDNN	Standard deviation of the normal R-R intervals
RMSSD	Root mean square differences of successive R-R intervals
QT analysis	
QT Min	Minimum value of the QT interval
QT Mean	Mean of the QT intervals
QT Max	Maximum value of the QT interval
QTc Min	Minimum value for the QT interval corrected for extreme heart rate
QTc Mean	Mean of the QT intervals corrected for extreme heart rate
QTc Max	Maximum value for the QT interval corrected for extreme heart rate
QTc $\geq$ 450	QT interval percentage that is greater than 450 ms corrected for extreme heart rate
Supraventricular ectopy	
Longest R-R	Maximum difference between two R-R peaks

These ECG features should now be analyzed in order to find the correlation between them and the

presence of FD. With this study, it may be possible to find a predictor among these 15 features. In Table 3, are described the normal reference range of values used to control the outcome of the ECG. These values indicate the range of values in which a patient can be considered healthy.

Table 3: ECG normal reference range of values (Rijnbeek et al., 2014; Umetani et al., 1998).

Electrocardiogram features	**Normal reference range of values**	
	Male	**Female**
Heart Rate	65 to 74	66 to 72
QT	378 to 398	390 to 400
QTc	409 to 430	418 to 432
ASDNN 5 (ms)	43 to 88	38 to 66
SDNN 5 (ms)	117 to 182	114 to 147
SDANN (ms)	104 to 162	102 to 133
RMSSD (ms)	22 to 53	22 to 43

3.3 Echocardiogram

The data provided by the echocardiogram consists of 22 features described in Table 4.

The typical reference range of values for echocardiogram features is referred to in Table 5 and can be used to compare the actual result of a patient to check for abnormality in the outcome of the echocardiogram.

Table 4: Echocardiogram features.

Echocardiogram features	
MV E/A Ratio	Mitral valve ratio between early diastole (E Wave) and atrial contraction (A Wave)
MV A Vel	Mitral valve A Wave blood flow velocity
MV Dec T	Mitral valve deceleration time
MV E Vel	Mitral valve E Wave blood flow velocity
E' Lateral	E Wave using tissue doppler imaging at lateral mitral annulus position
E' Septal	E Wave using tissue doppler imaging at septal mitral annulus position
E/E' Lateral	Ratio between E Wave and E' Wave measured at lateral mitral annulus position
E/E' Medial	Ratio between E Wave and E' Wave at medial mitral annulus position
E/E' Septal	Ratio between E Wave and E' Wave measured at septal mitral annulus position
LVPWd	Left ventricular posterior wall thickness
ISVd	Interventricular septum thickness at end-diastole
LVIDd	Left ventricular internal dimension at end-diastole
LADiam/SC	Left atrial diameter measured at subcostal position
AoDiam	Aorta diameter
S' Lateral	Peak systolic velocity using tissue doppler tissue doppler imaging at lateral mitral annulus position
LVdMassInd ASE	Left ventricular mass at end-diastole indexed to body surface area according to the American Society of Echocardiography (ASE)
LADiam	Left Atrial diameter
S' Septal	Peak systolic velocity using tissue doppler imaging at septal mitral annulus position
A' Septal	A Wave using tissue doppler imaging at septal mitral annulus position
A' Lateral	A Wave using tissue doppler imaging at lateral mitral annulus position
LVDdMass ASE	Left ventricular mass at end-diastole according to the ASE
LVIDd/SC	Left ventricular internal dimension at end-diastole at subcostal position

Table 5: Echocardiogram normal reference range of values (Caballero et al., 2015; El Missiri et al., 2016).

Echocardiogram features	Normal reference range of values
MV E/A Ratio	0.86 to 1.88
MV E Vel (cm/s)	0.61 to 0.93
MV A Vel (cm/s)	0.43 to 0.77 (must be smaller than MV E Vel)
MV Dec T (ms)	138.6 to 237.4
E' Lateral (cm/s)	9.5 to 17.5
E' Septal (cm/s)	7.3 to 13.3
E/E' Lateral	4.0 to 8.2
E/E' Medial	4.6 to 8.6
E/E' Septal	5.5 to 10.3
S' Septal (cm/s)	6.7 to 9.5
A' Septal (cm/s)	7.4 to 11.4
S' Lateral (cm/s)	7.4 to 12.2
LVPWd (mm)	7.65 to 10.07
IVSd (mm)	7.78 to 10.06
LVIDd (mm)	43.56 to 52.16
LADiam/SC (mm/m^2)	Not defined
AoDiam (mm)	23.22 to 27.88
LVdMassInd(ASE) (g/m^2)	43 to 115
LADiam (mm)	23.74 to 30.44
LVdMass(ASE) (g)	95.78 to 192.81
LVIDd/SC (mm/m^2)	Not defined

Chapter 4

Methodology

4.1 Multiple Regression Normalization

Gait characteristics of a subject are affected by his demographics properties including height, weight, age, and sex, as well as by walking speed (Wahid et al., 2016; Wahid et al., 2015) or stride length (Alcock et al., 2018; Fernandes et al., 2019; Ferreira et al., 2019). To normalize the data regression models according to Wahid et al.'s method (Wahid et al., 2016; Wahid et al., 2015) were used. Comparing to other methods, such as dimensionless equations and detrending methods, MR normalization revealed better results on reducing the interference of subject-specific physical characteristics and gait variables (Mikos et al., 2018; Wahid et al., 2016; Wahid et al., 2015), thereby improving gait classification accuracy using machine learning methods (Fernandes et al., 2019; Wahid et al., 2015).

First, to control the multicollinearity among predictor variables within this multiple regression, Variance Inflation Factor (VIF) was calculated (Thompson et al., 2017). This test measures the colinearity of the physical characteristics (age, weight, height, and sex), speed, and stride length, being a value of VIF greater than 5 an indicator of this strong correlation.

Spatio-temporal gait variables and foot clearance variables were normalized as follows:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \tag{4.1}$$

where $\hat{y}_i$ represents the prediction of the dependent variable for the ith observation; x_{ij} represents the jth physical property of the ith observation including age, weight, height, sex, speed or stride length, β_0 represents the intercept term, β_j represents the coefficient for the jth physical property and ε_i represents the residual error for the ith observation. The model's coefficients are estimated using the physical properties and the mean values of the gait variables of the healthy controls. Although at least 20 subjects per independent variable are recommended in multiple linear regression (Katz, 2011), based on similar studies (Mikos et al., 2018; Wahid et al., 2016) with higher sample sizes, MR models were computed for all combinations with 1, 2, and 3 independent variables using a bi-square weight function. For the models with all significant independent variables (p-value<0.05) Akaike's information criterion (AIC) (Burnham & Anderson, 2004) and R-squared metrics were used to select the best-fitted model. Statistical assumptions of a linear regression including linearity, normality, and homoscedasticity were verified.

In each subject group, the best fitted MR models are used to normalize each stride gait variable by dividing the original value y_i by the predicted gait variable $\hat{y}_i$ from (4.1), as follow:

$$y_i^n = \frac{y_i}{\hat{y}_i} \tag{4.2}$$

where y_i^n represents the normalized value for the ith observation.

After normalizing all strides of each of the 16 gait variables, the mean and the standard deviation (SD) of each variable (each gait time series) for all subjects were calculated. In this work, the SD value is used to measure the variability of each gait variable.

4.2 Statistical tests

Independent T tests or non-parametric Mann–Whitney U tests (if the variable is not normally distributed) were used to compare differences between two independent groups. The normality of data was determined with Shapiro–Wilk tests. Fisher Exact T test was used to compare the differences between the groups when the data is categorical.

The theory of statistical test of significance essentially involves the setting of a null hypothesis and

competing for an alternative hypothesis. The null hypothesis is a hypothesis of "no difference among groups" or " a sample came from a normally distributed population". To decide if the null hypothesis is rejected or not, the p-value was calculated. The p-value means the probability of obtaining test results at least as extreme as the results actual observed, under the assumption that the null hypothesis is corrected. In practice, the significance level (α) is stated in advance to determine how small the p-value must be in order to reject the null hypothesis. Conventionally, it is considered statistically significant as p-value < α when α=0.05 and statistically highly significant as p-value < α when α=0.001 (Kirkwood & Sterne, 2003).

4.3 Machine Learning

Machine Learning (ML) differs from typical programming because the goal here is to make computers learn from data, as said by Arthur Samuel (1959): "field of study that gives computers the ability to learn without being explicitly programmed." Since this is an era where data can be accessed pretty easily, ML algorithms can revolutionize everything.

Typically, to write a piece of code, we write all the rules necessary to make it logical as in Figure 3.

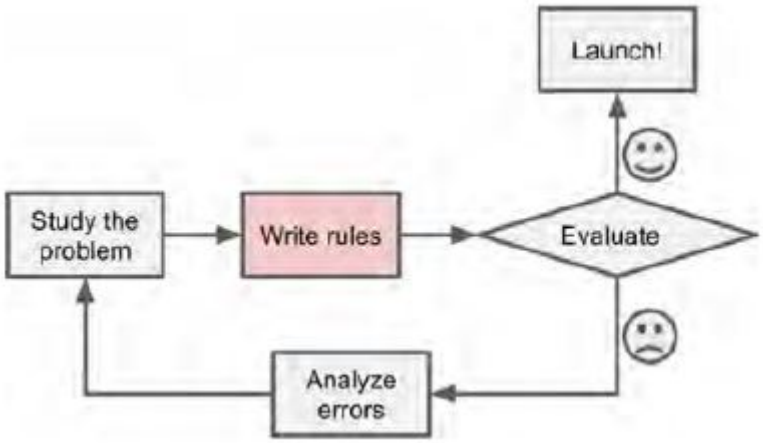

Figure 3: Traditional approach to data analysis project (adapted from (Géron, 2019)).

In a complex situation, the code can be pretty complicated and heavy. Another solution is to write a ML algorithm that can detect a pattern and learn by itself, making it much shorter, easier to maintain, and

more accurate (Figure 4).

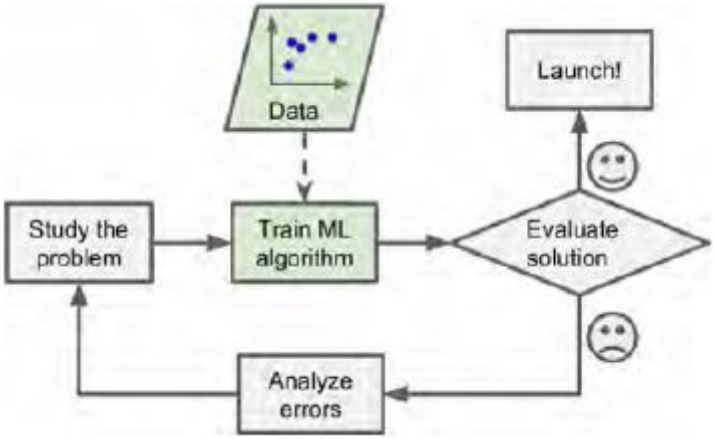

Figure 4: Machine learning approach to data analysis project (adapted from (Géron, 2019)).

There is also the advantage of finding patterns in a large amount of data, that human eyes cannot detect (data mining).

The first step is to know exactly what is the objective of the ML algorithm. This is important because it will determine how to frame the problem, what algorithms to select and what performance measure should be used to evaluate the model.

Second step consists of getting the data. Without data, ML is not possible because the computer needs "something" to learn from. Usually, it is provided a dataset (database table where every column of the table represents a specific variable (or feature), and each row corresponds to a sample) to the ML algorithm. Missing data, duplicate data, outliers, a large number of variables, or highly correlated variables are possible issues that originate at the source and that should be overcome to ensure good data. The performance of ML is greatly affected by the quality of the data. Different techniques such as missing data imputation, outlier detection, dimensionality reduction data transformations, are first applied to get data with quality for ML modeling (Gudivada et al., 2017).

After the dataset is ready, it is time to create the test, train, and validation set:

- **Training Dataset**: The sample of data used to fit the model. The actual dataset that we use to train the model. The model reads and learns from this data.

- **Validation Dataset**: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. This data is used to fine-tune the model hyperparameters hence the model never "learns" from it.

- **Test Dataset**: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. The test set provides the standard used to evaluate the model.

Next, it is important to keep the relevant features so that the training is not clouded with irrelevant information because having many features will not help to have a better outcome (exactly the opposite). The model will only be capable of learning if the training data contains enough relevant features and not too many irrelevant ones. So it becomes critical to its success coming up with a good set of features to train on. This process is called feature engineering and involves:

- **Feature selection**: selecting the most useful features to train on among existing features creating a subset of the original attributes (features or variables);

- **Feature extraction** (or Feature projection): combining existing features to produce a more useful one, reducing this way the high-dimensional spaces into a space with fewer dimensions achieved by dimension reduction algorithms;

- Creating **new features** by gathering new data.

4.4 Feature selection

It is possible to state that feature selection is an inevitable part of a classifier design. For instance, the classifier-independent univariate filter methods have similar trends. Filter methods such as the T Test, Mann Whitney U Test or Fisher Exact T Test have better or similar performance with wrapper methods for harder problems. This improved performance is usually accompanied by significant peaking (Hua et al., 2009).

In this study, a hybrid method (filter method followed by a wrapper method) was employed to select the most relevant gait and cardiac characteristics. First, a filter method based on Mann Whitney U tests and Spearman's correlation between the variables was used to selected 10 gait characteristics. The selected 10 gait or cardiac characteristics are the ones that present a higher U-value and do not present a high correlation between them ($|\rho|$<0.90).

Before applying a wrapper method, the selected 10 variables were scaled to have zero mean and unit variance. Based on previous work (Rehman et al., 2019), the wrapper was developed using the Recursive Feature Elimination (RFE) technique with three different ML classifiers: Logistic Regression (LR), SVM with Linear kernel, and RF. RFE has some advantages over other filter methods (Iguyon & Elisseeff, 2003). RFE is an iterative method where features are removed one by one without affecting the training error. The selection of the optimal number of features for each model is based on the evaluation metric F1 score (see definition in section 4.6) evaluated through 5-fold cross-validation, that consists in a procedure that divides a limited dataset into 5 non-overlapping folds, where each fold is at some point of the iteration used as a held-back test set, whilst all other folds collectively are used as a training dataset. The gait characteristics' importance was quantified using the model itself (feature importance for LR and SVM with linear kernel and information gain for RF). The F1 score was used to assess the performance of the different gait characteristics combinations.

4.5 Models theoretical basis

4.5.1 Logistic Regression

Some regressions can be used for classification. Logistic Regression is commonly used to estimate the probability that an instance belongs to a particular class. If the estimated probability is greater than 50%, then the model predicts that the belongs to that class, or else it predicts that it does not. Therefore, Logistic Regression is a binary classifier.

To compute this, Logistic Regression computes a weighted sum of input features (plus a bias term),

but instead of outputting the result directly, it outputs the logistic of this result.

The Logistic Regression model estimated probability (vectorized form) is computed by the equation:

$$\hat{p} = h_\theta(x) = \sigma(\theta^T \cdot x) \tag{4.3}$$

The logistic is a sigmoid function that outputs a number between 0 and 1. Is computed by the equation:

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{4.4}$$

Once the Logistic Regression model has estimated the probability $\hat{p}$ = h_θ(x) that an instance x belongs to the positive class, it can make its prediction $\hat{y}$ easily.

$$\hat{y} = \begin{cases} 0, & \text{if } \hat{p} < 0.5, \\ 1, & \text{if } \hat{p} \geq 0.5 \end{cases} \tag{4.5}$$

Notice that σ(t)< 0.5 when t $<$ 0, and σ(t) $\geq$ 0.5 when t $\geq$ 0, so a Logistic Regression model predicts 1 if $\theta^T \cdot$ x is non-negative, and 0 if it is negative.

Now there is only one more question, how is it trained? The objective of the training is to set the parameter vector θ so that the model estimates high probabilities for positive instances (y = 1) and low probabilities for negative instances (y = 0). This idea is captured by the cost function below for a single training instance x.

$$c(\theta) = \begin{cases} -log(\hat{p}), & \text{if } y = 1, \\ -log(1 - \hat{p}), & \text{if } y = 0 \end{cases} \tag{4.6}$$

This cost function makes sense because –log(t) grows very large when t approaches 0, so the cost will be large if the model estimates a probability close to 0 for a positive instance, and it will also be very large if the model estimates a probability close to 1 for a negative instance. On the other hand, –log(t) is close to 0 when t is close to 1, so the cost will be close to 0 if the estimated probability is close to 0 for a negative instance or close to 1 for a positive instance, which is precisely the goal.

The cost function over the whole training set is simply the average cost of overall training instances.

It can be written in a single expression called log loss:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^i \cdot log(\hat{p}^i) + (1 + y^i) \cdot log(1 - \hat{p}^i] \tag{4.7}$$

4.5.2 Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a very powerful and versatile Machine Learning model, capable of performing linear or nonlinear classification, regression, and even outlier detection (Géron, 2019). SVM's are based on the concept of a decision boundary that separates two different classes of data in order to discriminate classes. A separating hyperplane is constructed in the training phase by using an input training data set containing data samples. The hyperplane that best separates the samples belonging to the two classes is called a maximum-margin hyperplane that forms the decision boundary. The class samples that are on the boundary are called Support Vectors (SVs).

These SVs obtained from the training phase are then used in the classification phase to classify new data. The pattern x lies on the hyperplane in the feature space can be described by equation 2.9, where w is a normal vector to the hyperplane and b is the model bias:

$$x_i w^T + b = 0 \tag{4.8}$$

By selecting the two hyperplanes described in equation (4.9) and equation (4.10), the data points are separated in the margin region, and the aim is to maximize the distance between them.

$$x_i w^T + b = +1, y_i = +1 \tag{4.9}$$

$$x_i w^T + b = -1, y_i = -1 \tag{4.10}$$

The distance between the two hyperplanes is $\frac{2}{||w||}$:

$$d_+ - d_- = \frac{|1-b|}{||w||} - \frac{|-1-b|}{||w||} = \frac{2}{||w||} \tag{4.11}$$

and the goal is to minimize $||w||$ since the aim is to maximize the margin $\frac{2}{||w||}$. Also, for every $i \in (1, n), x_i$ and y_i follows the constraints:

$$x_i w^T + b \geq +1, y_i = +1 \tag{4.12}$$

$$x_i w^T + b \leq -1, y_i = -1 \tag{4.13}$$

Equations 4.12 and 4.13 are equivalent to the following equation:

$$y_i(x_i w^T + b) - 1 \geq 0, \forall i \tag{4.14}$$

The final SVM decision function is:

$$f(x) = sgn(xw^T + b) \tag{4.15}$$

Support vector classification relies on this notion of linearly separable data. However, in practice data is often very far from being linearly separable, and we need to transform the data into a higher dimensional space in order to fit a support vector classifier.

The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions and separate data by class differently from each other. These functions can be different types:

- **Linear**: $K(x, z) = (x \cdot z)$
- **Polynomial**: $K(x, z) = ((x \cdot z)) + 1)^d$
- **Gaussian Radial Basis Function (RBF)**: $K(x, z) = exp(-\frac{||x-z||^2}{2\sigma^2})$

where in Linear and Polynomial function $x \cdot z$ is the cross product.

4.5.3 Decision trees

Decision Trees are a class of very powerful Machine Learning Model. It is capable of achieving high accuracy while being highly interpretable.

As the name goes, it uses a tree-like model of decisions. The "knowledge" learned by a decision tree, through training, is directly formulated into a hierarchical structure, with leaves (nodes) and branches (splits). This structure holds and displays the knowledge in such a way that it can easily be understood, which means that is possible to fully understand where our accuracy and errors are coming from, what type of data the model would do well with, and how the output is influenced by the values of the features.

Another great advantage of Decision Tree Models is that require very little data preparation. Many Machine Learning models may require heavy data pre-processing or complex regularisation schemes. Decision trees work quite well after tweaking a few of the parameters.

Also, the cost of using the tree for inference is logarithmic in the number of data points used to train the tree. That is a huge plus since it means that having more data will not make a huge dent in the inference speed.

Decision Tree models are created using two steps: induction (where the tree is built. Is, also, set all of the hierarchical decision boundaries based on the data provided. Because of the nature of the decision trees, they can be prone to major overfitting) and pruning (the process of removing the unnecessary structure from a decision tree, effectively reducing the complexity to combat overfitting with the added bonus of making it even easier to interpret).

Induction

Determine the"best feature" in the dataset to split the data on consists in choose the features to use and the specific split, using a greedy algorithm to minimize a cost function. This algorithm consists of a trial and error method, where different choices are made until the lowest cost solution is found. Better the choices made on the tries, less wasting computations on testing out split point will exist.

For a regression tree, it can be used a simple squared error as our cost function:

$$E = \sum (y - \hat{y})^2 \tag{4.16}$$

where y is our ground truth and $(\hat{y})$ is our predicted value. The sum of all the samples in our dataset is the total error.

For a classification tree, it can be used the Gini Index Function:

$$E = \sum [p_k \cdot (1 - p_k)] \tag{4.17}$$

where p_k is the proportion of training instances of class k in a particular prediction node.

A node should ideally have an error value of zero, which means that each split outputs a single class 100% of the time. This value is the objective because once reached that particular node, the output is known whether on one side of the decision boundary or the other.

As a problem usually has a large set of features, it results in a large number of splits, which in turn gives a huge, slow, and overfitted to our training dataset. That way is necessary to implement some predefined stopping criterion to halt the construction of the tree.

The most common stopping method is to use a minimum count of the number of training examples assigned to each leaf node. That way, if the count is less than a minimum value then the split is not accepted and the node is taken as a final leaf node.

If all leaf nodes become final, the training stops. The smaller the minimum count, finer splits, and more information will be given but is also prone to overfitting on the dataset. A larger minimum count can make the stop happen too early. Doing so, the minimum value is usually set based on the dataset, depending on how many examples are expected to be in each class.

Another way is to set the maximum depth of the tree, which means the length of the longest path from a root to a leaf (Prashant Gupta, 2017).

Pruning

As previously mentioned, because of the nature of training decision trees they can be prone to major overfitting. Because of the difficulty of setting a correct value for the minimum number of instances per

node, the result ends up that many splits will be redundant and unnecessary to increase the accuracy.

Tree Pruning is a technique that leverages this splitting redundancy to remove (prune) the unnecessary splits in the tree. In other, pruning compresses part of the tree from strict and rigid decision boundaries into ones that are more smooth and generalize better. Being the complexity directly linked to the number of splits in the tree, this method can reduce effectively the tree complexity.

Pruning can start at either root or the leaves. The simplest method of pruning, called reduced error pruning, starts at leaves and removes each node with the most popular class in that leaf. This change is kept if it does not deteriorate accuracy. There are more sophisticated pruning methods, called cost complexity pruning, where a learning parameter (alpha) is used to weigh whether nodes can be removed based on the size of the sub-tree. This is also known as the weakest link pruning. Another cost complexity pruning method consists in go through each node in the tree and evaluate the effect of removing it on the cost function. If the change is not significant then it is possible to prune it.

Random Forests

Random Forests (Breiman, 2001) are based on Decision Trees. Random Forest is used for supervised learning which means learning from labeled data and making predictions based on the learned patterns.

Random Forests randomly selects observations/rows and specific features/variables to build multiple decision trees from and then averages the results while a decision tree is built on an entire dataset, using all the features/variables of interest. Random Forest also allows for controlling the number of trees built along the process.

4.5.4 K-Nearest Neighbors

Nearest neighbor classification, also known as K-nearest neighbors (KNN), is based on the idea that the nearest patterns to a target pattern x', the pattern for which the label is seeked, deliver useful label information. KNN assigns the class label of the majority of the K-nearest patterns in data space (Kramer,

2013). In $\mathbb{R}$, the Minkowski metric can be employed:

$$||x' - x_j||^p = (\sum_{i=1}^{q} |(x_i)' - (x_i)_j|^p)^{1/p} \tag{4.18}$$

which corresponds to the euclidean distance for p=2. For binary classification, KNN is defined as:

$$f_{KNN}(x') = \begin{cases} 1 & \text{if } \sum_{i \in Nk(x')} y_i \geq 0 \\ -1 & \text{if } \sum_{i \in Nk(x')} y_i < 0 \end{cases} \tag{4.19}$$

with neighborhood size K and with the set of indices $Nk(x)$ of the K-nearest patterns.

The choice of K defines the locality of KNN. For a smaller K little neighborhoods arise in regions, where patterns from different classes are scattered. For bigger choices of K, patterns with labels in the minority are ignored. Therefore, the big question when building a KNN model is what K to choose to achieve the best neighborhood and, consequently, the best performance.

4.5.5 Convolutional Neural Networks

Convolutional Neural Networks (CNN's) were developed back in 1989 by LeCun to perform automatic recognition of handwritten zip code digits by the U.S. Postal Service (LeCun et al., 1989). These networks are intended to process data of a grid-like structure nature. Time-series can be processed by this model if they are thought as a 1D grid of values with different time-steps. Also, imaging data can be thought of as 2D or a 3D grid, depending on the color channel is included (Papernot et al., 2016).

CNN's key idea is the convolution operations performed to input data. A CNN is fully filled with three processing layers: convolutional layers, pooling layers, and fully connected layers (Papernot et al., 2016).

Convolutional layers

This is the first layer of a CNN model and is based on the discrete convolution operation. This operation can be described as an input D represented by a 2D array of size $n_1 * n_2$ convolved with a filter (also called kernel) of size $a * a$. The filter slides across the input D according to the stride parameter. The output

of a convolutional layer is a collection of feature maps of size $q_1 * q_2$, the number of generated feature maps is given by the number of filters i (Koushik, 2016). The i^{th} feature map of the l^{th} convolutional layer denoted $C_i^{(l)}$ is computed as follows:

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{ij}^{(l-1)} * C_j^{(l-1)} \tag{4.20}$$

where $B_i^{(l)}$ is the bias, $K_{ij}^{(l-1)}$ is the filter of size $a * a$ that connects the j^{th} feature map in layer $(l-1)$ with the i^{th} feature map in layer l and $C_j^{(l-1)}$ represents the j^{th} feature of layer $l-1$. The input of the first convolutional layer $(l=1)$ is the input data D, that is $C_1^{(0)} = D$.

Following the convolution layer, an activation function is applied to the feature maps and the output of the i^{th} feature map is:

$$Y_i^{(l)} = g(C_i^{(l)}) \tag{4.21}$$

where $g(x)$ represents an activation function (ReLU for this work).

Pooling layers

Following the convolutional layer is always the pooling layer and its objective is to replace the feature maps' output values at a certain location with a statistical summary of the nearby values reducing the dimensionality of the feature maps.

This objective is achieved using a mask of size $b * b$ to perform a pooling operation on each of the feature maps. From the most common pooling operations available, maximum pooling operation was selected for this work and it outputs the maximum value within a neighborhood (Koushik, 2016; Papernot et al., 2016).

This pooling operation can also increase the robustness of the model to noise and distortions since after this operation the representations of the feature maps are able to become approximately invariant to small translations of the input. Therefore, pooling layer is essential to improve the performance of the network for unseen data, but, no learning is done at the pooling layers. The only objective is to summarize the output responses over a neighborhood (Papernot et al., 2016).

Fully connected layers

Third and final layer of CNN is the fully connected layer. When convolutional and pooling layers are finished, the output of the last pooling layer is flattened and given as input to fully connected layers. A cost function is used to measure the discrepancy between the output of the network, the class labels, and the weights of the CNN are updated using backpropagation and Adam optimization method (Papernot et al., 2016).

4.5.6 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a class of neural networks specialized in processing sequential data. They are built with a chain-like structure and can be used to process time series and predict their outcome (Papernot et al., 2016). The key idea is to share parameters across different parts of a model, sharing the same parameters across the different time steps (Graves, 2012; Papernot et al., 2016).

A sequence that contains vectors $x^{(t)}$ with the time index t ranging from 1 to τ, the value of the hidden node can be defined as (Papernot et al., 2016):

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \theta) \tag{4.22}$$

where $h^{(t)}$ represents the state of a hidden node and θ represents the parameters of the network. the hidden state is then able to save a "memory" of previous inputs and therefore influence the output of the network.

The hidden nodes receive two input signals (Graves, 2012). An RNN with I input nodes, H hidden nodes, and K output nodes, the forward pass can be computed as follows:

$$u_h^{(t)} = \sum_{i=1}^{I} w_{ih} x_i^{(t)} + \sum_{h=1}^{H} w_{hh} z_h^{(t-1)} \tag{4.23}$$

where w_{ih} represents the weights between the input vector x and the hidden node h, w_{hh} represents the weights between hidden node h at time step $t-1$ and t and u_t represents the internal value of h

at time step t. the output value $z_h^{(t)}$ of the hidden node h is computed by applying an activation function $g(x)$:

$$z_h^{(t)} = g(u_h^{(t)}) \tag{4.24}$$

The output value $\alpha_k^{(t)}$ at time step t is then computed as:

$$\alpha_k^{(t)} = \sum_{h=1}^{H} w_{hk} z_h^{(t)} \tag{4.25}$$

$$z_k^{(t)} = g(a_k^{(t)}) \tag{4.26}$$

The loss of a RNN is the sum of the time step losses:

$$J(\theta) = \sum_{t=1}^{\tau} J(\theta) \tag{4.27}$$

where the $J(\theta)$ represents the cost function.

The training process of a RNN is made through backpropagation and this is how the gradients of the cost function respective to the parameters are calculated. In this specific case, the algorithm is backpropagation through time (BPTT). This algorithm unfolds an RNN into a Multilayer Perceptron allowing the application of standard back-propagation (Haykin, 2009). Using this, the gradients are easy to compute but very difficult to train due to vanishing or exploding gradient problems where the influence of the early time-steps either decays or blows up exponentially as the sequence is processed by the network (Allen-Zhu et al., 2018; Bengio et al., 1994; Sutskever et al., 2014).

The problem is that when applying backpropagation from $z_h^{(t)}$ to $z_h^{(t-1)}$ a multiplication by W_{hh}^T is performed. This means that many factors of W_{hh}^T are involved in the gradient computation of $z_h^{(0)}$. The value of the gradients will either vanish or explode depending on the magnitude of W_{hh}^T. This can be described as (Sutskever et al., 2014):

$$\frac{\partial J}{\partial W_{hh}} = \sum_{t=1}^{\tau} dz_k^{(t)} z_h^{(t-1)T} \tag{4.28}$$

where

$$dz_k^{(t)} = (\prod_{t=1}^{\tau} W_{hh}^T g'^{(z_k^{(t)})}) \quad (4.29)$$

If the values of W_{hh}^T are small then the contributions of the older inputs will rapidly tend to zero as t increases. The algorithm closest to solve this issue is Long Short-Term Memory (LSTM).

Long Short-Term Memory

Long Short-Term Memory (LSTM) is part of the RNNs algorithms. It is based on the idea of creating paths through time that derivatives that neither vanish nor explode.

It was developed in 1997 by Hochreither and Schmidhuber and since then it has been further developed and applied to various fields with a high degree of success. It works by a cell state which has a linear self-loop controlled by different gates (forget and input gate) and the output is computed by the output gate. The operations of an LSTM (Papernot et al., 2016) can be described as:

$$\textbf{Forget Gate: } f_i^{(t)} = b_i^f + \sum_{j=1}^{I} U_{i,j}^f x_j^{(t)} + \sum_{j=1}^{K} W_{i,j} h_j^{(t-1)} \quad (4.30)$$

$$f_i^{(t)} = \sigma(f_i^{(t)}) \quad (4.31)$$

$$\textbf{Input Gate: } s_i^{(t)} = \sigma(b_i^s + \sum_{j=1}^{I} U_{i,j}^S x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^s h_j^{(t-1)}) \odot \tilde{C}_i^{(t)} \quad (4.32)$$

$$\tilde{C}_i^{(t)} = tanh(b_i^c + \sum_{j=1}^{I} U_{i,j}^c x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^c h_j^{(t-1)}) \quad (4.33)$$

$$\textbf{Cell State: } C_i^{(t)} = (f_i^{(t)} \odot C_i^{(t-1)}) + s_i^{(t)} \quad (4.34)$$

$$\textbf{Output Gate: } o_i^{(t)} = \sigma(b_i^o + \sum_{j=1}^{I} U_{i,j}^o x_j^{(t)} + \sum_{j=1}^{K} W_{i,j}^o h_j^{(t-1)}) \quad (4.35)$$

$$h_i^{(t)} = o_i^{(t)} \odot tanh(C^{(t)}) \tag{4.36}$$

where $\odot$ is the symmetric tensor product and b, U, W are the bias, input weights, and recurrent weights of the LSTM gate respectively. σ represents the sigmoid function, $tanh$ represents the hyperbolic tangent function, $f_i^{(t)}$ represents the forget gate for time step t and cell i, $s_i^{(t)}$ represents the input gate for time step t and cell i, $C_i^{(t)}$ represents the cell state for time step t and cell i, $o_i^{(t)}$ represents the output gate for time step t and cell i and $h_i^{(t)}$ represents the hidden value of hidden node i at time step t.

As RNNs, LSTM can be trained using backpropagation and gradient descent. The vanishing gradient problem is tackled because the LSTM cell state provides a way for the gradient to flow allowing the network to learn long term dependencies (Papernot et al., 2016).

To go further into the analysis, CNN and LSTM approach were implemented with the group of features selected above from the recursive feature selection method to check if these two methods can improve the metrics. The features selected included its mean and standard deviation, but for CNN and LSTM the all stride values of the gait variables were used on the iteration. To perform these models, an algorithm iterating across different combinations of the number of neurons, learning rate, and kernel were developed to achieve the best model. The process for both CNN and LSTM is described in the flowchart in Figure 5.

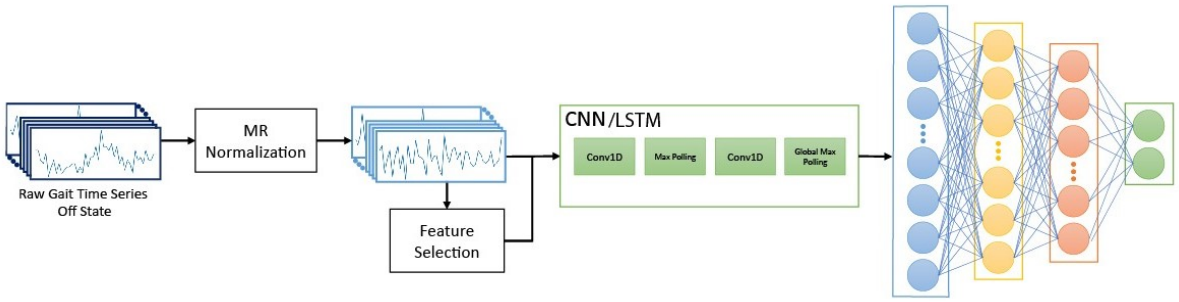

Figure 5: CNN and LSTM flowchart.

From Figure 5, first, from the raw full stride series of the gait assessment the MR normalization described in Section 4.1. was performed. From this outcome, a feature selection algorithm using recursive feature elimination was developed to select which gait variables are used in CNN and LSTM models.

4.6 Machine Learning metrics

All classifiers hyperparameters were tuned using randomized search and grid search method with 5-fold cross-validation: LR regularization strength constraint and type of penalty according to regularization strength constraint; SVM regularization parameter, gamma, and different types of kernel (with different degrees); RF maximum depth, minimum samples leaf, minimum samples split, and the number of estimators; finally, KNN number of neighbors, weights, and metric. All classifiers were implemented in Python programming language using Scikit-learn library (Pedregosa et al., 2011).

After tweaking your models for a while, eventually, the system performs sufficiently well. Now is the time to evaluate the final model on the test set. Often classification accuracy is used to measure the performance of the model, however, it is not enough to truly judge it. The most common metrics to evaluate a classification model are:

- **accuracy**: the ratio of correct predictions;
- **precision**: the ratio of correctly predicted positive observations to the total predicted positive observations;
- **sensitivity/recall**: the ratio of positive instances that are correctly detected by the classifier;
- **specificity**: the ratio of negative instances that are correctly detected by the classifier;
- **F1 Score**: the harmonic mean between precision and recall whose range goes from 0 to 1 (Gu et al., 2009). It tells how precise the classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances):

$$F1\ Score = \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{4.37}$$

Part III :

Results, Discussion, Conclusion, and Future Work

Chapter 5

Gait assessment

5.1 Participants

Gait assessment data from seventy-six subjects (28 males) were collected. Of this group, thirty-nine of the patients have FD (14 males) and thirty-seven are healthy subjects (14 males). Of the thirty-nine FD patients, twenty-five patients have White Matter Lesions (WMLs)(7 males). The subject demographics of the groups of FD with WMLs, FD without WMLs, and controls are summarized in Table 6. For all FD patients, the exclusion criteria were: less than eighteen years of age, the presence of resting tremor, moderate-severe dementia (CDR $>$ 2), depression, extensive intracranial lesions, or neurodegenerative disorders, musculoskeletal disease, and rheumatological disorders. Local hospital ethics committee approved the protocol of the study, submitted by ICVS/UM and Center Algoritmi/UM. Written consent was obtained from all subjects or their guardians.

Table 6: Demographic variables for FD patients with and without WMLs and Controls.

	FD patients with WMLs (n = 24)	FD patients without WMLs (n = 15)	Controls (n = 37)
Age (years)	59.29 (15.99)	36.93 (10.99)	52.76 (22.91)
Male (%)	29%	47%	38%
Weight (kg)	66.03 (10.47)	64.69 (6.67)	66.51 (9.12)
Height (m)	1.59 (0.07)	1.66 (0.09)	1.62 (0.09)

Data is presented as mean (standard deviation)

5.2 Gait Normalization

The gait normalization was performed based on the physical properties and the mean values of the gait variables of the 37 healthy controls. Variance Inflation Factor (VIF) was employed to measure the colinearity of the physical characteristics (age, weight, height, and sex), speed, and stride length and the results are presented in Table 7.

Table 7: Variance inflation factor for physical characteristics (age, weight, height, and sex), speed, and stride length.

	Age	Weight	Height	Sex	Speed	Stride Length
	2.24	1.97	3.76	1.28	6.05	9.89
VIF	1.99	1.93	3.37	1.27	-	2.06
	1.82	1.97	2.9	1.28	1.26	-

Since when all variables are used both speed and stride length had a VIF value higher than 5 (an indicator of strong correlation), they can not be used simultaneously. When used separately no VIF value is higher than 5. Therefore, from the gait variables speed and stride length, only stride length is used on further investigation.

All possible combinations for each feature were computerized by a developed algorithm, allowing to find the best possible model for each feature. The models created for each feature are summarized in Table 8 (right foot) and in Table 38 (left foot).

The models obtained for both feet shown in Table 8 and Table 38 are very similar, therefore, all the

Table 8: Multiple linear regression models using only significative independent variables for each gait variable on the right foot.

	MR normalization						AIC	Adjusted R^2
		Age	Weight	Height	Speed	Gender	Stride length		
Spatial-temporal variables									
Cycle duration	= +0.84	-0.0011		+0.372	-0.286			-143.71	0.771
Cadence	= +137.86	+0.128		-42.44	+33.41			197.76	0.776
Stride length	= -0.23	-0.0014		+0.502	+0.611			-126.29	0.928
Stance	= +67.71				-5.933			146.57	0.227
Swing	= +32.28				5.933	0.362		146.57	0.227
Loading	= +16.72		-0.129		+2.68	+2.41		154.30	0.363
Foot flat	= +63.20	+0.079	+0.162		-17.73			190.02	0.662
Pushing	= +21.81	-0.067	-0.076		+14.88			181.53	0.600
Double support	= +35.78				-11.89			187.90	0.277
Peakswing	= +154.26				+178.691			340.04	0.570
Foot clearance variables									
Strike angle	= +16.14			-13.82		+3.83	+22.86	195.22	0.515
Lift-off angle	= -28.99	+0.178					-35.40	214.51	0.767
Maximum heel clearance	= +0.137	-0.00065		+0.085		+0.030		-147.28.06	0.449
Maximum toe clearance 1	= +0.105						-0.027	-170.17	0.006
Minimum toe clearance	= +0.013	+0.0003						-244.59	0.427
Maximum toe clearance 2	= +0.023	-0.00057				+0.024	+0.115	-181.17	0.736

following analyses will be performed only for the right foot. After normalizing all strides of each of the sixteen gait variables, two datasets were created:

1. Gait measures: the mean and the standard deviation (SD) of each variable (each gait time series) for all subjects. The SD value is used to measure the variability of each gait variable;

	Patient	Cycle Duration	Cadence	Stance	Swing	Loading	Foot Flat	Pushing	Double Support
Mean	N167	0,973314948	0,983298187	1,011724801	0,938906889	0,901196095	0,934988597	1,222782097	1,155275267
Standard deviation	N167	0,143578609	0,14474563	0,148760537	0,162373981	1,009738007	0,213624752	0,39186564	0,480788303

Figure 6: Example of the mean and variability (SD) for patient N167 and his gait variables: cycle duration, cadence, stance, swing, loading, foot flat, pushing and double support.

2. Gait time series: the all normalized stride values of each gait variable of all subjects (all the time series).

Patient	Cycle Duration	Cadence	Stance	Swing	Loading	Foot Flat	Pushing	Double Support
N167	0,997181133	0,996118302	1,040776511	0,937819908	0,676034895	0,925959164	1,253554577	1,246000363
N167	0,998465903	0,994775824	1,033255207	0,949334147	0,680640881	0,921572655	1,260131708	1,030657468
N167	0,982811233	1,010689029	1,02382903	0,963659942	0,639115485	0,916070563	1,293030295	1,006426641
N167	0,982145827	1,011185872	1,046062537	0,929954287	0,741745251	0,918394758	1,235977167	1,096916111
N167	1,001348782	0,991624059	1,032423945	0,950859412	0,796351255	0,909970438	1,225041902	1,175434265
N167	0,970355959	1,023223349	1,036071185	0,94542962	0,644513198	0,941259523	1,230039827	1,084151096
N167	1,001091397	0,991684467	1,033579246	0,949759942	0,805666695	0,948895592	1,131916939	1,152255131
N167	0,995410454	0,997444617	1,035409808	0,946468132	0,80069528	0,933263045	1,171950811	1,096762915
N167	0,979809425	1,013257212	1,0241022	0,96365454	0,824420668	0,926338213	1,173574751	1,124706252
N167	0,986202869	1,006877398	1,039180844	0,940587111	0,802917009	0,942138421	1,153861973	1,126008695
N167	0,973815008	1,019625101	1,038760935	0,941305252	0,814529485	0,919460519	1,195567813	1,124257841
N167	0,982005507	1,011038077	1,050932407	0,923055811	0,85733466	0,9601581	1,088171316	1,136222944
N167	0,983637124	1,009437312	1,036791265	0,944296677	0,75033157	0,938553762	1,185734907	1,100756651
N167	0,983778151	1,009288758	1,044931827	0,931978219	0,801713884	0,944344225	1,148886589	1,095671398
N167	0,996227622	0,996533928	1,050107396	0,924538911	0,791762267	0,957064624	1,124580557	1,126082236
N167	0,994191518	0,998600755	1,025307846	0,961833914	0,811486915	0,911662961	1,210307671	1,220227733
N167	0,982495704	1,010454222	1,041256717	0,937901084	0,810071885	0,900931653	1,231664029	1,127554322
N167	0,989373193	1,003463901	1,038424245	0,942053736	0,805143235	0,891438186	1,255418774	1,081912251
N167	0,979533715	1,013546654	1,048600941	0,926697133	0,747721067	0,917486588	1,228540703	1,110999528
N167	1,01649225	0,976640176	1,036617275	0,945065431	0,84537385	0,923348947	1,166549701	1,081967007
N167	0,982529797	1,010606754	1,020279512	0,969272469	0,762776797	0,925951783	1,206993517	1,088105209
N167	0,99174522	1,001395717	1,036022461	0,945224312	0,798770154	0,931208691	1,180905837	1,132821715
N167	0,989241709	1,004038392	1,035308513	0,946216518	0,685751738	0,916197057	1,269152783	1,087277561
N167	0,987776369	1,006522899	1,023986684	0,963058881	0,802004719	0,920201094	1,212811927	1,054331378
N167	0,97183983	1,022802304	1,05257576	0,919192558	0,624211563	0,965289635	1,197019532	1,205883176
N167	1,027943686	0,966208929	1,035488419	0,945966169	0,880117584	0,895071637	1,220719947	1,055155134
N167	1,00137658	0,991971933	1,019995359	0,969497459	0,743898537	0,927001183	1,218741153	1,078344994
N167	1,000530686	0,99285329	1,027603941	0,957866121	0,68179911	0,956942915	1,18380025	1,152048374
N167	0,995887382	0,997477345	1,040407118	0,93832793	0,789356718	0,94978478	1,147300065	1,112270682
N167	1,00982806	0,983503421	1,024020574	0,963449775	0,736506988	0,910573715	1,255894941	1,038680505
N167	0,98990981	1,00333048	1,035499298	0,945951038	0,685459544	0,95499892	1,184912885	1,098995209
N167	1,007988012	0,985187087	1,013967619	0,978788901	0,746561661	0,940186355	1,185899883	1,043633289
N167	0,978267969	1,015372489	1,012781989	0,980509858	0,767205845	0,93080076	1,198869573	1,053597052
N167	0,989185447	1,004098097	1,035292438	0,946238898	0,800072435	0,941594878	1,158912759	1,167745313
N167	0,989207586	1,003584708	1,0366813	0,944858422	0,693541567	0,955686097	1,173039864	1,127901363
N167	0,997094057	0,995725977	1,035883475	0,945814314	0,742724759	0,895848092	1,277129019	1,094402484
N167	0,990203158	1,002699125	1,03225816	0,95121938	0,74949852	0,914069879	1,236928515	1,117098412
N167	0,979907031	1,013154823	1,048708865	0,926553317	0,86255235	0,930938405	1,145867906	1,130351806
N167	1,002100493	0,991015981	1,042174757	0,935905141	0,78626325	0,917829348	1,215294818	1,109167602
N167	0,999502481	0,993287701	1,036559039	0,944884403	0,798666808	0,923979755	1,190525086	1,096183633
N167	0,998197459	0,994708638	1,021883541	0,966871267	0,80796591	0,886554963	1,268083542	1,069747363
N167	0,98847944	1,004491133	1,047892198	0,927491664	0,738767494	0,924154557	1,221697192	1,113901286
N167	0,998316904	0,994931153	1,048987347	0,925357824	0,726341322	0,93244793	1,214533395	1,125777027
N167	0,995543134	0,997308749	1,043473901	0,934282948	0,737722514	0,951932032	1,162659846	1,136383235
N167	0,987807111	1,005354925	1,025246303	0,961632662	0,752906568	0,933566139	1,197522488	1,056877851
N167	0,982714992	1,011373449	1,03503663	0,946206567	0,742611773	0,949114336	1,174721738	1,152561192
N167	0,987779275	1,00585415	1,044411603	0,932056166	0,734562405	0,959102633	1,154586435	1,144895627

Figure 7: Example of a gait time series for patient N167 and his gait variables: cycle duration, cadence, stance, swing, loading, foot flat, pushing and double support.

5.3 FD patients with WMLs vs controls

To the differentiation task of FD with WMLs vs Controls, 24 subjects from the control group, aged-matched with the group of FD patients with WMLs, were selected. The physical characteristics of these two groups are described in Table 9.

Table 9: Demographic variables for FD patients with WMLs and Controls.

	FD patients with WMLs (n = 24)	**Controls (n = 24)**	**p-value**
Age (years)	59.29 (15.99)	59.13 (21.69)	0.451
Male (%)	29%	29%	1
Weight (kg)	66.03 (10.47)	66.10 (10.27)	0.0.934
Height (m)	1.59 (0.07)	1.63 (0.07)	0.053

Data is presented as mean (standard deviation).

The comparison between the mean value of gait features in FD patients with WMLs vs controls for the raw gait data and the MR normalized gait is represented in Figure 8, where significant differences in gait features between FD patients with WMLs and controls are indicated with one asterisk (*$p<0.001$), and whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot. When using gait raw data: stance, swing, loading, pushing, double support, strike angle, maximum heel clearance, maximum toe clearance 1, and maximum toe clearance 2 are significantly different between FD patients with WMLs and controls (for more details see Table A.37). After normalization, loading is not statistically significant while cycle duration, cadence, foot flat, stride length, lift-off angle, and minimum toe clearance are now statistically significant.

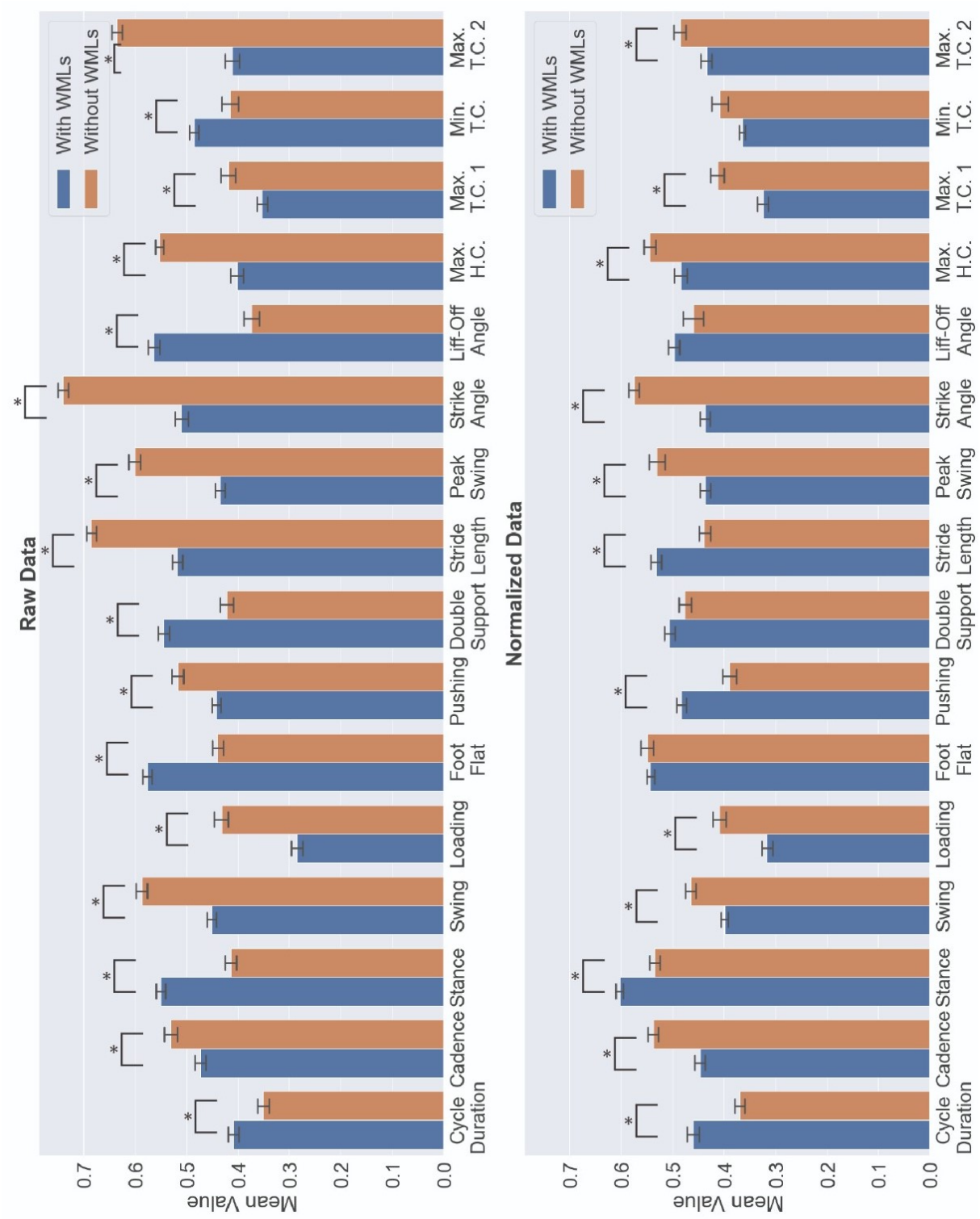

Figure 8: Comparison between the mean value of gait features in FD patients with WMLs vs controls. Data are shown for the raw gait data and the MR normalized gait. Max.: maximum; Var.: variability; Min.: minimum; H.C. : heel clearance; T.C. : toe clearance

5.3.1 Feature selection

Recursive feature elimination (RFE) technique was performed on the 10 remaining features from the filter method: cycle duration, swing, pushing, maximum toe clearance 1, maximum toe clearance 2, loading variability, foot flat mean and variability, and maximum heel clearance mean and variability. The heat map of these 10 features selected from the filter method algorithm is shown in Figure 9.

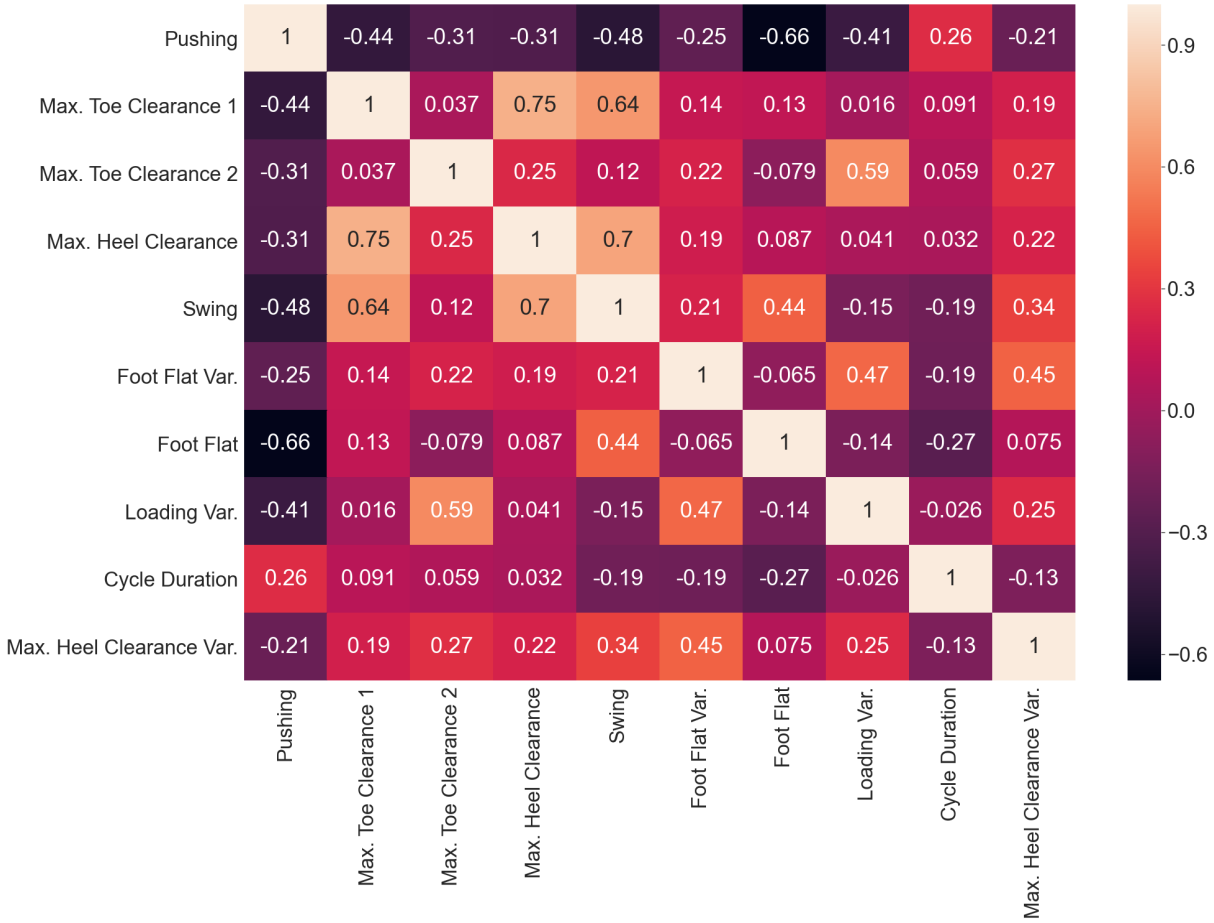

Figure 9: Gait correlation heat map in FD patients with WMLs vs controls.
Max.: maximum; Var.: variability; Min.: minimum.

The results are summarized in Figure 10 and Table 10. Four gait characteristics were selected by LR with a F1 score of 61.08%, while SVM selected 7 features with a F1 score of 60.42% and RF selected 7 features with a F1 score of 68.55%. In Table 10, are the training and validation accuracies for the optimal models of each algorithms. RF had the higher accuracy in the training while the three models showed

Table 10: Results for each of 3 machine learning algorithms (Random Forest, SVM with linear kernel, and Logistic Regression) with recursive feature elimination for FD patients with WMLs vs controls.

	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**
Logistic Regression	67.66 ± 2.63	63.50 ± 14.46
SVM (Linear Kernel)	71.79 ± 4.27	65.50 ± 18.19
Random Forest	80.24 ± 2.91	65.00 ± 14.14

SD: standard deviation

similar accuracies in validation, being SVM the highest.

Figure 10 (right side) shows the contribution of each gait characteristic in the classification model. The common gait characteristics among the selected features by each model were three (Top 3): foot flat, cycle duration, and pushing. The common gait characteristics from LR and SVM were four (Top 4): cycle duration, pushing, and foot flat mean and variability, illustrated in Figure 11.

5.3.2 Gait classification based on gait measures

These Top 3 and Top 4 were evaluated with five classification models (LR, SVM Linear kernel, SVM RBF kernel, RF, and KNN) to identify the optimal combination of gait characteristics and the classification model with better performance.

As shown in Table 11, RF, SVM RBF Kernel, and KNN performance increased slightly by reducing the number of features from 4 to 3 while SVM Linear Kernel had the same accuracy in both cases and LR had a decrease in the model's validation accuracy. With Top 4, higher accuracies are observed in LR and RF related to Top 3. KNN and SVM Linear Kernel show equal validation accuracy with Top 3 and Top 4, although KNN displays higher training accuracy with Top 3. The better performance is observed for RF with Top 3 and Top 4 achieving a validation accuracy of 71.50% and a training accuracy of 77.61% in Top 3 and a validation accuracy of 64.50% and a training accuracy of 70.18% in Top 4.

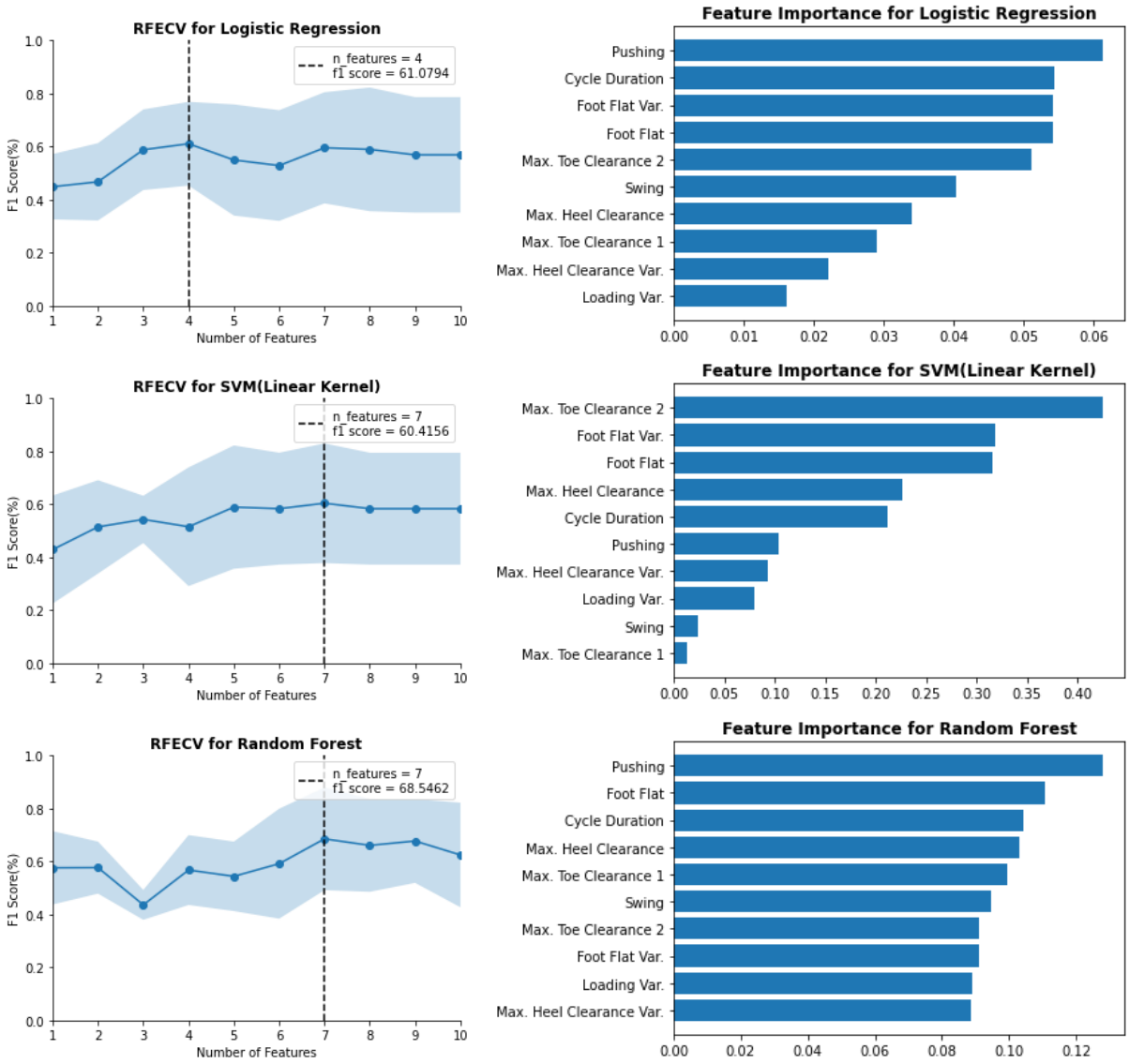

Figure 10: F1 score vs number of features for selection of optimal numbers of gait characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients with WMLs vs controls. Recursive feature elimination was used through the 5-fold cross-validation (RFECV).
Max.: maximum; Var.: variability; Min.: minimum.

5.3.3 Gait classification based on gait time series

The groups selected were Top 3 (same as Top LR and Top 4): pushing, cycle duration, and foot flat; Top SVM: pushing, cycle duration, foot flat, maximum toe clearance 2 and maximum heel clearance; and

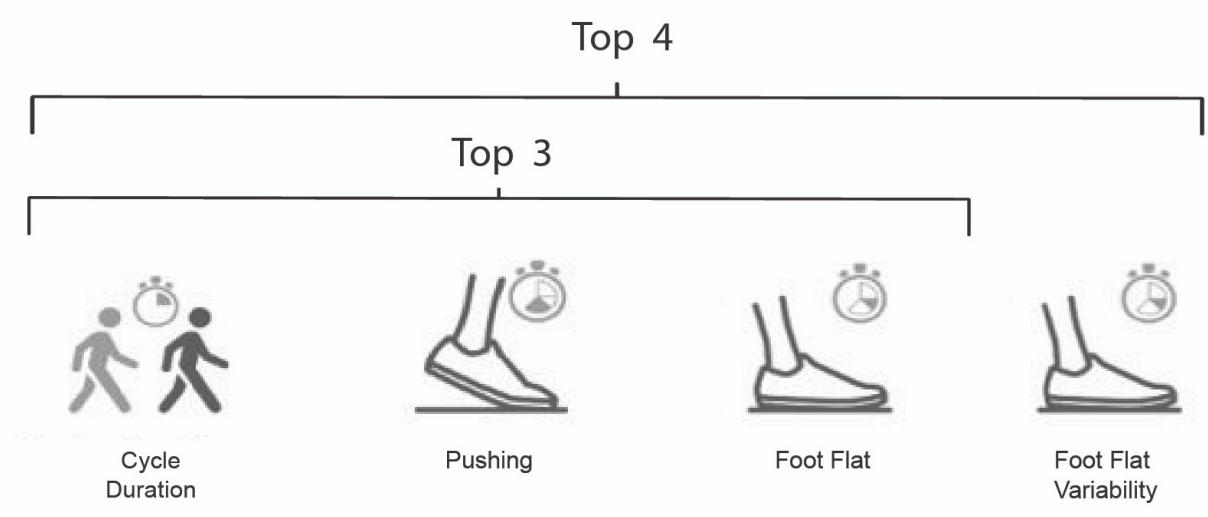

Figure 11: Top 3 and Top 4 after performing RFE on gait features for FD patients with WMLs vs controls.

Table 11: Classification accuracy on training and validation data for top common gait characteristics in FD patients with WMLs vs controls with LR, SVM (linear and RBF kernel), RF, and KNN.

	Top 3		**Top 4**	
	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Training accuracy% (Mean ± SD)**	**Validation accuracy% (Mean ± SD)**
LR	64.58 ± 2.05	63.00 ± 12.49	67.66 ± 2.63	63.50 ± 14.46
SVM (Linear Kernel)	67.63 ± 5.24	61.50 ± 15.46	65.00 ± 6.09	61.50 ± 16.70
SVM (RBF Kernel)	60.92 ± 5.06	64.50 ± 10.05	66.11 ± 4.92	61.50 ± 14.11
RF	77.61 ± 2.04	71.50 ± 10.91	70.18 ± 3.67	64.50 ± 4.58
KNN	71.04 ± 7.43	60.57 ± 13.08	74.38 ± 3.64	60.12 ± 15.30

LR: Logistic Regression, RBF: radial basis function, RF: Random Forest, KNN: K-Nearest Neighbour, SD: standard deviation

Top RF: pushing, cycle duration, foot flat, maximum toe clearance 1, maximum heel clearance, swing, and maximum toe clearance 2.

Table 12 shows the results for CNN and LSTM performed based on the gait time series of the FD patients with WML vs controls. Top RF achieved the best performance for both CNN and LSTM with a validation accuracy of 71.61% and 69.39%, respectively. Comparing CNN and LSTM algorithms, CNN achieved the highest percentage of validation accuracy for Top 5 and Top SVM, while Top 3 had the highest validation accuracy in LSTM method.

Finally, comparing the results from Table 10, Table 11, and Table 12, Top 3 had a better performance

Table 12: CNN and LSTM results on selected gait features from the gait time series for FD patients with WMLs vs controls.

	CNN		LSTM	
	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Training accuracy% (Mean ± SD)**	**Validation accuracy% (Mean ± SD)**
Top 3	62.79 ± 3.69	58.89 ± 12.40	62.18 ± 7.14	61.39 ± 8.82
Top SVM	76.72 ± 8.82	73.83 ± 16.93	75.57 ± 9.28	69.11 ± 10.59
Top RF	83.11 ± 6.71	71.61 ± 17.35	75.03 ± 12.76	69.39 ± 15.54

CNN: Convolutional Neural Networks, LSTM: Long-Short Term Memory, SD: standard deviation

using the first approach with RF algorithm with a validation accuracy of 71.50%. On the other hand, in Top SVM and Top RF both CNN and LSTM outperformed the rest of the algorithms, being CNN the algorithm with the best metrics achieving 73.83% and 71.61%, respectively. Regarding the standard deviation of the validation accuracy, there are no major differences between the results from all the models.

5.4 FD patients without WMLs vs controls

To FD without WMLs vs Controls differentiation task was selected from the control group 15 subjects aged-matched with the group of FD patients without WMLs. The physical characteristics of these two groups are described in Table 13.

Table 13: Demographic variables for FD patients without WMLs and Controls.

	FD patients without WMLs (n = 15)	**Controls (n = 15)**	**p-value**
Age (years)	36.93 (10.99)	37.00 (13.12)	0.87
Male (n(%))	47%	47%	1
Weight (kg)	64.69 (6.67)	69.73 (13.82)	0.325
Height (m)	1.66 (0.09)	1.71 (0.08)	0.246

Data is presented as mean (standard deviation).

The comparison between the mean value of gait features in FD patients without WMLs vs controls for

the raw gait data and the MR normalized gait is represented in Figure 12, where significant differences in gait features between FD patients without WMLs and controls are indicated with one asterisk (*$p<0.001$), and whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot. When using gait raw data: cycle duration, cadence, loading, foot flat, pushing, peak swing, strike angle, lift-off angle, maximum heel clearance, and minimum toe clearance are statistically significant between FD patients without WMLs and controls (for more details see Table A.37). After normalization, pushing is not statistically significant while stride length is now statistically significant.

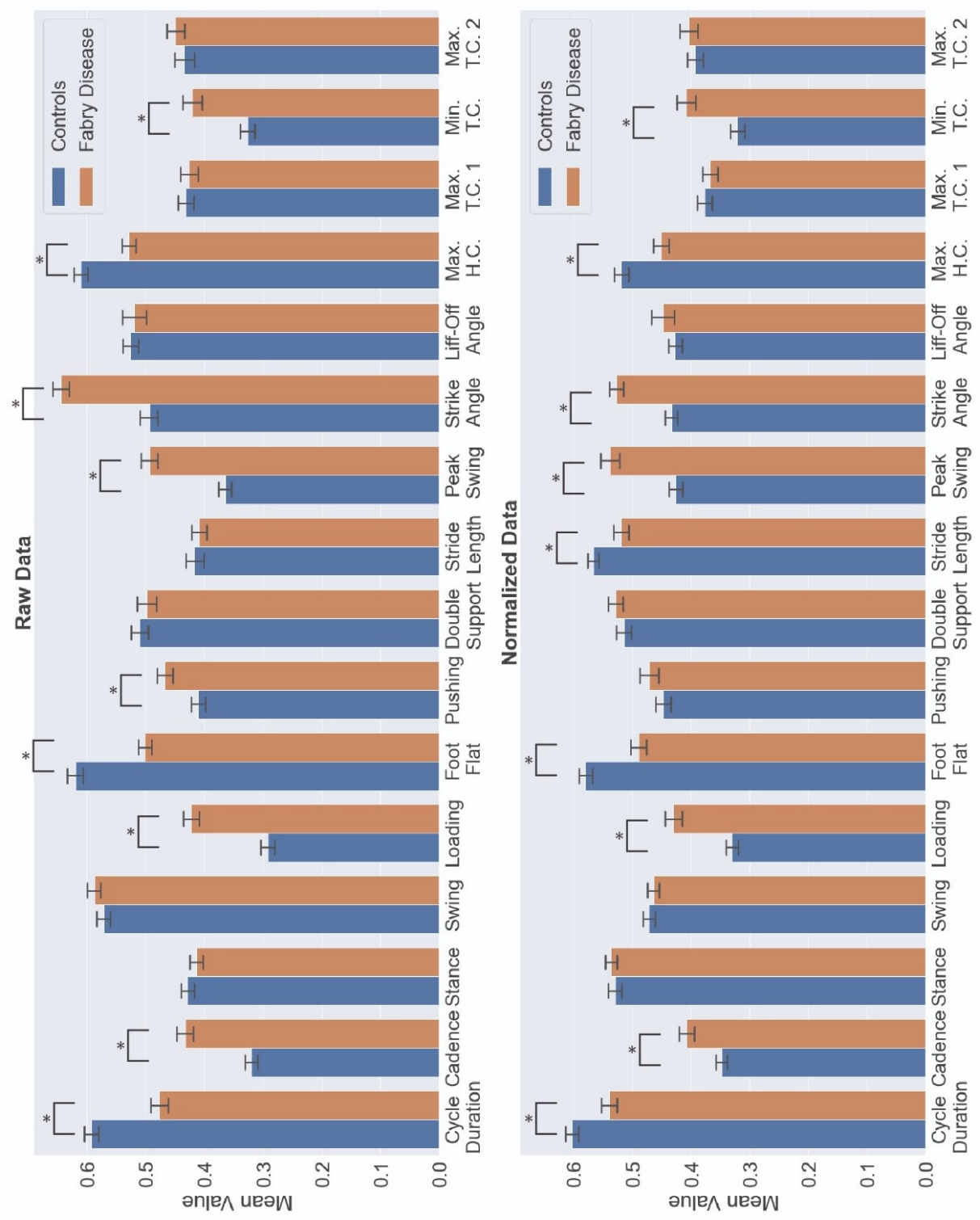

Figure 12: Comparison between the mean value of gait features in FD patients without WMLs vs controls. Data are shown for the raw gait data and the MR normalized gait. Max.: maximum; Var.: variability; Min.: minimum; H.C. : heel clearance; T.C. : toe clearance

5.4.1 Feature selection

Recursive feature elimination algorithm was performed on the 10 remaining features from filter method: foot flat, peak swing, strike angle, minimum toe clearance mean and variability, stride length variability, loading mean and variability, lift-off angle variability, and maximum heel clearance variability. The heat map of these 10 features selected from the filter method algorithm is shown in Figure 13. Results are stated in Table 14 and Figure 14.

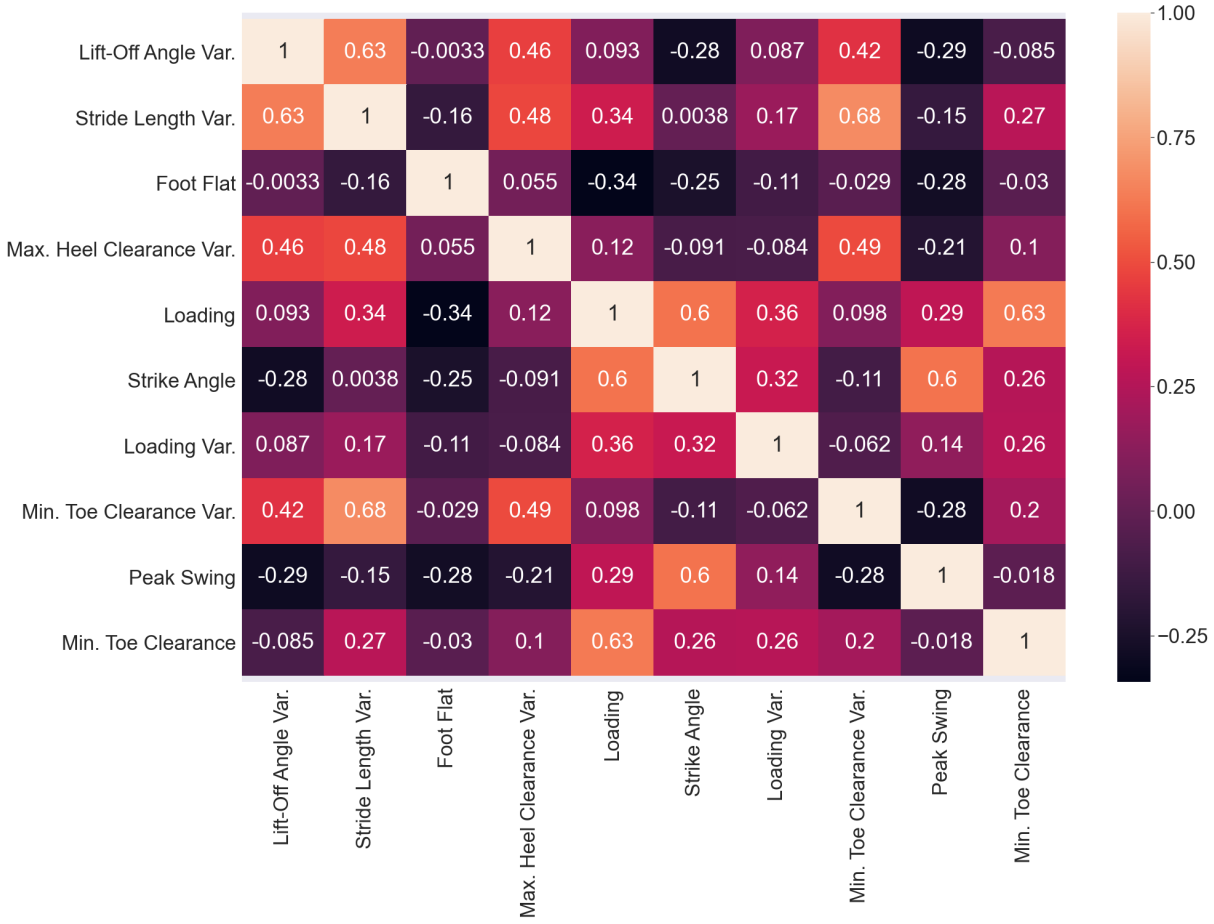

Figure 13: Gait correlation heat map in FD patients without WMLs vs controls.
Max.: maximum; Var.: variability; Min.: minimum.

The results show that LR and SVM selected 5 features as the optimal number of gait characteristics with a F1 score of 72.31% and 74.90%, respectively, and RF selected 6 features with a F1 score of 79.81%. Table 14 presents the training and validation accuracies for the optimal models of each algorithm. LR

presents the higher training and validation accuracies.

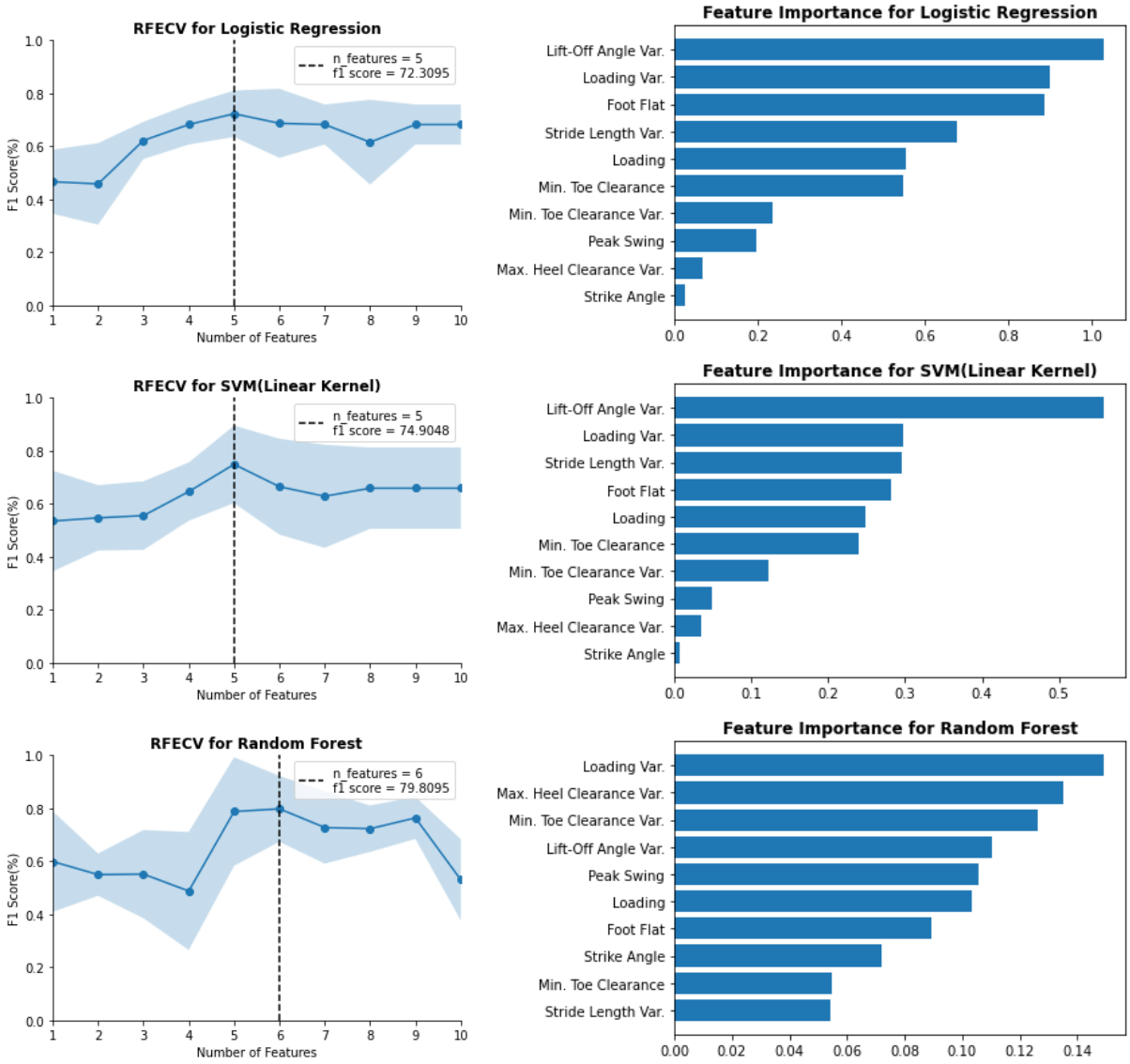

Figure 14: F1 score vs number of features for selection of optimal numbers of gait characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients without WMLs vs controls. Recursive feature elimination was used through the 5-fold cross-validation (RFECV).
Max.: maximum; Var.: variability; Min.: minimum.

Taking into account the contribution of each gait characteristic in the classification model (Figure 14, right side), the common features were 3 (Top 3): loading mean and variability and lift-off angle variability. The common gait characteristics from LR and SVM were 5 (Top 5): foot flat, stride length variability,

Table 14: Results for each of 3 machine learning algorithms (LR, SVM with linear kernel, and RF) with optimal gait characteristics' from recursive feature elimination for FD patients without WML vs controls.

	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**
LR	88.33 ± 3.12	76.67 ± 8.16
SVM (Linear Kernel)	85.00 ± 3.33	76.67 ± 8.16
RF	94.17 ± 4.25	70.00 ± 12.47

SD: standard deviation

loading mean and variability, and lift-off angle variability (see Figure 15).

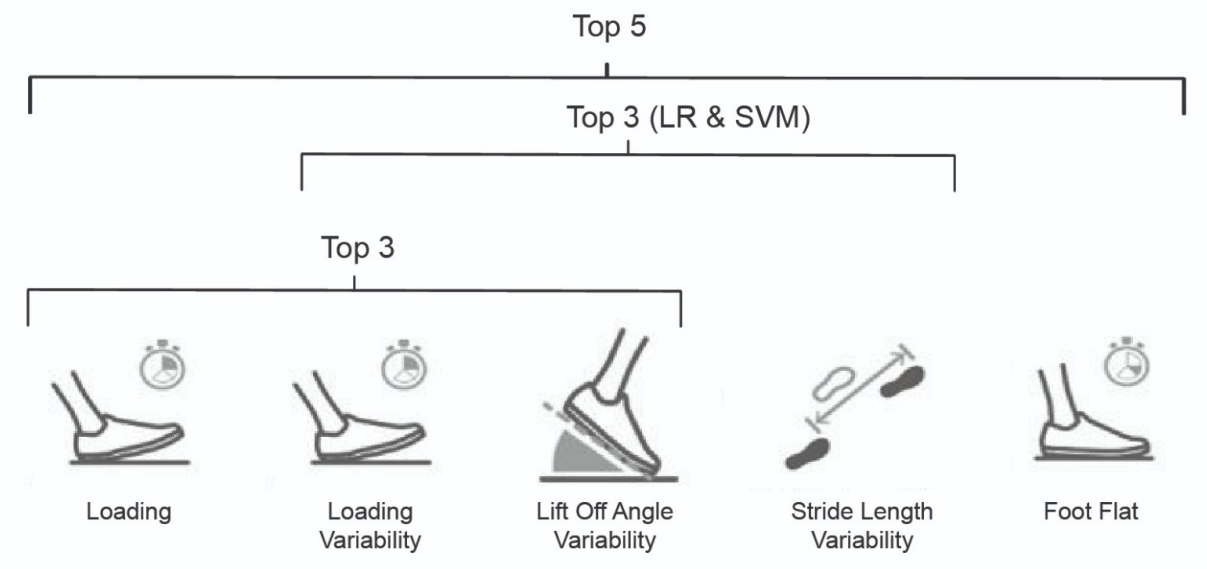

Figure 15: Top 3, Top 5, and Top 3 (LR & SVM) after performing RFE on gait features for FD patients without WMLs vs controls.

5.4.2 Gait classification based on gait measures

These Top 3 and Top 5, as well as the Top 3 from LR and SVM (stride length variability, loading variability and lift-off angle variability), as shown in Figure 15, were evaluated with five classification models (LR, SVM Linear kernel, SVM RBF kernel, RF, and KNN) to identify the optimal combination of gait

Table 15: Classification accuracy on training and validation data for top common gait characteristics in FD patients without WMLs vs controls with LR, SVM (linear and RBF kernel), RF, and KNN.

	Top 3		Top 5		Top 3 (LR & SVM)	
	Training accuracy% (Mean ± SD)	Validation accuracy% (Mean ± SD)	Training accuracy% (Mean ± SD)	Validation accuracy% (Mean ± SD)	Training accuracy% (Mean ± SD)	Validation accuracy% (Mean ± SD)
LR	83.33 ± 4.56	80.00 ± 12.47	88.33 ± 3.12	76.67 ± 8.16	80.83 ± 2.04	80.00 ± 12.47
SVM (Linear Kernel)	85.00 ± 3.33	73.33 ± 13.33	85.00 ± 3.33	76.67 ± 8.16	83.33 ± 3.73	76.67 ± 8.16
SVM (RBF Kernel)	80.00 ± 4.08	80.00 ± 19.44	87.50 ± 4.56	80.00 ± 6.67	78.33 ± 4.86	70.00 ± 12.47
RF	92.50 ± 7.17	70.00 ± 16.33	95.83 ± 3.73	70.00 ± 12.47	90.83 ± 1.67	73.33 ± 8.16
KNN	91.67 ± 2.64	83.33 ± 10.54	91.67 ± 2.04	86.67 ± 8.16	93.33 ± 4.56	86.07 ± 12.47

LR: Logistic Regression, RBF: radial basis function, RF: Random Forest, KNN: K-Nearest Neighbour, SD: standard deviation

characteristics and the classification model with better performance.

From Table 15, KNN achieved the highest validation accuracy in all three different top's with a validation accuracy of 83.33% in Top 3, 86.67% in Top 5 from LR and SVM, and 86.07% in Top 3 from LR and SVM. The validation accuracy in KNN and SVM Linear Kernel increased by changing from Top 3 to Top 5 while the other models decreased (except SVM RBF Kernel who remained equal). Going from Top 5 to Top 3 from LR and SVM, produced a slight decrease in SVM RBF Kernel and increase in LR and RF (SVM Linear Kernel and KNN had no change). Overall, KNN showed higher mean validation accuracy followed by SVM RBF Kernel.

5.4.3 Gait classification based on gait time series

The groups selected were Top 5 (same as Top LR and Top SVM): lift-off angle, loading, foot flat, and stride length; Top RF: loading, maximum heel clearance, minimum toe clearance, lift-off angle, and peak swing; Top 3: lift-off angle and loading; and Top 3 LR & SVM: stride length, loading, and lift-off angle.

Table 16 shows the results for CNN and LSTM performed based on gait time series of the FD patients with WML vs controls. Top 5 achieved the best performance when using CNN algorithm with a validation accuracy of 80.48% while Top RF achieved the best performance for LSTM with a validation accuracy of 67.62%. Comparing CNN and LSTM algorithms, CNN outperformed LSTM in all group of features.

Finally, comparing the results from Table 14, Table 15 and Table 16, KNN achieved the highest vali-

Table 16: CNN and LSTM results on selected gait features from the gait time series for FD patients without WMLs vs controls.

	CNN		**LSTM**	
	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Training accuracy% (Mean ± SD)**	**Validation accuracy% (Mean ± SD)**
Top RF	84.53 ± 10.38	69.05 ± 25.95	82.23 ± 21.98	67.62 ± 23.65
Top 3	85.37 ± 14.89	65.24 ± 18.53	59.73 ± 19.75	48.57 ± 22.08
Top 5	87.93 ± 10.12	80.48 ± 19.69	75.00 ± 13.01	60.48 ± 14.17
Top 3 (LR & SVM)	77.33 ± 9.28	64.76 ± 11.21	71.73 ± 16.00	55.24 ± 17.78

CNN: Convolutional Neural Networks, LSTM: Long-Short Term Memory, SD: standard deviation

dation for Top 3, Top 5, and Top 3 (LR & SVM) with a validation accuracy of 83.33%, 86.67%, and 86.07%, respectively. For Top RF, using RF algorithm the validation accuracy was higher with 70%. Regarding the standard deviation of the validation accuracy, it showed higher for all the results in both CNN and LSTM.

5.5 FD patients with vs without WMLs

For the analysis of FD patients with WMLs vs FD patients without WMLs all patients were included. The physical characteristics of these two groups are described in Table 17.

Table 17: Demographic variables for FD patients with and without WMLs.

	FD patients with WMLs (n = 24)	**FD patients without WMLs (n = 15)**	**p-value**
Age (years)	59.29 (15.99)	36.93 (10.99)	$<$ **0.001**
Male (%)	29%	47%	0.318
Weight (kg)	66.03 (10.47)	64.69 (6.67)	0.743
Height (m)	1.59 (0.07)	1.66 (0.09)	**0.011**

Data is presented as mean (standard deviation).

The comparison between the mean value of gait features in FD patients with vs without WMLs for the

raw gait data and the MR normalized gait is represented in Figure 16, where significant differences in gait features between FD patients with and without WMLs are indicated with one asterisk (*$p<0.001$), and whiskers represent 95% confidence interval (CI) values. The data was scaled between 0 and 1 to fit onto the same plot. When using gait raw data all variables are statistically significant between FD patients with and without WMLs (for more details see Table A.37). After normalization foot flat, double support, lift-off angle, and maximum toe clearance 1 are not statistically significant.

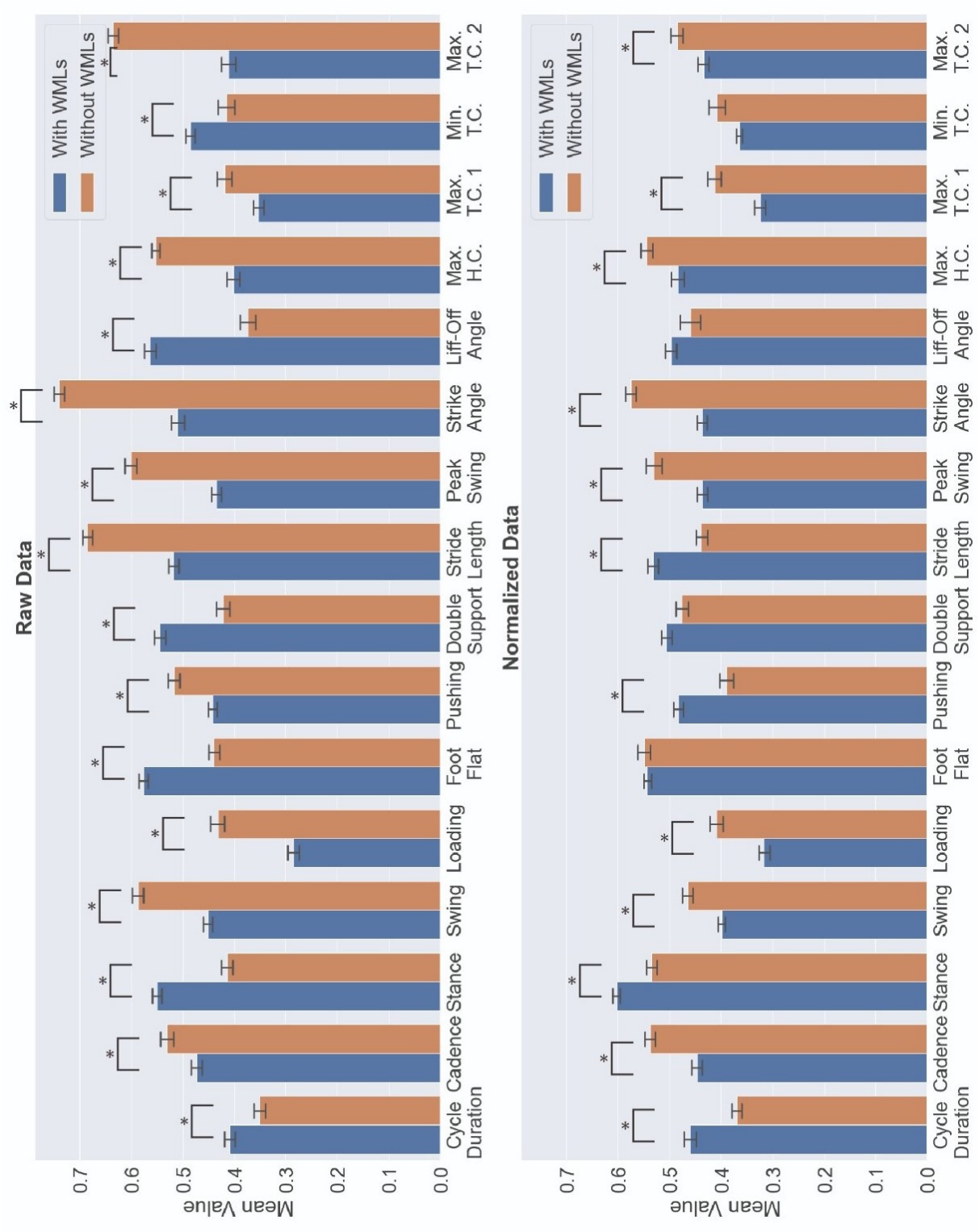

Figure 16: Comparison between the mean value of gait features in FD patients with vs without WMLs. Data are shown for the raw gait data and the MR normalized gait. Max.: maximum; Var.: variability; Min.: minimum; H.C. : heel clearance; T.C. : toe clearance

5.5.1 Feature selection

Recursive feature elimination algorithm was performed on the 10 remaining features from filter method: swing, minimum toe clearance variability, loading variability, double support variability, maximum toe clearance 1, foot flat variability, peak swing, pushing, strike angle, and lift-off angle variability. The heat map of these 10 features selected from the filter method algorithm is shown in Figure 17.

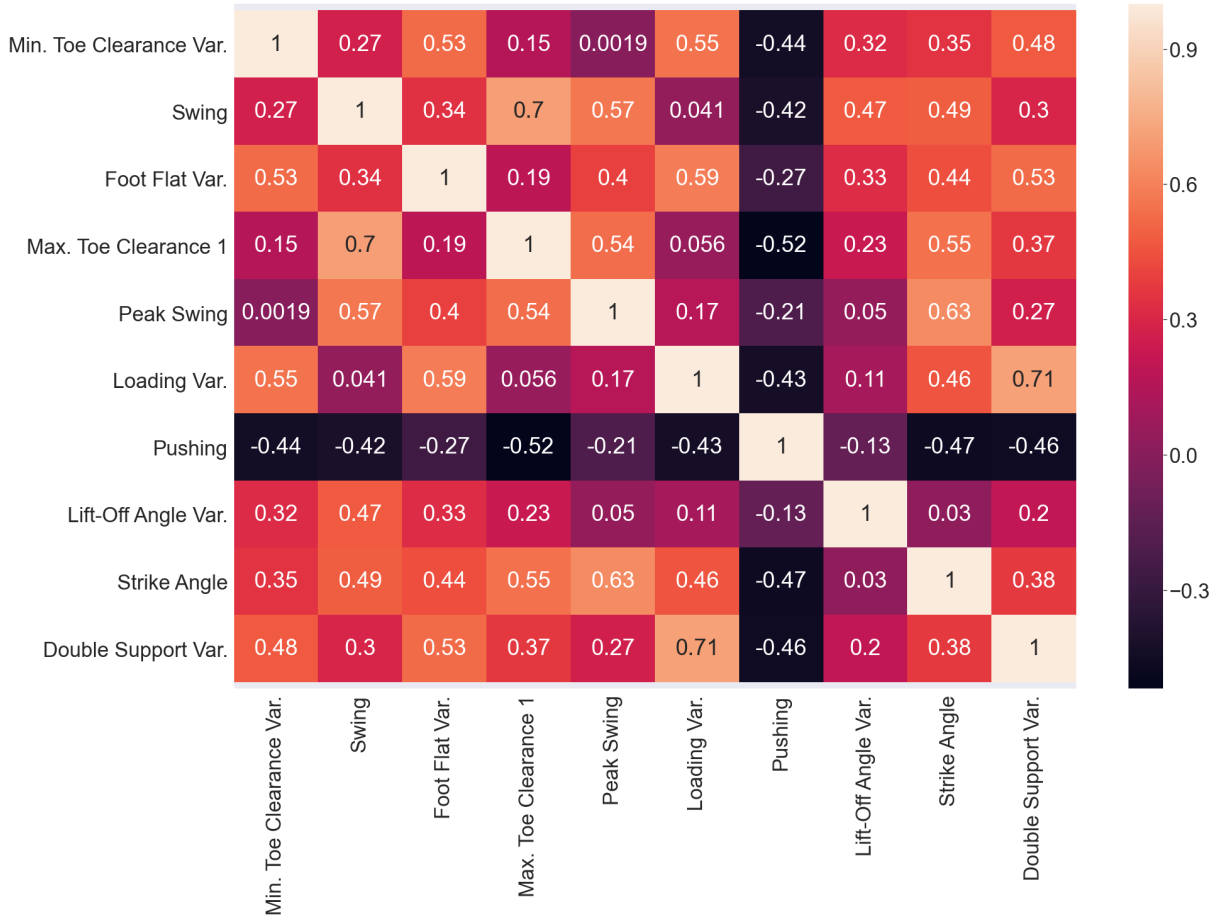

Figure 17: Gait correlation heat map in FD patients with vs without WMLs.
Max.: maximum; Var.: variability; Min.: minimum.

Results are stated in Table 18 and Figure 18. LR selected 7 features as the optimal number of gait characteristics with a F1 score of 72.51%, SVM selected 6 features with a F1 score of 64.13% and RF selected 6 features with a F1 score of 67.79%. Table 14 summarizes the training and validation accuracies for the optimal models of each algorithm. The higher training accuracy is obtained with RF while the higher

validation accuracy is obtained with both LR.

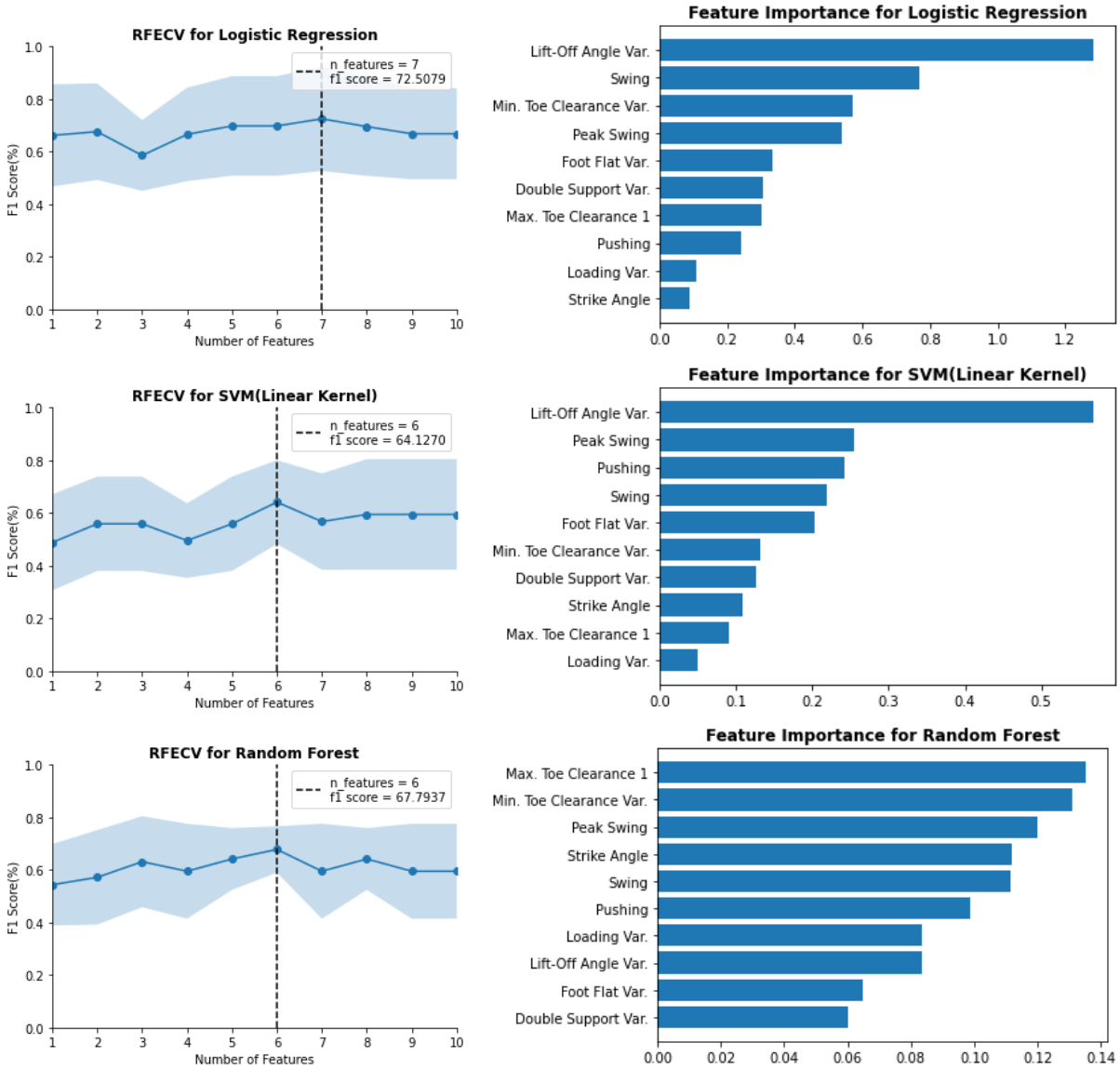

Figure 18: F1 score vs number of features for selection of optimal numbers of gait characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients with vs without WMLs. Recursive features elimination was used through the 5-fold cross-validation (RFECV).
Max.: maximum; Var.: variability; Min.: minimum.

Looking across each gait characteristic in the classification model (Figure 18, right side), the common features for LR, SVM, and RF were 3 (Top 3): minimum toe clearance variability, peak swing, and swing. The common gait characteristics from LR and SVM were 5 (Top 5): lift-off angle variability, minimum toe

Table 18: Results for each of 3 machine learning algorithms (Random Forest, SVM with linear kernel, and Logistic Regression) with recursive feature elimination for FD patients with vs without WMLs.

	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**
LR	83.86 ± 3.59	80.76 ± 5.01
SVM (Linear Kernel)	83.06 ± 1.60	77.90 ± 10.54
RF	83.86 ± 3.68	71.71 ± 15.86

SD: standard deviation

clearance variability, foot flat variability, peak swing, and swing, as shown in Figure 19.

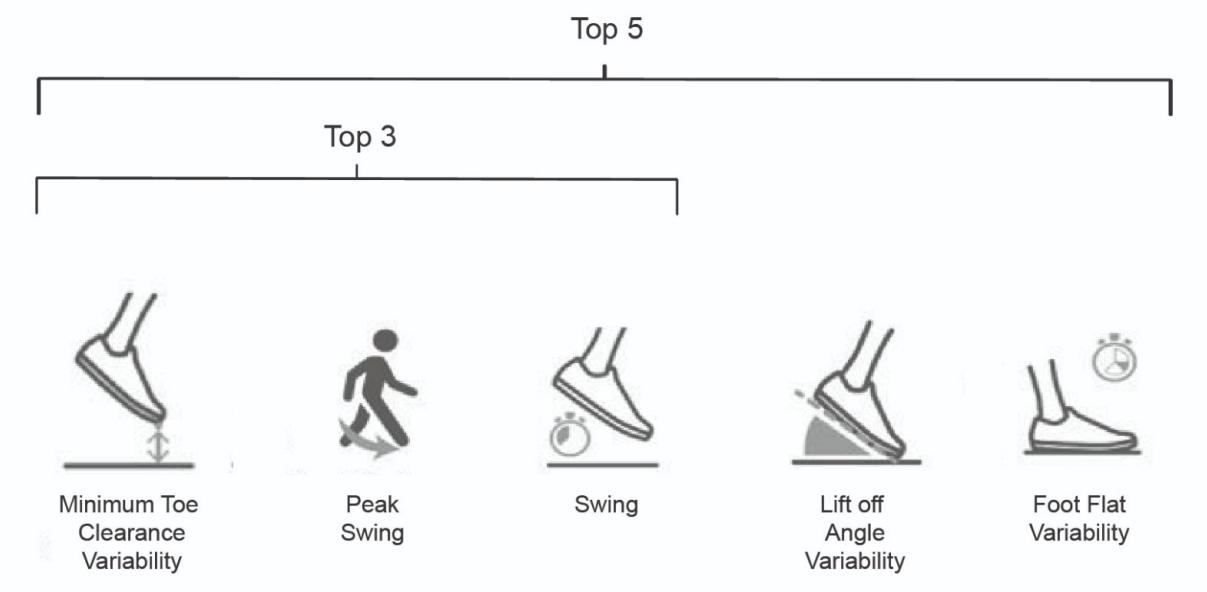

Figure 19: Top 3 and Top 5 after performing RFE on gait features for FD patients with vs without WMLs.

5.5.2 Gait classification based on gait measures

Top 3 and Top 5 were evaluated with five classification models (LR, SVM Linear kernel, SVM RBF kernel, RF, and KNN) to identify the optimal combination of gait characteristics and the classification model with better performance. Results are summarized in Table 19. In both Top 3 and Top 5 from LR

Table 19: Classification accuracy on training and validation data for top common gait characteristics in FD patients with vs without WMLs with LR, SVM (linear and RBF kernel), RF, and KNN.

	Top 3		**Top 5**	
	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Training accuracy% (Mean ± SD)**	**Validation accuracy% (Mean ± SD)**
LR	72.71 ± 5.85	68.57 ± 15.09	84.63 ± 1.94	80.76 ± 5.01
SVM (Linear Kernel)	68.55 ± 4.72	68.38 ± 18.16	80.09 ± 3.82	68.38 ± 12.90
SVM (RBF Kernel)	61.31 ± 1.12	61.52 ± 4.33	61.31 ± 1.12	61.52 ± 4.33
RF	82.32 ± 6.32	71.71 ± 9.40	83.75 ± 6.23	68.86 ± 14.29
KNN	71.72 ± 5.78	67.24 ± 11.35	78.25 ± 4.70	73.90 ± 16.41

LR: Logistic Regression, RBF: radial basis function, RF: Random Forest, KNN: K-Nearest Neighbour, SD: standard deviation

and SVM, LR had the highest validation accuracy with 68.57% and 80.76%, respectively. Reducing the number of features from 5 to 3, LR and KNN had a small decrease, SVM Linear and RBF Kernel remain with the same and RF had an increase in the validation accuracy. Overall, LR showed as the best algorithm achieving a validation accuracy of 72.71% and a training accuracy of 68.57% in Top 3 and a validation accuracy of 84.63%, and a training accuracy of 80.76% in Top 5.

5.5.3 Gait classification based on gait time series

Based on the results obtained in the feature selection the groups of variables selected were: Top LR: lift-off angle, swing, minimum toe clearance, peak swing, foot flat, double support, and maximum toe clearance 1; Top SVM: lift-off angle, peak swing, pushing, swing, foot flat, and minimum toe clearance; Top RF: maximum toe clearance 1, minimum toe clearance, peak swing, strike angle, swing, and pushing; Top 3: peak swing, swing, and minimum toe clearance; and Top 5: lift-off angle, foot flat, peak swing, swing, and minimum toe clearance.

Table 20 shows the results for CNN and LSTM performed on the stride series of the FD patients with vs without WMLs. Top 3 achieved the best performance when using CNN algorithm with a validation accuracy of 81.43% whilst Top RF achieved the best performance for LSTM with a validation accuracy of 73.93%. Comparing CNN and LSTM algorithms, CNN outperformed LSTM for Top LR, Top SVM, Top 3

Table 20: CNN and LSTM results on selected gait features from the gait time series for FD patients with vs without WMLs.

	CNN		**LSTM**	
	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Training accuracy% (Mean ± SD)**	**Validation accuracy% (Mean ± SD)**
Top LR	88.53 ± 11.87	78.93 ± 16.34	87.22 ± 8.81	68.93 ± 10.98
Top SVM	89.19 ± 9.00	71.07 ± 16.88	79.54 ± 8.39	66.43 ± 17.40
Top RF	86.01 ± 14.70	70.71 ± 24.50	78.27 ± 12.32	73.93 ± 14.79
Top 3	95.63 ± 8.75	81.43 ± 14.49	75.08 ± 6.41	71.43 ± 13.27
Top 5	92.40 ± 7.80	75.71 ± 27.26	76.37 ± 7.34	68.93 ± 7.63

CNN: Convolutional Neural Networks, LSTM: Long-Short Term Memory, SD: standard deviation

and Top 5, being only outperformed in Top RF.

Finally, comparing the results from Table 18, Table 19, and Table 20, CNN achieved the highest validation for Top 3 with a validation accuracy of 81.43% and LR achieved the best performance for Top 5 with a validation accuracy of 80.76%. For Top LR, using LR algorithm the validation accuracy was higher with 80.76%. For Top RF, using LSTM achieved the best validation accuracy with 73.93%. Finally, for Top SVM, the better performance was obtained using SVM algorithm with a validation accuracy of 77.90%. Regarding the standard deviation of the validation accuracy, using CNN and LSTM the standard deviation values were slightly higher than with the other algorithms.

5.6 Discussion

Based on previous literature (Aich et al., 2018; Fernandes et al., 2020; Mannini et al., 2016; Pradhan et al., 2015; Rehman et al., 2019; Wahid et al., 2015) different classification models were evaluated with different sets of gait characteristics selected using a filter method based on Mann Whitney U tests and Spearman's correlation between the variables followed by a recursive feature elimination wrapper method with RF, SVM Linear kernel, and LR. Sixteen gait time series were obtained by two wearable sensors. All strides were normalized before developing any ML model according to previous studies (Fernandes

et al., 2020; Mikos et al., 2018; Wahid et al., 2016). Then, for each gait time series the mean and the standard deviation (as a variability measure) were calculated obtaining 32 gait characteristics. From the feature selection analysis, foot flat, pushing, and maximum toe clearance 2 were identified as important characteristics to classify FD with WMLs. While stride length variability, loading variability, and lift-off angle variability, followed by loading, foot flat, and minimum toe clearance were identified as important gait characteristics to distinguish FD without WMLs from aged-matched healthy adults. Previous work (Fernandes et al., 2020) reveals that FD patients (with and without WMLs together) present lower percentages in foot flat and higher in pushing comparing with healthy adults.

For FD patients with WMLs versus controls, validation accuracy of 62-72% and a similar training accuracy of 58-83% was achieved through the five selected classification models based on Top 3 gait characteristics, showing RF classifier the best performance with validation and training accuracy of 72% and 78%, respectively. With one more feature (foot flat variability) RF and LR revealed good performance with an accuracy of 65% and 64% for validation and 70% and 68% for training. These results corroborated the hypothesis that the gait characteristics can be used to distinguish FD patients with WMLs from controls. This goes in line with the premise that gait is a final outcome of WMLs (Snir et al., 2019; Starr, 2003; Zheng et al., 2012).

Surprisingly, in the FD patients without WMLs versus controls classification higher training and validation accuracies of 60-93% and 49-86%, respectively, were obtained based on Top 3, Top 5 or Top 3 LR & SVM. KNN classifier displayed the best performance based on Top 3, with an accuracy of 83% for validation and 92% for training. By increasing the feature set for Top 5 (adding stride length variability and foot flat), overall validation accuracy of 61–87% was achieved, where for the SVM Linear Kernel, KNN, and CNN classifiers the accuracy slightly increased. Finally, reducing the number of features to 3 with Top 3 LR & SVM showed KNN as the best classifier with a top accuracy of 86% for validation and 93% for training. Also, reducing the number of features increased the performance of LR and RF. Similarly, in (Rehman et al., 2019) an increase in the model accuracy was observed with feature reduction.

FD patients with vs without WMLs overall performance across all algorithms stood with 62-81% for validation and 61-96% for training. When using Top 5, LR produced the best performance with an 81% validation accuracy and 85% training. Reducing the number of features to 3 (Top 3) improved the accuracy of the best performance to 81% for validation and 96% for training. Further, feature selection (reduction)

plays an important role to deal with the problem of model overfitting, reduces training time, enhancing the overall ML performance and implementation.

These results suggest that selected gait characteristics could be used as clinical features for supporting diagnoses of FD patients even without WMLs from younger ages since the mean age of these patients is 36.93 ± 10.99 years.

Due to the number of subjects involved in this study, all dataset was used in the training and validation of the models and any independent (external) dataset was used for checking the model performances. To test the robustness of classification models based on the selected gait characteristics further research with independent datasets is needed.

Chapter 6

Electrocardiogram

6.1 Participants and Holter data

One hundred and fourteen FD patients (44 males) were evaluated with a 24-hour ambulatory ECG (Holter) exam. From this group, 61 have WMLs (23 males). The group of patients with WMLs was significantly older than the group of patients without WMLs (Table 21). The age of about half of patients with WMLs ranged from 49 to 67 years and 91 years is the age of the oldest one while the age of 50% of FD patients without WMLs ranged from 34 to 55 years years and the oldest patient has 76 years old (Figure 20).

Table 21: Demographic characteristics for 114 FD patients with and without WMLs.

	With WMLs (n = 61)	**Without WMLs (n = 53)**	**p-value**
Age (years)	55.28 (16.52)	44.13 (14.90)	$<$ **0.001**
Male (%)	38%	40%	0.849

Data is presented as mean (standard deviation) plus p values from Mann Whitney U Test.

Regarding the Holter data, Table 22 shows that there is a significant difference between the two groups on six ECG features: heart rate maximum, QT mean, minimum, and maximum, QT corrected mean, and QT corrected $\geq$ 450.

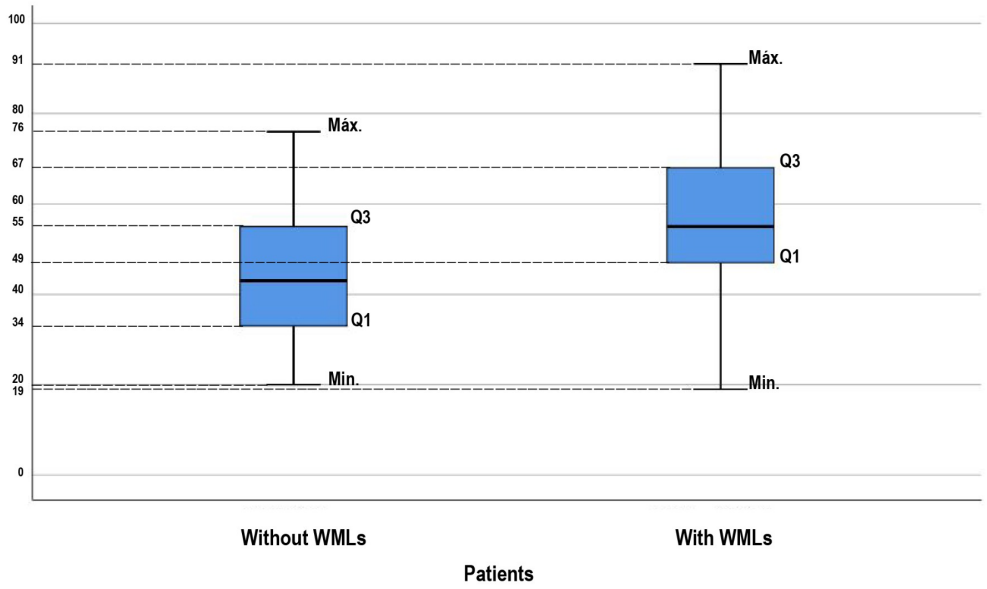

Figure 20: Boxplot of age stratified for the presence of WMLs in FD patients.

Table 22: ECG features for 114 FD patients with and without WMLs.

	With WMLs (n = 61)	**Without WMLs (n = 53)**	**p-value**
HR Min	50.39 (5.88)	50.45 (6.84)	0.75
HR Mean	74.69 (7.32)	77.4 (8.43)	0.0197
HR Max	121.45 (12.5)	134.52 (16.91)	$<$ **0.001**
ASDNN 5	58.83 (22.58)	59.92 (19.18)	0.301
SDANN 5	119.17 (30.07)	120 (41.36)	1
SDNN	137.37 (32.71)	136.56 (44.73)	0.88
RMSSD	58.74 (52.9)	44.09 (29.6)	0.242
QT Min	309.08 (26.91)	292.5 (27.46)	**0.005**
QT Mean	403.12 (33.96)	384.48 (27.04)	**0.003**
QT Max	490.14 (76.93)	466.36 (67.49)	**0.019**
QTc Min	374.19 (34.3)	373.29 (32.07)	0.483
QTc Mean	443.73 (29.48)	430.26 (24.93)	**0.007**
QTc Max	556.33 (69.52)	545.43 (73.44)	0.197
QTc $\geq$ **450**	35.33 (35.34)	19 (27.16)	**0.012**
Longest R-R	1.84 (0.95)	1.61 (0.4)	0.952

Data is presented as mean (standard deviation) plus p-values from T test or Mann Whitney U Test for normally or non-normally distributed continuous variables or Fisher Exact T Test for categorical variables.

From the univariate logistic regression, heart rate maximum, QT minimum, QT mean, QT corrected mean, and QT corrected $\geq$ 450 present a significant association with the presence of WMLs. However, when adjusted by age no ECG features present a significant association with the presence of WMLs (Table 23).

Table 23: Univariate and multivariate (adjusted for age) logistic regression model to predict the presence of WMLs.

	Univariate Logistic Regression	**Adjusted for age**
	p-value	**p-value**
HR Min	0.739	0.467
HR Mean	0.228	0.613
HR Max	$<$ **0.001**	0.210
ASDNN 5	0.487	0.144
SDANN 5	0.827	0.331
SDNN	0.876	0.244
RMSSD	0.098	0.236
QT Min	**0.005**	0.061
QT Mean	**0.003**	0.361
QT Max	0.108	0.376
QTc Min	0.887	0.333
QTc Mean	**0.004**	0.367
QTc Max	0.304	0.362
QTc $\geq$ 450	**0.004**	0.720
Longest R-R	0.304	0.759

So, since age influences the presence of WMLs the group of patients was subgrouped into three age classes: age 19-39 years, age 40-59 years, and age > 59 years (Table 24). In each age class, there was no significant difference in age distribution between FD with WMLs and FD patients without WMLs. However, there is a big difference between the number of patients in age classes 19–39 and 60-91 years. While in subjects aged 19-39 years the number of patients without WMLs is higher in subjects aged 60-91 years the majority have WMLs. Due to this difference, these two age classes were not included in further study.

Focusing only on patients between 40 and 59 years inclusive (age distribution is shown in Figure 21) same tests done earlier were repeated and the results are shown in Table 25 and Table 27.

Table 24: Number of FD patients with and without WMLs per age classes

	With WMLs	Without WMLs	Total	p-value
[19,40[	8	25	33	0.411
[40,60[	23	25	48	0.331
[60,91]	30	3	33	0.683

P-values from Mann Whitney U Test.

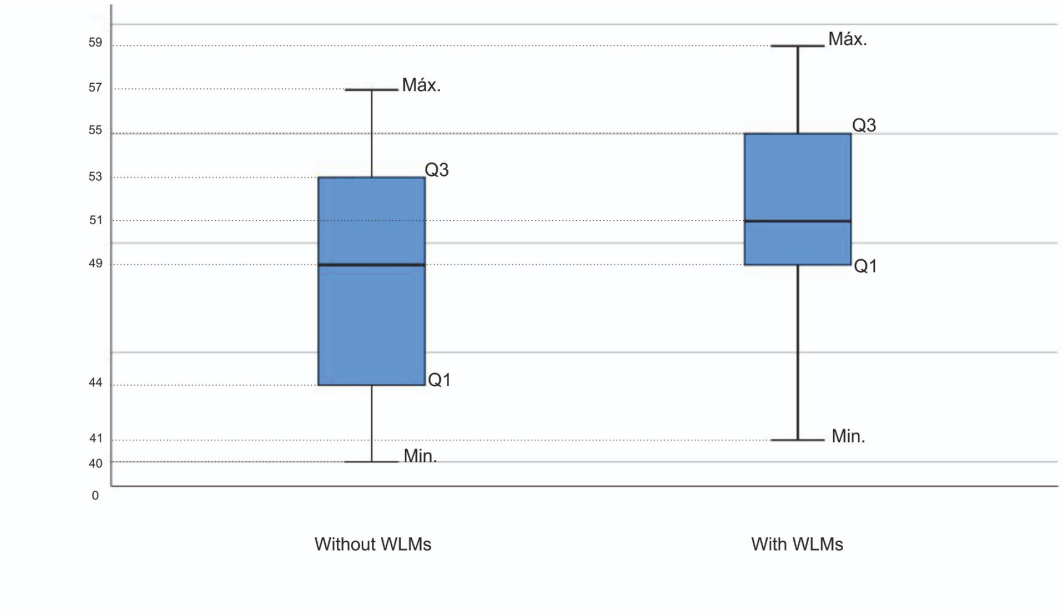

Figure 21: Boxplot of age stratified for the presence of WMLs in FD patients with ages between 40 and 59 years old, inclusive.

Table 25: Demographic characteristics for patients between 40 and 59 years old, inclusive.

	With WMLs (n = 23)	Without WMLs (n = 25)	p-value
Age (years)	50.61 (1.078)	49.00 (1.134)	0.331
Male (%)	35%	40%	0.552

Data is presented as mean (standard deviation) plus p-values from Mann Whitney U Test.

In these subgroups of patients the age and sex differences are not statistically significant (Table 25).

From the univariate logistic regression analysis, two ECG features were significantly associated with the presence of WMLs: heart rate variability SDANN5 and SDNN (Table 26). These two features were further analyzed using multivariate logistic regression (adjusted for age or sex), and both remained as

Table 26: Univariate and multivariate (adjusted for age or sex) logistic regression model to predict the presence of WMLs for patients between 40 and 59 year old, inclusive.

	Univariate Logistic Regression	**Adjusted for age**	**Adjusted for sex**
	p-value	**p-value**	**p-value**
HR Min	0.080	0.067	0.104
HR Mean	0.220	0.316	0.300
HR Max	0.437	0.721	0.589
ASDNN 5	0.149	0.091	0.190
SDANN 5	**0.002**	**0.001**	**0.002**
SDNN	**0.004**	**0.002**	**0.004**
RMSSD	0.151	0.122	0.163
QT Min	0.154	0.231	0.224
QT Mean	0.649	0.882	0.766
QT Max	0.301	0.166	0.281
QTc Min	0.423	0.380	0.393
QTc Mean	0.524	0.414	0.534
QTc Max	0.247	0.169	0.333
QTc $\geq$ 450	0.460	0.300	0.361
Longest R-R	0.329	0.401	0.289

independent predictors of the presence of WMLs in the adjusted models (see Table 27).

Table 27: Association of heart rate variability SDANN5 and SDNN with WMLs in FD patients age ranged from 40 to 59 years.

	With WMLs (n = 23)	**Without WMLs (n = 25)**	**Univariate Logistic Regression**	**Adjusted for age**		**Adjusted for sex**	
			p-value	**p-value**		**p-value**	
SDANN 5	133.18 (7.02)	106.11 (4.61)	**0.002**	group age	**0.001** 0.105	group sex	**0.002** 0.41
SDNN	145.43 (7.28)	119.51 (4.93)	**0.004**	group age	**0.002** 0.052	group sex	**0.004** 0.621

Data is presented as mean (standard deviation).

6.2 Feature selection with RFE

Going further with the analysis of the electrocardiogram, the same feature selection technique used earlier with the gait assessment was employed again to develop a machine learning model that best fits the data. As done earlier, a filter method was performed to select the 10 most significant features (Mann Whitney U Test) not correlated ($|\rho| < 0.80$).

Recursive feature elimination algorithm was performed on the 10 remaining features from the filter method: heart rate minimum, mean and maximum, QT minimum, mean and maximum, QT corrected $\geq$ 450 and minimum, ASDNN 5, and SDANN 5. The heat map of these 10 features selected from the filter method algorithm is shown in Figure 22.

Results are stated in Table 28 and Figure 23. LR selected 6 features as the optimal number of gait characteristics with a F1 Score of 64.94%, SVM selected 6 features with a F1 Score of 60.07% and RF selected 6 features with a F1 Score of 68.12%. Table 28 presents the training and validation accuracies for the optimal models of each algorithm. RF presents the higher training and validation accuracies.

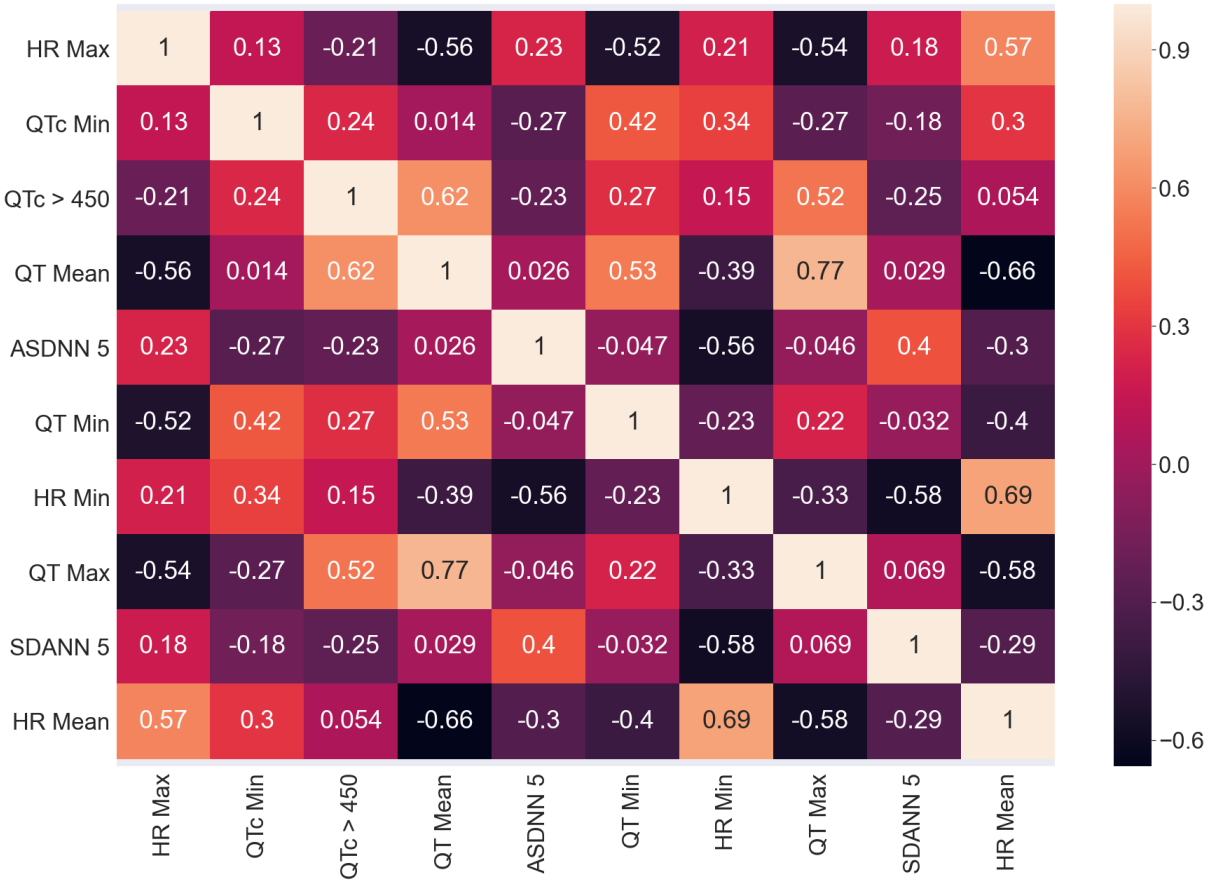

Figure 22: ECG correlation heat map in FD patients.

Table 28: Results for each of 3 machine learning algorithms (LR, SVM with linear kernel and RF) with optimal ECG characteristics' from recursive feature elimination for FD patients with vs without WMLs.

	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**
Logistic Regression	81.71 ± 1.77	79.04 ± 3.66
SVM (Linear Kernel)	78.76 ± 1.12	77.80 ± 2.98
Random Forest	84.75 ± 0.09	79.08 ± 3.42

SD: standard deviation

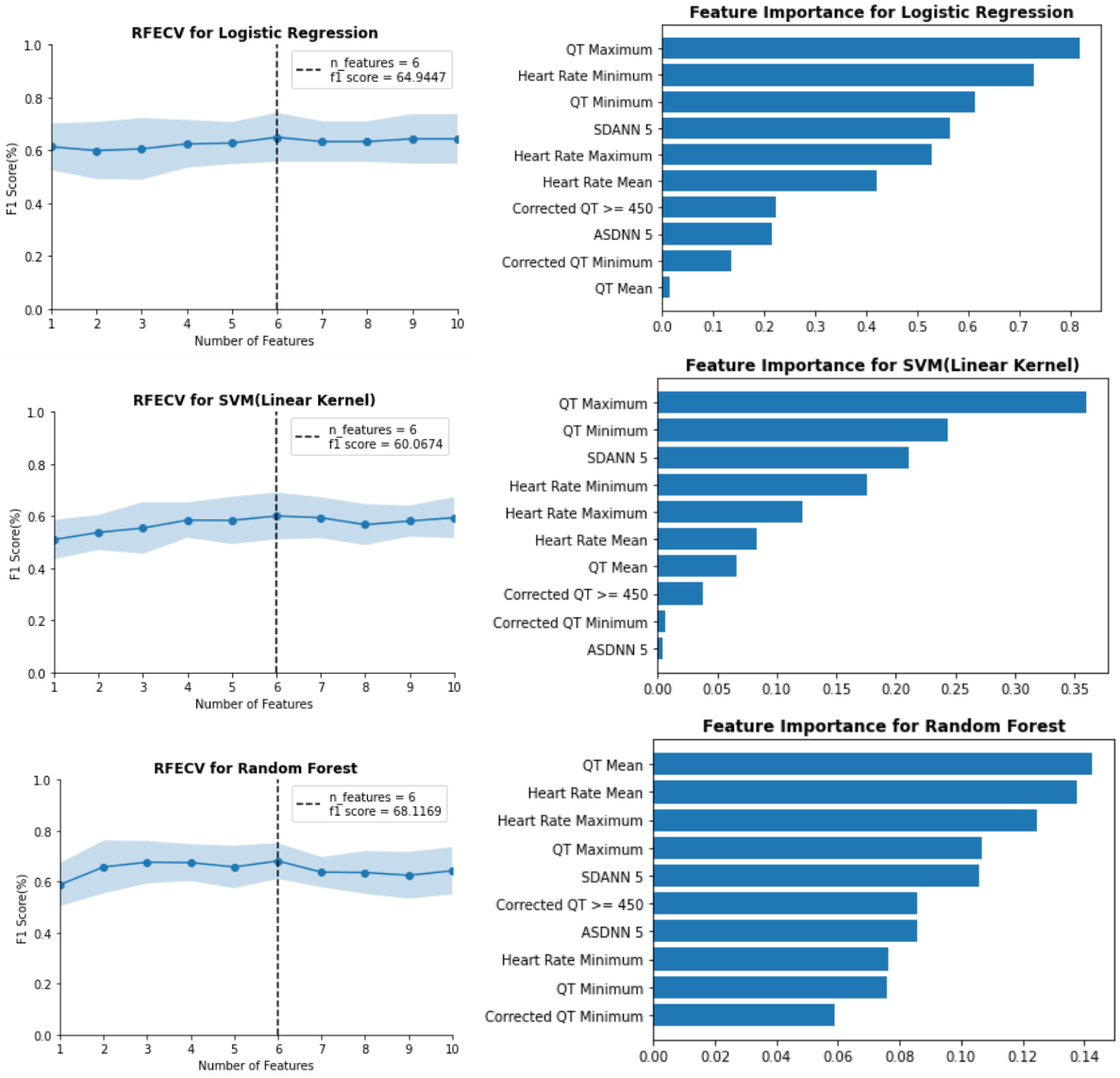

Figure 23: F1 score vs number of features for selection of optimal numbers of ECG characteristics (left) and feature importance results (right) obtained based on Logistic Regression, Support Vector Machine (SVM) Linear kernel and Random Forest in FD patients with vs without WMLs. Recursive features elimination was used through the 5-fold cross-validation (RFECV).
Max.: maximum; Min.: minimum.

6.3 Electrocardiogram classification

Taking into account the contribution of each ECG feature in the classification model (Figure 23, right side), the common features were 4 (Top 4): heart rate mean and maximum, QT maximum, and SDANN 5. Multiple logistic regression analysis revealed that SDANN5 and SDNN features were significantly associated with the presence of WMLs even after adjustment for age or sex (Table 27). This Top 4 and SDANN 5 (since SDANN 5 and SDNN are highly correlated, $\rho = 0.85$) were evaluated with five classification models (LR, SVM Linear kernel, SVM RBF kernel, RF, and KNN) to identify the optimal combination of ECG features and the classification model with better performance. Results are displayed in Table 29. Overall, the validation accuracies were higher when the input was the Top 4, where RF achieved the highest performance with a validation accuracy of 79.72%, followed by KNN with a validation accuracy of 77.75%. LR and SVM (both Linear and RBF kernel) had similar validation accuracy with 75.80%, 76.46%, and 75.17%, respectively. Every one of the 5 models created had a validation accuracy of above 75% with a standard deviation in the range 0.72–3.45%, which means that all models showed a good and consistent performance. Although the validation accuracy values obtained with SDANN 5 as input was lower compared with Top 4, the difference was very small. Expect with RF, the validation accuracy was above 74% with a standard deviation in the range 0.67-0.72.

Table 29: Classification accuracy on training and validation data for top common ECG characteristics and SDANN 5 in FD patients with vs without WMLs with LR, SVM (linear and RBF kernel), RF and KNN.

	Top 4 common features		**SDANN 5**	
	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Training accuracy% (Mean ± SD)**	**Validation accuracy% (Mean ± SD)**
LR	76.91± 1.45	75.80 ± 3.45	74.51 ± 0.23	74.52 ± 0.72
SVM (Linear Kernel)	76.31 ± 0.40	76.46 ± 2.41	74.51 ± 0.16	74.53 ± 0.67
SVM (RBF Kernel)	76.47 ± 0.72	75.17 ± 1.30	74.28 ± 0.44	73.00 ± 0.81
RF	84.10 ± 0.72	79.72 ± 1.40	77.56 ± 0.81	66.10 ± 6.00
KNN	82.18 ± 1.37	77.75 ± 1.57	71.51 ± 0.78	71.37 ± 1.14

LR: Logistic Regression, RBF: radial basis function, RF: Random Forest, KNN: K-Nearest Neighbour, SD: standard deviation

6.4 Gait + Electrocardiogram analysis

To end both gait and ECG study's, a joint model of both was created to check if it would improve their performance. First, the model with the best performance for each dataset was selected: Top LR for gait which achieved a validation accuracy of 80.76% with lift-off angle variability, swing, minimum toe clearance variability, peak swing, foot flat variability, double support variability, and maximum toe clearance 1; and Top 4 with RF for ECG with a validation accuracy of 79.72% and as input the features QT maximum, heart rate mean and minimum, and SDANN 5. These models were re-built without using a group of 9 patients that were assessed on both gait and ECG, that was used only for test purpose. This group of 9 patients was also used to test a model which combined the features mentioned above for both gait's Top LR and ECG's Top 4 with RF, which achieved the best performance, respectively, using LR and RF algorithm. In Table 30 are represented the training, validation accuracy, and testing accuracy for gait's Top LR, ECG's Top 4 with RF, and gait + ECG best model combined using LR and RF. The highest testing accuracy was achieved by gait + ECG using LR with 77.78%, while gait's Top LR, ECG's Top RF, and gait + ECG model using RF had 66.67% accuracy. The SVM Linear Kernel, SVM RBF Kernel, and KNN classifiers were also evaluated with this set of gait + ECG features (the results can be consulted in Table A.39) but the classification accuracies were not superior to LR.

Table 30: Classification accuracy on training and validation data plus testing accuracy for group of 9 patients that were assessed on both gait and ECG with gait's best model (Top LR), ECG's best model (Top RF), Gait ECG best model combined with LR and RF.

	Training accuracy% (Mean ± SD)	**Validation accuracy% (Mean ± SD)**	**Testing accuracy%**
Gait's best model Top LR	88.73 ± 1.50	77.90 ± 5.43	66.67
ECG's best model Top 4 with RF	83.91 ± 1.33	79.31 ± 2.38	66.67
Gait + ECG best model combined with LR	84.09 ± 0.80	79.72 ± 2.92	77.78
Gait + ECG best model combined with RF	93.53 ± 5.46	71.24 ± 11.61	66.67

LR: Logistic Regression, RF: Random Forest

In Table 31, are discriminated the predictions of each model for each individual patient, who are described from Patient 1 to Patient 9.

Table 31: Gait and ECG plus gait + ECG best models prediction for 9 patients selected for test.

Patient	**Age (years)**	**Sex**	**Presence of WMLs**	**Gait's best model Top LR Prediction**	**ECG's best model Top RF Prediction**	**Gait + ECG best model combined**	
						With LR	**With RF**
1	42	Male	Yes	Yes	Yes	Yes	Yes
2	44	Female	No	Yes	No	No	No
3	49	Male	No	No	No	No	No
4	51	Female	No	No	Yes	No	Yes
5	51	Female	No	No	Yes	No	Yes
6	52	Female	Yes	No	Yes	No	Yes
7	52	Male	No	No	No	No	No
8	53	Female	Yes	No	Yes	No	Yes
9	54	Female	Yes	Yes	No	Yes	No

LR: Logistic Regression, RF: Random Forest

6.5 Discussion

The results from statistical analysis using all patients showed that the ECG outcome differences between patients with and without WMLs are affected by age (Table 22 and Table 23). As reported in other studies (Körver et al., 2018b; Rost et al., 2016), the patients with WMLs were significantly older compared to patients without WMLs (mean age 55.3 $\pm$ 16.5 vs 44.1 $\pm$ 14.9 years, p<0.001). The prevalence of WMLs reported in FD corresponds to that found in individuals in the general population, especially in the elderly (Alber et al., 2019; Körver et al., 2018b). With age, people tend to have their heart functions decreased (Strait & Lakatta, 2012) which reflects on different values on the ECG features. To further analysis, participants were divided into age groups: young adulthood (19 to 39 years), middle-aged (40 to 59 years), and older adulthood (59 years and older). Most patients above 60 years old presented WMLs while most of the younger patients did not present WMLs. So the remaining analysis done in this chapter was centered on the middle-aged. In patients aged 40-59 years, SDANN 5 and SDNN were the only ECG features to demonstrate an association with WMLs in logistic regression models (Table 26 and Table 27). The FD patients with WMLs presented higher values in these two heart variability measures compared to FD patients without WMLs (Table 27). Independently of age and gender, SDANN 5 and SDNN were revealed to be good independent predictors of the presence of WMLs. Using only SDANN 5 (since SDANN 5 and SDNN are highly correlated, $\rho = 0.85$) as input a good classification performance was obtained (Table 29). Expect with RF, the validation accuracy was above 74% with a standard deviation in the range 0.67-0.72.

When using Top 4 (heart rate mean and maximum, QT maximum, and SDANN 5) as input, FD patients with vs without WMLs classification overall performance across all algorithms stood with 75-80% for validation and 77-84% for training. RF produced the best performance with an 80% validation accuracy and 84% training. These results suggest that selected electrocardiogram characteristics are associated with the presence of WMLs in FD patients in the middle-aged.

Gait's Top LR: lift-off angle variability, swing, minimum toe clearance variability, peak swing, foot flat variability, double support variability, and maximum toe clearance 1; and ECG's Top 4: QT maximum, heart rate mean and minimum, and SDANN 5 were used to create a joint dataset to test the hypothesis

of improving the accuracy by using gait and ECG together. To test this hypothesis nine patients were left outside as a test group for this hypothesis. Gait plus ECG using LR achieved the best accuracy with 77.78%. Analyzing the results in Table 31, patients 1,3, and 7 were classified correctly by the four classifiers (LR based on gait, RF based on ECG, LR based on gait+ECH, and RF based on gait+ECG). Patients 4, 5, and 9 were classified correctly when gait or gait+ECG as input and LR algorithm were used but these patients were not well classified when ECG or gait+ECG as input and RF algorithm were used. While patients 6 and 8 were classified correctly when ECG or gait+ECG as input and RF algorithm were used but these patients were not well classified when gait or gait+ECG as input and LR algorithm were used. These results suggest that the presence of WMLs may be associated with gait patterns and ECG characteristics, or just one of them. Patient 2 was not correctly classified when gait was used as input but when gait+ECG was used as input this patient was well classified using in both cases LR algorithm. Furthermore, the validation accuracy was higher when gait+ECG as input and LR algorithm was applied. Then, there is some evidence that suggests the use of the two datasets, gait and ECG features, with LR as the classification algorithm to support the identification of WMLs.

Chapter 7

Echocardiogram

7.1 Participants

The dataset used in this section was the result of an echocardiogram on 93 FD patients (38 males). From this group, 49 of the patients have WMLs (25 males). As observed in the previous group of FD patients (see Section 6.1) the patients with WMLs were significantly older than the patients without WMLs (Table 32). Furthermore, sex difference was found between the two groups of patients (Table 32). By observing the boxplot (Figure 24), we find that the two groups of patients have a different distribution of age. While the age of about half of patients with WMLs ranged from 30 to 50 years, the age of 50% of patients with WMLs ranged from 50 to 70 years.

Regarding the echocardiogram data, only two features, LVIDd/SC and A'Lateral, did not present a significant difference between the two groups of patients (Table 33).

Table 32: Demographic characteristics for 93 FD patients with and without WMLs.

	With WMLs (n = 49)	**Without WMLs (n = 44)**	**p-value**
Age (years)	57.24 (14.52)	39.02 (14.62)	**<0.001**
Male (%)	51%	30%	**0.029**

Data is presented as mean (standard deviation) plus p values from Mann Whitney U Test.

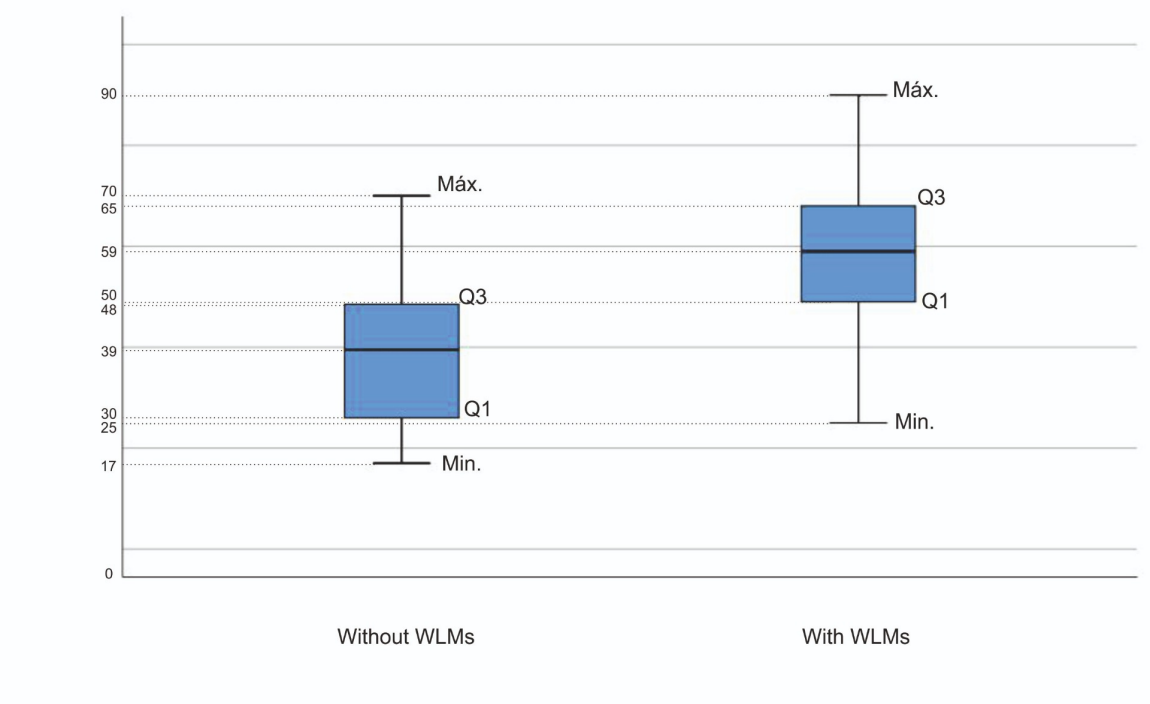

Figure 24: Box plot of FD patients with vs without WMLs according to age.

From the univariate logistic regression, also the same two features did not present statistical significance with the presence WMLs (Table 34). All the features were then adjusted by age and sex. While when adjusting by sex no changes are noticed, adjusting by age, except for LVIDd, all echocardiogram features become no significant predictor of the presence of WMLs.

To decrease this influence of age, the group of patients was subgrouped into the same three age classes as in Section 6.1 (Table 34). There was no significant differences in age between the two group of patients in all three age classes. However, the number of patients in group of patients in age classes 19–39 and 60-91 are very different. Therefore, these two age classes were not included in further study. Focusing only on patients between 40 and 59 years inclusive (age distribution is shown in Figure 25) the univariate logistic regression analysis was performed again. The results are shown in Table 36. Only MV E/A Ratio, MV A Vel and S' Lateral were significantly associated with the presence of WMLs and just the association of S' Lateral with the presence of WMLs retained significance in the adjusted models for age. Furthermore, in the three models adjusted for age, the variable age was still a significant predictor of WMLs.

Table 33: Echocardiogram features for 93 FD patients with and without WMLs.

	With WMLs (n = 49)	Without WMLs (n = 44)	p-value
MV E/A Ratio	1.01 (0.56)	1.48 (0.53)	**<0.001**
MV A Vel	0.81 (0.19)	0.60 (0.18)	**<0.001**
MV Dec T	245.58 (63.34)	202.08 (36.74)	**<0.001**
MV E Vel	0.74 (0.17)	0.82 (0.13)	**0.014**
E' Lateral	9.36 (4.32)	14.16 (5.21)	**<0.001**
E' Septal	7.37 (3.25)	10.70 (3.68)	**<0.001**
E/E' Lateral	9.27 (3.47)	6.53 (2.66)	**<0.001**
E/E' Medial	10.56 (3.96)	7.86 (2.98)	**<0.001**
E/E' Septal	11.72 (5.11)	8.73 (3.51)	**0.001**
LVPWd	11.19 (3.29)	9.04 (2.36)	**<0.001**
ISVd	12.55 (4.18)	9.82 (3.32)	**0.001**
LVIDd	42.74 (4.56)	46.07 (3.57)	**0.001**
LADiam/SC	21.69 (3.55)	19.52 (2.92)	**0.001**
AoDiam	32.44 (3.69)	30.13 (4.23)	**0.006**
S' Lateral	8.47 (2.67)	9.75 (2.82)	**0.013**
LVdMassInd ASE	108.67 (47.35)	86.16 (37.70)	**0.020**
LADiam	37.06 (5.76)	34.29 (5.51)	**0.024**
S' Septal	7.06 (1.85)	7.80 (1.63)	**0.027**
A' Septal	9.24 (2.08)	8.34 (2.11)	**0.036**
A' Lateral	10.43 (2.48)	9.48 (2.87)	0.108
LVDdMass ASE	108.07 (82.25)	153.78 (71.95)	**0.040**
LVIDd/SC	25.11 (2.79)	26.06 (2.42)	0.063

Data is presented as mean (standard deviation) plus significance on T test, Mann Whitney U Test for normally or non-normally distributed continuous variables and Fisher Exact T Test for categorical variables.

Table 34: Univariate and multivariate (adjusted for age) logistic regression model to predict the presence of WMLs.

	Univariate Logistic Regression	Adjusted for age	Adjusted for sex
	p-value	p-value	p-value
MV E/A Ratio	**<0.001**	0.774	**<0.001**
MV A Vel	**<0.001**	0.146	**<0.001**
MV Dec T	**<0.001**	0.222	**<0.001**
MV E Vel	**0.015**	0.745	**0.022**
E' Lateral	**<0.001**	0.879	**<0.001**
E' Septal	**<0.001**	0.711	**<0.001**
E/E' Lateral	**<0.001**	0.663	**<0.001**
E/E' Medial	**<0.001**	0.868	**0.001**
E/E' Septal	**0.02**	0.711	**0.004**
LVPWd	**0.001**	0.787	**0.001**
ISVd	**0.001**	0.861	**0.003**
LVIDd	**<0.001**	**0.016**	**<0.001**
LADiam/SC	**0.002**	0.926	**0.001**
AoDiam	**0.006**	0.857	**0.017**
S' Lateral	**0.027**	0.270	**0.044**
LVdMasInd ASE	**0.014**	0.366	**0.033**
LADiam	**0.021**	0.819	**0.036**
S' Septal	**0.044**	0.293	0.081
A' Septal	**0.040**	0.296	0.054
A' Lateral	0.090	0.810	0.130
LVDdMass ASE	**0.048**	0.299	0.120
LVIDd/SC	0.084	0.241	0.113

Table 35: Division of patients according to respective ages plus age significance value for each age interval performed with Mann Whitney U Test.

	With WMLs	Without WMLs	Total	p-value
[17,40[	5	22	27	0.416
[40,60[	24	19	43	0.082
[60,91[	20	3	23	0.404

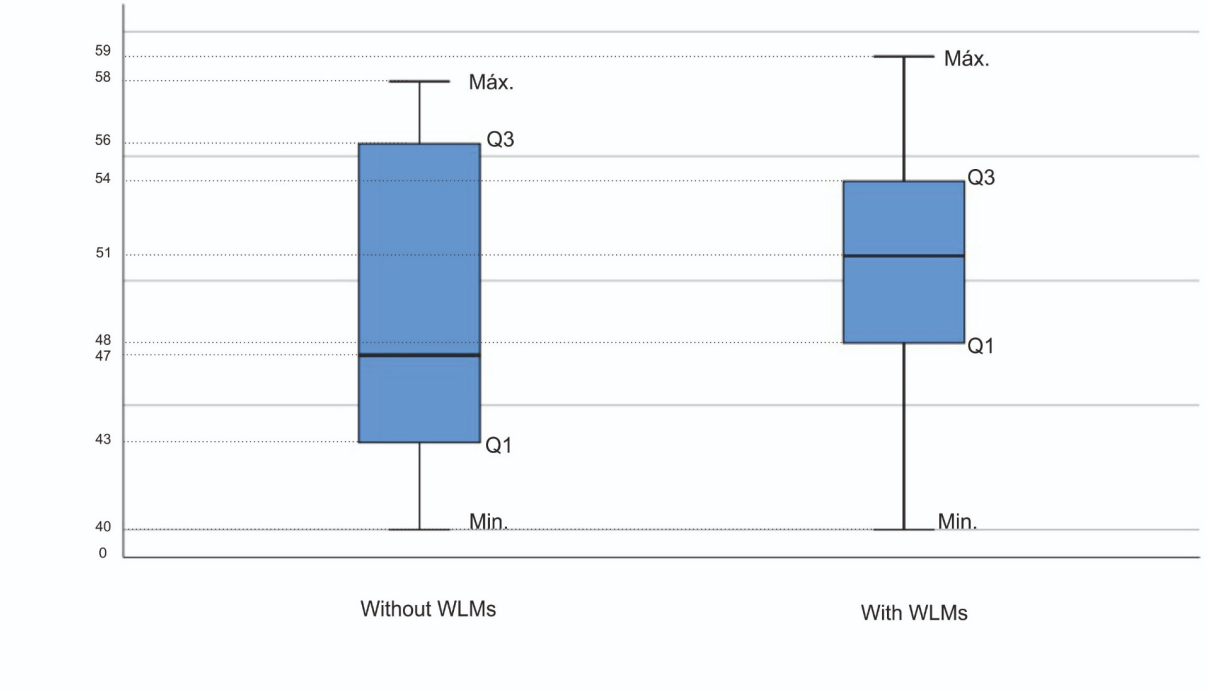

Figure 25: Box plot of patients with vs without WMLs in the interval of ages between 40 and 59 years, inclusive.

Table 36: Echocardiogram features presented as mean $\pm$ standard deviation, statistically significant after performing univariate logistic regression and adjusted logistic regression for age to predict the presence of WMLs for patients between 40 and 59 years, inclusive.

	With WMLs	**Without WMLs**	**Univariate Logistic Regression**	**Adjusted for age**		**Adjusted for sex**	
MV E\A Ratio	1.02 ± 0.33	1.31 ± 0.42	**0.017**	group age	0.068 **0.013**	group sex	**0.003** 0.091
MV A Vel	0.79 ± 0.18	0.68 ± 0.17	**0.04**	group age	0.142 **0.039**	group sex	**<0.001** 0.219
S' Lateral	9.25 ± 2.88	8.42 ± 2.35	**0.027**	group age	**0.039** **<0.001**	group sex	**<0.001** 0.309

7.2 Discussion

In agreement with the results presented in the previous chapter, age was found to be significantly associated with the presence of WMLs. In the group of patients analyzed in this chapter, the patients with WMLs were 18 years older than the patients without WMLs on average. As mentioned before, with age, people tend to have their heart functions decreased (Strait & Lakatta, 2012) which also reflects on different values on the echocardiogram features. The results (Table 33) show that the difference in echocardiogram features between the two FD patients groups with WMLs and without WMLs are due to the age difference. Furthermore, after age stratification, in the group aged from 40 to 59 years, the results (Table 36) reveal that the values of the echocardiogram features are still affected by age.

From this first analysis, the echocardiogram features not reveal to be good predictors of the presence of WMLs in FD patients. Then, it is not expected to get higher classification performance with the echocardiogram features comparing with the performance achieved previously with gait and ECG data. However, further research is needed to confirm this hypothesis. But, as the echocardiogram and ECG data are from two different groups of FD patients, this analysis is left for future work.

Chapter 8

Conclusion, limitations, and future work

8.1 General conclusions

The major goals of this dissertation were to evaluate the effectiveness of machine learning methods for the differential diagnosis of WMLs in FD patients based on a dataset from one of three different sources: gait evaluation, electrocardiogram exam (Holter), or echocardiogram exam, and, to investigate how the integration of information from more than one source can improve the predictive system. To achieve these aims, several steps were performed including the evaluation of a MR normalization strategy for the de-correlation of physical properties, speed, stride length, and gait variables, the implementation of various feature selection methods using hybrid methods, the implementation and evaluation of different classification algorithms based on different datasets.

To the best of our knowledge, this was the first study that explores gait characteristics and cardiac data and their discriminate power in FD patients with WMLs from FD patients without WMLs.

Using gait data, for the discrimination of FD patients with WMLs from age-matched healthy adults the model with higher accuracy (72%) was achieved with RF classifier with the variables foot flat, pushing, and cycle duration as input. For the discrimination of FD patients without WMLs from age-matched healthy adults, the best-suited model was achieved using KNN based on the variables loading, foot flat, stride length variability, loading variability, and lift-off angle variability, with an accuracy of 87%. Finally, to

discriminate the presence of WMLs in FD patients using gait characteristics, the models that fit best were using CNN based on the time series of the variables minimum toe clearance, swing, and peak swing, with an accuracy of 82%, and using LR algorithm based on 5 gait measures lift-off angle variability, swing, minimum toe clearance variability, peak swing, and foot flat variability, with an accuracy of 81%.

When using ECG features for the differential diagnosis of the presence of WMLs in FD patients the best-suited model was using RF with QT maximum, heart rate mean and maximum, and SDANN 5, with an accuracy of 80%.

To analyze if the use of both gait and ECG features as input could improve the identification of the presence of WMLs, the best models described above: LR algorithm based on gait measures and RF algorithm based on ECG features were evaluated on a test group of nine patients. Six patients were well classified when the LR algorithm was performed based on gait or based on gait+ECG. However, with the RF algorithm based on ECG features 6 patients were well classification and this number increase to 7 when gait+ECG was used as input. This result shows some evidence that the use of both datasets could improve the performance of identification of WMLs in FD patients.

The results from logistic regression analysis revealed that age was independently associated with the presence of WMLs, and the echocardiogram features that revealed some significant association with the presence of WMLs showed to be affected by age. Then, it is not expected to get higher classification performance with the echocardiogram features comparing with the performance achieved previously with gait and ECG data. However, further research is needed to confirm this hypothesis.

8.2 Limitations

Several limitations were found across this study. Firstly, the developed MR models were based on a fairly small number of control subjects (n = 37), still comparable to other previous works (Mikos et al., 2018; Wahid et al., 2016; Wahid et al., 2015). A larger amount of subjects along with more independent variables might improve the robustness of the feature selection methods by reducing the correlation among variables and, therefore, improve the effectiveness of the classifiers. To validate the best set of gait

variables found by the feature selection methods further research with a higher number of subjects is required. For last, gait, ECG, and echocardiogram datasets were not collected for all the same patients, which made it impossible to develop a classification model with the three datasets as input.

8.3 Future work

The findings reported in this thesis are the first step to demonstrate the potential of machine learning techniques based on gait and cardiac variables as a complementary tool to understand the role of WMLs in the gait impairment of FD. For future research, a larger sample size will be used to confirm and extend these findings.

The implications of WMLs on gait compromise in FD or predictive value of each kinematic gait variable remain still elusive, warranting further investigation with a more enriched cohort.

Due to the number of subjects involved in this study, in most of the cases, all subjects were used in the training and validation of the models, and only in Section 6.4, a small independent group test was used for checking the model performances. To test the robustness of classification models based on the selected gait and cardiac characteristics further research with a larger independent group test is needed.

Since, in this study, the classification of FD was only evaluated based on gait, to go further with this study of the presence or absence of FD, is required to analyze: FD patients vs controls with ECG data, FD patients vs controls with echocardiogram data, and, finally, FD patients vs controls with gait + ECG + echocardiogram.

Part IV :

Appendix

Table 37: Value of gait assessment of 17 gait variables for 39 FD patients with vs without WMLs, 25 FD patients with WMLs vs controls and 14 FD patients without WMLs vs controls.

		With WML (n = 24)	Without WML (n = 15)	p-value	With WML (n = 24)	Control (n = 24)	p-value	Without WML (n = 15)	Control (n = 15)	p-value
Speed	**Mean**	1.28	1.44	**0.02***	1.28	1.28	0.992	1.44	1.41	0.511
	SD	0.056	0.063	0.3	0.056	0.056	0.801	0.063	0.067	0.635
	CV	4.45	4.399	0.92	4.45	4.34	0.691	4.399	4.71	0.401
Cycle duration	**Mean**	1.02	0.99	0.53	1.02	1.015	0.977	0.99	1.04	0.15
	SD	0.026	0.023	0.28	0.026	0.023	0.273	0.023	0.0199	0.427
	CV	2.5	2.28	0.35	2.5	2.22	0.225	2.28	1.92	0.210
Cadence	**Mean**	118.24	121.41	0.53	118.24	118.8	0.961	121.41	116.04	.15
	SD	2.89	2.73	0.55	2.89	2.63	0.318	2.73	2.22	0.194
	CV	2.46	2.24	0.39	2.46	2.21	0.248	2.24	1.91	0.210
Stride length	**Mean**	1.28	1.41	**0.01***	1.28	1.28	0.808	1.41	1.44	0.571
	SD	0.038	0.042	0.44	0.038	0.041	0.304	0.042	0.53	**0.035**
	CV	3.05	2.97	0.94	3.05	3.21	0.421	2.97	3.65	**0.019**
Stance	**Mean**	60.9	58.69	**0.004**	60.9	59.92	**0.043**	58.69	59.06	0.511
	SD	1.04	1.45	0.11	1.04	1.34	0.240	1.45	1.33	0.603
	CV	1.7	2.48	0.067	1.7	2.25	0.190	2.48	2.26	0.635
Swing	**Mean**	39.1	41.31	**0.004**	39.1	40.08	**0.043**	41.31	1.84	.511
	SD	1.04	1.45	0.11	1.04	1.34	0.240	1.45	1.33	0.603
	CV	2.68	3.48	0.13	2.68	3.34	0.337	3.48	3.24	0.635
Loading	**Mean**	12.17	14.62	**0.03**	12.17	12.62	0.705	14.62	12.03	0.061
	SD	1.15	1.63	**0.017**	1.15	1.38	0.108	1.63	1.12	0.089
	CV	9.15	11.29	0.13	9.15	10.82	0.080	11.29	9.21	0.401
Foot flat	**Mean**	53.52	49.44	**0.03**	53.52	55.38	0.299	49.44	53.28	**0.033***
	SD	1.91	2.11	0.38	1.91	2.36	**0.041**	2.11	2.15	0.946
	CV	3.66	4.34	0.15	3.66	4.36	0.204	4.34	4.14	0.839
Pushing	**Mean**	34.31	35.94	0.24	34.31	31.998	0.138	35.94	34.69	0.306
	SD	1.68	1.84	0.39	1.68	2.08	0.322	1.84	2.095	0.667
	CV	4.92	5.17	**0.041**	4.92	6.62	0.165	5.17	6.08	0.376
Double support	**Mean**	21.22	18.36	**0.03***	21.22	20.04	0.282	18.36	18.97	0.667
	SD	1.59	2.1	0.11	1.59	1.84	0.485	2.1	1.87	0.454
	CV	7.67	12.02	**0.041***	7.67	9.61	0.233	12.02	10.03	0.401
Peakswing	**Mean**	390.11	435.8	**0.016***	390.11	387.74	0.764	435.8	405.87	**0.041***
	SD	16.74	17.57	0.76	16.74	15.87	0.372	17.57	17.70	0.946
	CV	4.35	4.04	0.72	4.35	4.11	0.467	4.04	4.39	0.804
Strike angle	**Mean**	23.51	28.88	**0.004***	23.51	24.58	0.567	28.88	26.24	0.094
	SD	1.28	1.4	0.35	1.28	1.34	0.352	1.4	1.20	0.150
	CV	5.82	4.86	0.22	5.82	5.59	0.915	4.86	4.64	0.667
Lift-off angle	**Mean**	-65.94	-74.07	**0.017***	-65.94	-64.11	0.399	-74.07	-73.56	0.874
	SD	1.76	1.55	0.25	1.76	1.71	0.869	1.55	1.93	0.077
	CV	-2.73	-2.095	**0.038***	-2.73	-2.75	0.705	-2.095	-2.65	0.085
Maximum heel clearance	**Mean**	0.23	0.26	**0.005**	0.23	0.25	0.290	0.26	0.27	0.037
	SD	0.0075	0.0086	0.15	0.0075	0.0082	0.243	0.0086	0.011	0.062*
Maximum toe clearance 1	**Mean**	0.063	0.071	0.24	0.063	0.71	**0.029**	0.071	0.068	0.910
	SD	0.006	0.0073	0.12	0.006	0.0062	0.884	0.0073	0.0087	0.306
Minimum toe clearance	**Mean**	0.032	0.028	0.43	0.032	0.031	0.961	0.028	0.021	0.227
	SD	0.0044	0.0053	0.098	0.0044	0.0047	0.668	0.0053	0.0069	0.085
Maximum toe clearance 2	**Mean**	0.13	0.17	**0.003**	0.13	0.14	0.318	0.17	0.17	0.769
	SD	0.0082	0.0096	0.098	0.0082	0.0086	0.365	0.0096	0.0089	0.482

Data is represented in arithmetic mean, standard deviation and coefficient of variation plus statistical significance on non parametric Mann Whitney U Test for the right foot. *Statistically significant for the left foot.

Table 38: Multiple linear regression models using only significative independent variables for each gait variable on the left foot.

		MR normalization						AIC	Adjusted R^2
		Age	Weight	Height	Speed	Gender	Stride length		
Spatial-temporal variables									
Cycle duration	= 1.081	-0.0011		0.219	-0.274			-124.37	0.658
Cadence	= 110.572	0.132		-24.879	31.301			220.30	0.697
Stride length	= 0.301	-0.002		0.301	0.494			-116.39	0.893
Stance	= 66.496				-4.547	-1.109		150.62	0.229
Swing	= 33.504				4.547	1.109		150.62	0.229
Loading	= 16.522		-0.141		3.629	2.476		167.79	0.359
Foot flat	= 62.227	0.032	0.169		-15.921			195.63	0.604
Pushing	= 23.327	-0.031	-0.069		12.131			190.28	0.463
Double support	= 32				-8.804	-1.454		197.81	0.207
Peak swing	= 243.705			-22.362	137.775	4.883		338.19	0.589
Foot clearance variables									
Strike angle	= 13.508		-0.153			3.227	15.6465	182.29	0.600
Lift-off angle	= -42.221	0.179					-25.831	224.27	0.705
Maximum heel clearance	= 0.161	-0.0003				0.039	0.0718	-166.19	0.560
Maximum toe clearance 1	= 0.048	0.00026				0.014		-192.54	0.159
Minimum toe clearance	= 0.007	0.00042						-231.63	0.455
Maximum toe clearance 2	= 0.032	-0.0006				0.0289	0.106	-186.18	0.759

Table 39: Classification accuracy on training and validation data plus testing accuracy for a group of 9 patients that were assessed on both gait and ECG with Gait ECG best model combined with SVM Linear Kernel, SVM RBF Kernel, and KNN.

	Training accuracy% (Mean $\pm$ SD)	Validation accuracy% (Mean $\pm$ SD)	Testing accuracy%
Gait + ECG best model combined with SVM Linear Kernel	77.12 $\pm$ 3.33	72.14 $\pm$ 15.71	66.67
Gait + ECG best model combined with SVM RBF Kernel	77.78 $\pm$ 3.47	67.50 $\pm$ 20.47	55.56
Gait + ECG best model combined with KNN	80.59 $\pm$ 5.06	78.21 $\pm$ 13.19	66.67

SVM: Support Vector Machine, KNN: K-Nearest Neighbor

Bibliography

Aich, S., Choi, K.-W., Pradhan, P. M., Park, J., & Kim, H.-C. (2018). A Performance Comparison Based on Machine Learning Approaches to Distinguish Parkinson's Disease from Alzheimer Disease Using Spatiotemporal Gait signals. *Advanced Science Letters*, *24*(3), 2058–2062. https://doi.org/10.1166/asl.2018.11847

Alber, J., Alladi, S., Bae, H. J., Barton, D. A., Beckett, L. A., Bell, J. M., Berman, S. E., Biessels, G. J., Black, S. E., Bos, I., Bowman, G. L., Brai, E., Brickman, A. M., Callahan, B. L., Corriveau, R. A., Fossati, S., Gottesman, R. F., Gustafson, D. R., Hachinski, V., … Hainsworth, A. H. (2019). White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities. *Alzheimer's and Dementia: Translational Research and Clinical Interventions*, *5*, 107–117. https://doi.org/10.1016/j.trci.2019.02.001

Alcock, L., Galna, B., Perkins, R., Lord, S., & Rochester, L. (2018). Step length determines minimum toe clearance in older adults and people with Parkinson's disease. *Journal of Biomechanics*, *71*, 30–36. https://doi.org/10.1016/j.jbiomech.2017.12.002

Allen-Zhu, Z., Li, Y., & Song, Z. (2018). On the convergence rate of training recurrent neural networks.

Atul Mehta, Michael Beck, & Sunder-Plassmann, G. (2006). Fabry disease: perspectives from five years of FOS. *Oxford PharmaGenesis*.

Azevedo, O., Gago, M. F., Miltenberger-Miltenyi, G., Sousa, N., & Cunha, D. (2020). Fabry Disease Therapy: State-of-the-Art and Current Challenges. *International Journal of Molecular Sciences*, *22*(1), 206. https://doi.org/10.3390/ijms22010206

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buechner, S., Moretti, M., Burlina, A. P., Cei, G., Manara, R., Ricci, R., Mignani, R., Parini, R., Di Vito, R., Giordano, G. P., Simonelli, P., Siciliano, G., & Borsini, W. (2008). Central nervous system involvement in Anderson-Fabry disease: A clinical and MRI retrospective study. *Journal of Neurology, Neurosurgery and Psychiatry*, *79*(11), 1249–1254. https://doi.org/10.1136/jnnp.2008.143693

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. https://doi.org/10.1177/0049124104268644

Caballero, L., Kou, S., Dulgheru, R., Gonjilashvili, N., Athanassopoulos, G. D., Barone, D., Baroni, M., Cardim, N., De Diego, J. J. G., Oliva, M. J., Hagendorff, A., Hristova, K., Lopez, T., Magne, J., Martinez, C., De La Morena, G., Popescu, B. A., Penicka, M., Ozyigit, T., ... Lancellotti, P. (2015). Echocardiographic reference ranges for normal cardiac Doppler data: Results from the NORRE Study. *European Heart Journal Cardiovascular Imaging*, *16*(9), 1031–1041. https://doi.org/10.1093/ehjci/jev083

Caramia, C., Torricelli, D., Schmid, M., Munoz-Gonzalez, A., Gonzalez-Vargas, J., Grandas, F., & Pons, J. L. (2018). IMU-Based Classification of Parkinson's Disease from Gait: A Sensitivity Analysis on Sensor Location and Feature Selection. *IEEE Journal of Biomedical and Health Informatics*, *22*(6), 1765–1774. https://doi.org/10.1109/JBHI.2018.2865218

Desnick, R. J., Brady, R., Barranger, J., Collins, A. J., Germain, D. P., Goldman, M., Grabowski, G., Packman, S., & Wilcox, W. R. (2003). Fabry disease, an under-recognized multisystemic disorder: Expert recommendations for diagnosis, management, and enzyme replacement therapy. https://doi.org/10.7326/0003-4819-138-4-200302180-00014

El Missiri, A. M., El Meniawy, K. A. L., Sakr, S. A. S., & deen Mohamed, A. S. E. (2016). Normal reference values of echocardiographic measurements in young Egyptian adults. https://doi.org/10.1016/j.ehj.2016.01.002

Faust, O., & Ng, E. Y. (2016). Computer aided diagnosis for cardiovascular diseases based on ECG signals: A survey. https://doi.org/10.1142/S0219519416400017

Fernandes, C., Ferreira, F., Gago, M., Azevedo, O., Sousa, N., Erlhagen, W., & Bicho, E. (2020). Gait classification of patients with Fabry's disease based on normalized gait features obtained using multiple regression models, 2288–2295. https://doi.org/10.1109/bibm47256.2019.8983241

Fernandes, C., Ferreira, F., Lopes, R. L., Bicho, E., Erlhagen, W., Sousa, N., & Gago, M. F. (2021). *Discrimination of idiopathic parkinson's disease and vascular parkinsonism based on gait time series and levodopa effect* [In press].

Fernandes, C., Fonseca, L., Ferreira, F., Gago, M., Costa, L., Sousa, N., Ferreira, C., Gama, J., Erlhagen, W., & Bicho, E. (2019). Artificial Neural Networks Classification of Patients with Parkinsonism based on Gait. *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, 2024–2030. https://doi.org/10.1109/BIBM.2018.8621466

Ferreira, F., Gago, M. F., Bicho, E., Carvalho, C., Mollaei, N., Rodrigues, L., Sousa, N., Rodrigues, P. P., Ferreira, C., & Gama, J. (2019). Gait stride-to-stride variability and foot clearance pattern analysis in Idiopathic Parkinson's Disease and Vascular Parkinsonism. *Journal of Biomechanics*, *92*, 98–104. https://doi.org/10.1016/j.jbiomech.2019.05.039

Forte, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. https://doi.org/10.3389/fnins.2019.00710

G. Jignesh Chowdary, Suganya. G, P. M. (2020). EFFECTIVE PREDICTION OF CARDIOVASCULAR DISEASE USING CLUSTER OF MACHINE LEARNING ALGORITHMS. *Journal of Critical Reviews*, *7*(18), 2192–2201.

Galluzzi, S., Nicosia, F., Geroldi, C., Alicandri, A., Bonetti, M., Romanelli, G., Zulli, R., & Frisoni, G. B. (2009). Cardiac autonomic dysfunction is associated with white matter lesions in patients with mild cognitive impairment. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *64*(12), 1312–1315. https://doi.org/10.1093/gerona/glp105

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*.

Giugliani, R., Niu, D. M., Ramaswami, U., West, M., Hughes, D., Kampmann, C., Pintos-Morell, G., Nicholls, K., Schenk, J. M., & Beck, M. (2016). A 15-year perspective of the fabry outcome survey. https://doi.org/10.1177/2326409816666298

Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks (Studies in Computational Intelligence)* (Doctoral dissertation).

Grünfeld, J. P. (2003). How to improve the early diagnosis of Fabry's disease? *Kidney International*, *64*(3), 1136–1137. https://doi.org/10.1046/j.1523-1755.2003.00196.x

Gu, Q., Zhu, L., & Cai, Z. (2009). Evaluation measures of the classification performance of imbalanced data sets. *Communications in Computer and Information Science*, *51*, 461–471. https://doi.org/10.1007/978-3-642-04962-0_53

Gudivada, V. N., Ding, J., & Apon, A. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software*, *10*(1), 1–20.

Haykin, S. (2009). *Neural networks and learning Third Edition* (Vol. 127).

Hijazi, S., Page, A., Kantarci, B., & Soyata, T. (2016). Machine Learning in Cardiac Health Monitoring and Decision Support. *Computer*, *49*(11), 38–48. https://doi.org/10.1109/MC.2016.339

Hua, J., Tembe, W. D., & Dougherty, E. R. (2009). Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition*, *42*(3), 409–424. https://doi.org/10.1016/j.patcog.2008.08.001

Iguyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. https://doi.org/10.1162/153244303322753616

Joseph, F. (2017). *A study on Deep Machine Learning Algorithms for diagnosis of diseases* (tech. rep.).

Katz, M. H. (2011). *Multivariable analysis: A practical guide for clinicians and public health researchers*. Cambridge University Press. https://doi.org/10.1017/CBO9780511974175

Kirkwood, B. R., & Sterne, J. A. (2003). *Essential Medical Statistics*.

Körver, S., Vergouwe, M., Hollak, C. E., van Schaik, I. N., & Langeveld, M. (2018a). Development and clinical consequences of white matter lesions in Fabry disease: a systematic review. https://doi.org/10.1016/j.ymgme.2018.08.014

Körver, S., Vergouwe, M., Hollak, C. E., van Schaik, I. N., & Langeveld, M. (2018b). Development and clinical consequences of white matter lesions in Fabry disease: a systematic review. https://doi.org/10.1016/j.ymgme.2018.08.014

Koushik, J. (2016). Understanding Convolutional Neural Networks.

Kramer, O. (2013). Dimensionality Reduction with Unsupervised Nearest Neighbors. *Intelligent Systems Reference Library*, *51*. https://doi.org/10.1007/978-3-642-38652-7

Kubota, K. J., Chen, J. A., & Little, M. A. (2016). Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement Disorders*, *31*(9), 1314–1326. https://doi.org/10.1002/mds.26693

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Löhle, M., Hughes, D., Milligan, A., Richfield, L., Reichmann, H., Mehta, A., & Schapira, A. H. (2015). Clinical prodromes of neurodegeneration in Anderson-Fabry disease. *Neurology*, *84*(14), 1454–1464. https://doi.org/10.1212/WNL.0000000000001450

MacDermot, KD and Holmes, A., & Miners. (2001). Anderson-Fabry disease: clinical manifestations and impact of disease in a cohort of 98 hemizygous males. *Journal of medical genetics*, *38*, 769–775.

Maity, N. G., & Das, S. (2017). Machine learning for improved diagnosis and prognosis in healthcare. *IEEE Aerospace Conference Proceedings*. https://doi.org/10.1109/AERO.2017.7943950

Mannini, A., Trojaniello, D., Cereatti, A., & Sabatini, A. (2016). A Machine Learning Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington's Disease Patients. *Sensors*, *16*(1), 134. https://doi.org/10.3390/s16010134

Mauer, M., & Kopp, J. (2018). Fabry disease: Treatment.

Mikos, V., Yen, S. C., Tay, A., Heng, C. H., Chung, C. L. H., Liew, S. H. X., Tan, D. M. L., & Au, W. L. (2018). Regression analysis of gait parameters and mobility measures in a healthy cohort for subject-specific normative values. *PLoS ONE*, *13*(6), e0199215. https://doi.org/10.1371/journal.pone.0199215

Morgan, S. H., Cheshire, J. K., Wilson, T. M., MacDermot, K., & d.A. Crawfurd, M. (1987). Anderson-Fabry disease-family linkage studies using two polymorphic X-linked DNA probes. *Pediatric Nephrology*, *1*(3), 536–539. https://doi.org/10.1007/BF00849266

Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M., Neil, D., Saffari, A., Mead, R., Hautbergue, G. M., Holbrook, J. D., & Ferraiuolo, L. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. https://doi.org/10.1038/s41582-020-0377-8

Papernot, N., Faghri, F., Carlini, N., Goodfellow, I., Feinman, R., Kurakin, A., Xie, C., Sharma, Y., Brown, T., Roy, A., Matyasko, A., Behzadan, V., Hambardzumyan, K., Zhang, Z., Juang, Y.-L., Li, Z., Sheatsley, R., Garg, A., Uesato, J., ... McDaniel, P. (2016). Technical Report on the CleverHans v2.1.0 Adversarial Examples Library.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot* (tech. rep.).

Pradhan, C., Wuehr, M., Akrami, F., Neuhaeusser, M., Huth, S., Brandt, T., Jahn, K., & Schniepp, R. (2015). Automated classification of neurological disorders of gait using spatio-temporal gait parameters. *Journal of Electromyography and Kinesiology*, *25*(2), 413–422. https://doi.org/10.1016/j.jelekin.2015.01.004

Prakash, C., Kumar, R., & Mittal, N. (2018). Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review*, *49*(1), 1–40. https://doi.org/10.1007/s10462-016-9514-6

Prashant Gupta. (2017). Decision Trees in Machine Learning – Towards Data Science.

Ramaswami, U., Whybra, C., Parini, R., Pintos-Morell, G., Mehta, A., Sunder-Plassmann, G., Widmer, U., & Beck, M. (2006). Clinical manifestations of Fabry disease in children: Data from the Fabry Outcome Survey. *Acta Paediatrica, International Journal of Paediatrics*, *95*(1), 86–92. https://doi.org/10.1080/08035250500275022

Rehman, R. Z. U., Del Din, S., Guan, Y., Yarnall, A. J., Shi, J. Q., & Rochester, L. (2019). Selecting Clinically Relevant Gait Characteristics for Classification of Early Parkinson's Disease: A Comprehensive Machine Learning Approach. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-53656-7

Rijnbeek, P. R., Van Herpen, G., Bots, M. L., Man, S., Verweij, N., Hofman, A., Hillege, H., Numans, M. E., Swenne, C. A., Witteman, J. C., & Kors, J. A. (2014). Normal values of the electrocardiogram for ages 16-90 years. *Journal of Electrocardiology*, *47*(6), 914–921. https://doi.org/10.1016/j.jelectrocard.2014.07.022

Rost, N. S., Cloonan, L., Kanakis, A. S., Fitzpatrick, K. M., Azzariti, D. R., Clarke, V., Lourenco, C. M., Germain, D. P., Politei, J. M., Homola, G. A., Sommer, C., Üçeyler, N., & Sims, K. B. (2016). Determinants of white matter hyperintensity burden in patients with Fabry disease. *Neurology*, *86*(20), 1880–1886. https://doi.org/10.1212/WNL.0000000000002673

Satriano, A., Afzal, Y., Sarim Afzal, M., Fatehi Hassanabad, A., Wu, C., Dykstra, S., Flewitt, J., Feuchter, P., Sandonato, R., Heydari, B., Merchant, N., Howarth, A. G., Lydell, C. P., Khan, A., Fine, N. M., Greiner, R., & White, J. A. (2020). Neural-Network-Based Diagnosis Using 3-Dimensional Myocardial Architecture and Deformation: Demonstration for the Differentiation of Hypertrophic Cardiomyopathy. *Frontiers in Cardiovascular Medicine*, *7*, 584727. https://doi.org/10.3389/fcvm.2020.584727

Snir, J. A., Bartha, R., & Montero-Odasso, M. (2019). White matter integrity is associated with gait impairment and falls in mild cognitive impairment. Results from the gait and brain study. *NeuroImage: Clinical*, *24*. https://doi.org/10.1016/j.nicl.2019.101975

Starr, J. M. (2003). Brain white matter lesions detected by magnetic resosnance imaging are associated with balance and gait speed. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(1), 94–98. https://doi.org/10.1136/jnnp.74.1.94

Strait, J. B., & Lakatta, E. G. (2012). Aging-Associated Cardiovascular Changes and Their Relationship to Heart Failure. https://doi.org/10.1016/j.hfc.2011.08.011

Sunder-Plassmann, G., & Födinger, M. (2006). Diagnosis of Fabry disease: the role of screening and case-finding studies.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, *4*(January), 3104–3112.

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the Variance In flation Factor and Other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology*, *39*(2), 81–90. https://doi.org/10.1080/01973533.2016.1277529

Tsai, C. H., Ma, H. P., Lin, Y. T., Hung, C. S., Hsieh, M. C., Chang, T. Y., Kuo, P. H., Lin, C., Lo, M. T., Hsu, H. H., Peng, C. K., & Lin, Y. H. (2019). Heart Rhythm Complexity Impairment in Patients with Pulmonary Hypertension. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-47144-1

Umetani, K., Singer, D. H., McCraty, R., & Atkinson, M. (1998). Twenty-four hour time domain heart rate variability and heart rate: Relations to age and gender over nine decades. *Journal of the American College of Cardiology*, *31*(3), 593–601. https://doi.org/10.1016/S0735-1097(97)00554-8

Wahid, F., Begg, R., Lythgo, N., Hass, C. J., Halgamuge, S., & Ackland, D. C. (2016). A multiple regression approach to normalization of spatiotemporal gait features. *Journal of Applied Biomechanics*, *32*(2), 128–139. https://doi.org/10.1123/jab.2015-0035

Wahid, F., Begg, R. K., Hass, C. J., Halgamuge, S., & Ackland, D. C. (2015). Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE Journal of Biomedical and Health Informatics*, *19*(6), 1794–1802. https://doi.org/10.1109/JBHI.2015.2450232

Zheng, J. J., Lord, S. R., Close, J. C., Sachdev, P. S., Wen, W., Brodaty, H., & Delbaere, K. (2012). Brain white matter hyperintensities, executive dysfunction, instability, and falls in older people: A prospective cohort study. https://doi.org/10.1093/gerona/gls063