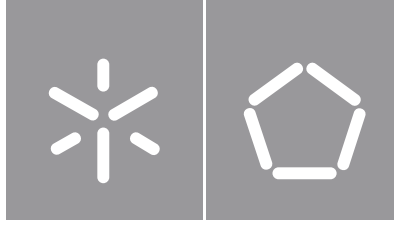


Universidade do Minho
Escola de Engenharia

Ana Rita Sousa Moura

Detecting important electrocardiogram characteristics for the diagnosis of Fabry disease via statistics and machine learning techniques



Universidade do Minho

Escola de Engenharia

Ana Rita Sousa Moura

**Detecting important electrocardiogram
characteristics for the diagnosis of Fabry
disease via statistics and machine learning
techniques**

Master's dissertation
Integrated Master's in Biomedical Engineering
Medical electronics

Work elaborated under the supervision of
Prof. Dr. Estela Bicho (Supervisor)
Dr. Flora Ferreira (Co-supervisor)

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial-SemDerivações

CC BY-NC-ND

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

0.1 Agradecimentos

Tenho que agradecer a algumas pessoas, pelo apoio que me proporcionaram, de diversas formas, durante a minha dissertação e todo o meu percurso académico.

Em primeiro lugar, agradeço à professora Estela Bicho por me ter aceite neste projeto e pela sua disponibilidade e à Doutora Flora Ferreira pela ajuda e orientação durante toda a dissertação, pela confiança, pela disponibilidade e pelos conhecimentos que me transmitiu. Agradeço também ao Doutor Miguel Gago e à Doutora Olga Azevedo, por terem lançado o desafio, fornecido os dados e ajudado nas questões clínicas que surgiram neste trabalho, sem eles esta dissertação não seria possível.

Aos meus pais e ao meu irmão pelos esforços que fizeram para me possibilitarem esta oportunidade, por me apoiarem e quererem sempre o melhor para a minha vida. Sem eles não tinha conseguido chegar aqui. Obrigada por tudo Mãe.

Um grande obrigada a todos os meus amigos e familiares, que durante estes anos da minha vida, estando perto ou longe, me acompanharam e incentivaram, em especial à Bárbara, à Helena, à Mariana, ao Nuno e à prima Carolina. Ao meu afilhado por alegrar os meus dias e à minha madrinha.

Não podia deixar de agradecer ao meu namorado, Tiago, que durante a minha dissertação e durante o curso, que percorremos juntos, esteve sempre presente para me ajudar, motivar e acima de tudo aturar. Esta conquista também é um pouco tua.

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.

0.2 Resumo

Título: Detecção de características do electrocardiograma importantes para o diagnóstico da doença de Fabry utilizando estatística e técnicas de *machine learning*.

A doença de Fabry (DF) é uma doença genética rara de depósito lisossômico que afeta a qualidade de vida e pode mesmo levar à morte prematura. O diagnóstico ainda tardio reduz a eficiência da terapia de reposição enzimática. Assim, é extremamente importante identificar biomarcadores que possam auxiliar no diagnóstico precoce da DF. Apesar dos progressos nos últimos anos, a DF continua a ser mal compreendida. Desta forma, esta dissertação teve como objetivo identificar características de electrocardiograma (ECG) importantes para o diagnóstico da DF, mais especificamente, para diferenciar os pacientes de DF com e sem lesões da matéria branca, e estes de pacientes com miocardiopatia hipertrófica sarcomérica. Para este fim, foram desenvolvidos e aplicados modelos estatísticos e de *machine learning* (ML).

Quinze características de ECG foram avaliadas usando diversos métodos de inferência estatística, sobretudo para identificar diferenças significativas entre os grupos tendo em conta o sexo e a idade. Dois métodos de seleção de atributos foram aplicados, um baseado nos valores do fator de inflação da variância (VIF), e uma eliminação recursiva de atributos usando *logistic regression*, *support vector machine* (SVM) linear e *random forest* como classificadores. Depois, avaliou-se a performance de cinco algoritmos de ML - *logistic regression*, SVM com função de base radial (RBF), *random forest* e *k-nearest neighbor* (KNN) - a distinguir os diferentes grupos, para identificar o melhor modelo para cada problema de classificação.

A idade revelou ser significativamente diferente entre os grupos e estar relacionada com as variáveis de ECG. Após subdividir em três faixas etárias, algumas características revelaram-se importantes para distinguir os grupos, não sendo afetadas pela idade. Com base nas características de ECG selecionadas, os resultados mostraram boa taxa de acerto, com os melhores resultados a variar de 80% a 85%, obtidos usando SVM RBF, *random forest* e KNN.

Estas descobertas demonstram o potencial das técnicas de ML baseadas em características de ECG como ferramenta complementar ao diagnóstico da DF, com e sem lesões da matéria branca, o que pode ser útil para reduzir a demora no diagnóstico.

Palavras-chave: Classificação, Doença de Fabry, Eletrocardiograma, Estatística, *Machine Learning*.

0.3 Abstact

Title: Detecting important electrocardiogram characteristics for the diagnosis of Fabry disease via statistics and machine learning techniques.

Fabry disease (FD) is a rare genetic lysosomal storage disorder that affects life quality and may even lead to premature death. The mean delay between the onset of symptoms and the diagnosis is still very high, reducing enzyme replacement therapy's efficiency. Therefore, it is extremely important to identify biomarkers that could assist in the early diagnosis of FD. Despite all the progress in recent years, FD remains misunderstood. Thus, this dissertation aimed to identify important electrocardiogram (ECG) characteristics for diagnosing FD, and more specifically, to differentiate FD with white matter lesions (WMLs) from FD without WMLs patients and these from patients with sarcomeric hypertrophic cardiomyopathy. To this end, statistics and machine learning (ML) models have been developed and applied.

Fifteen ECG variables were evaluated using several statistical inference methods mainly to identify significant differences between groups, considering sex and age. Two feature selection methods were applied, one based on variance inflation factor (VIF) values, and a recursive feature elimination using logistic regression, linear support vector machine (SVM), and random forest classifiers. Then, the performance of five ML algorithms - logistic regression, linear SVM, radial basis function (RBF) SVM, random forest, and K-nearest neighbor (KNN) - at distinguishing the different groups were evaluated to identify the best model for each classification problem.

Age was found to be significantly different between groups and to be related to the ECG variables values. After subdivision into three categories by age, some ECG characteristics were identified as important to distinguish the groups and unaffected by age. Based on selected ECG characteristics, the results showed good classification accuracies, with the best scores ranging from 80% to 85%, obtained using RBF SVM, random forest, or KNN.

These findings demonstrate the potential of ML techniques based on ECG characteristics as a complementary tool for diagnosing FD with or without WMLs that could be useful to reduce the diagnosis delay.

Keywords: Classification, Electrocardiogram, Fabry Disease, Machine Learning, Statistics.

Contents

0.1	Agradecimientos	iii
0.2	Resumo	v
0.3	Abstact	vi
List of Figures		xi
List of Tables		xii
Acronyms		xvi
 I : Introduction and State of the Art		 1
1	Introduction	2
1.1	Motivation	2
1.2	Goals	3
1.3	Dissertation Structure	3
2	Fabry disease and hypertrophic cardiomyopathy disease	5
2.1	Fabry Disease	5
2.1.1	Description	6
2.1.2	Phenotype and Clinical Manifestations	6
2.1.3	Diagnosis	8
2.1.4	Treatment	8
2.2	Hypertrophic Cardiomyopathy	9

2.2.1	Description	9
2.2.2	Phenotype and Clinical Manifestations	10
2.2.3	Diagnosis	11
2.2.4	Treatment	12
3	Electrocardiogram and machine learning diagnostic value	13
3.1	Electrocardiogram characteristics as data for classification	13
3.2	ECG-based machine learning algorithms for classification	18
II :	Materials and Methods	21
4	Study data	22
4.1	Study Sample	22
4.2	24-Hour Holter	23
5	The implemented statistical methods	26
5.1	Descriptive statistics	26
5.2	Statistical inference: Hypothesis testing	27
5.2.1	Normality	27
5.2.2	Homogeneity of variances	28
5.2.3	Differences between independent samples: quantitative data	29
5.2.4	Differences between independent samples: qualitative data	33
5.3	Logistic Regression Analysis	35
6	The implemented Machine Learning methods	37
6.1	Cross-validation	39
6.2	Hyperparameter Tuning	40
6.3	Feature Selection	41
6.3.1	Variance Inflation Factor	42
6.3.2	Recursive Feature Elimination	43

6.4	Machine Learning Classification Algorithms	44
6.4.1	Logistic Regression	46
6.4.2	Support Vector Machines	47
6.4.3	Random Forest	48
6.4.4	K-Nearest Neighbors	50
6.5	Performance Metrics	51
III :	Results, Discussion, Conclusion, and Future Work	55
7	Statistical Results	56
7.1	Results from patients over 18 years old	56
7.2	Results from patients aged 40 to 59 years (inclusive)	64
7.3	Results from patients over 59 years old	69
7.4	Discussion	71
8	Machine Learning Results	74
8.1	Results from patients aged 40 to 59 years (inclusive)	74
8.2	Results from patients over 59 years old	83
8.3	Discussion	86
9	Conclusion, limitations and future work	90
9.1	General conclusions	90
9.2	Limitations and future work	91
	Bibliography	103
	Appendices	105
A	Statistical Analysis Appendix	105

List of Figures

1	Methodology machine learning-based flowchart	37
2	Example of support machine classification results	47
3	Decision tree structure	49
4	Confusion Matrix structure	52
5	Example of ROC curve and AUC	53
6	RFECV results for the different machine learning models of the HCM vs FD without WMLs patients aged 40 to 59 (inclusive)	76
7	RFECV results for the different machine learning models of the HCM vs FD with WMLs patients aged 40 to 59 (inclusive)	79
8	RFECV results for the different machine learning models of the FD with WMLs vs FD without WMLs patients aged 40 to 59 (inclusive)	81
9	RFECV results for the different machine learning models of the HCM vs FD with WMLs patients aged above 59.	85

List of Tables

1	Demographic characteristics of the study sample	23
2	Patients count per group and age class	23
3	24-Hour Holter variables	24
4	Heart rate and QT analysis normal reference range of values.	25
5	Heart rate and heart rate variability normal reference range of values.	25
6	Variables summary and comparison of means/medians between HCM, FD with WMLs and FD without WMLs groups	58
7	Statistically significant logistic regression results of HCM and FD without WMLs groups .	61
8	Statistically significant logistic regression results of HCM and FD with WMLs groups . . .	62
9	Statistically significant logistic regression Results of FD without WMLs and FD with WMLs groups	63
10	Variables summary and comparison of means/medians between HCM, FD with WMLs and FD without WMLs patients aged 40 to 59 (inclusive)	65
11	Statistically significant logistic regression results of HCM and FD without WMLs patients aged 40 to 59 (inclusive)	67
12	Statistically significant logistic regression results of HCM and FD with WMLs patients aged 40 to 59 (inclusive)	68
13	Statistically significant logistic regression results of FD without WMLs and FD with WMLs patients aged 40 to 59 (inclusive)	68
14	Variables summary and comparison of means/medians for HCM and FD with WMLs patients aged above 59	70

15	Statistically significant logistic regression results of HCM and FD with WMLs groups with ages above 59	71
16	VIF values for patients aged 40 to 59 (inclusive)	75
17	Models's classification performance metrics per number of features for HCM and FD without WMLs patients aged 40 to 59 (inclusive)	77
18	Models's classification performance metrics per number of features for HCM and FD with WMLs patients aged 40 to 59 (inclusive)	80
19	Models's classification performance metrics per number of features for FD with WMLs and FD without WMLs patients aged 40 to 59 (inclusive)	83
20	VIF values for patients aged above 59	84
21	Models's classification performance metrics per number of features for HCM and FD with WMLs patients aged above 59	86
22	Normality of HCM, FD with WMLs and FD without WMLs groups and homogeneity of variance results	105
23	Statistically non-significant logistic regression results of HCM and FD without WMLs groups	106
24	Statistically non-significant logistic regression results of HCM and FD with WMLs groups .	107
25	Statistically non-significant logistic regression results of FD with WMLs and FD without WMLs groups	108
26	Normality of HCM, FD with WMLs and FD without WMLs groups and homogeneity of variance results, for patients aged between 40 and 59 (including)	109
27	Statistically non-significant logistic regression results of HCM and FD without WMLs groups with ages between 40 and 59 (including)	110
28	Statistically non-significant logistic regression results of HCM and FD with WMLs groups with ages between 40 and 59 (including)	111
29	Statistically non-significant logistic regression results of FD with WMLs and FD without WMLs groups with ages between 40 and 59 (including)	112
30	Normality of HCM and FD with WMLs groups and homogeneity of variance results, for patients aged above 59	113

31	Statistically non-significant logistic regression results of HCM and FD with WMLs groups with ages above 59	114
32	Hyperparameters values/status per machine learning model and classification problem for patients aged between 40 and 59 (inclusive)	115
33	Accuracy values per machine learning model and classification problem for patients aged between 40 and 59 (inclusive)	116
34	Hyperparameters values/status per machine learning model and number of features for HCM and FD without WMLs patients aged 40 to 59 (inclusive)	117
35	Hyperparameters values/status per machine learning model and number of features for HCM and FD with WMLs patients aged 40 to 59 (inclusive)	118
36	Hyperparameters values/status per machine learning model and number of features for FD with WMLs and FD without WMLs patients aged 40 to 59 (inclusive)	119
37	Hyperparameters values/status per machine learning model and classification problem for patients aged above 59	120
38	Accuracy values per machine learning model and classification problem for patients aged above 59	120
39	Hyperparameters values/status per machine learning model and number of features for HCM and FD with WMLs patients aged above 59	121

Acronyms

ANOVA Analysis of variance.

AUC Area under curve.

ECG Electrocardiogram.

ERT Enzyme replacement therapy.

FD Fabry disease.

GLA Galactosidase alpha.

HCM Hypertrophic cardiomyopathy.

HR Heart rate.

HRV Heart rate variability.

KNN K-nearest neighbor.

LOOCV Leave-one-out cross-validation.

LR Logistic regression.

LVH Left ventricular hypertrophy.

ML Machine learning.

MRI Magnetic resonance imaging.

OR Odds ratio.

RBF Radial basis function.

RF Random forest.

RFE Recursive feature elimination.

SVM Support vector machine.

VIF Variance inflation factor.

WMLs White matter lesions.

Part I :

Introduction and State of the Art

Chapter 1

Introduction

This Chapter presents the motivation of this dissertation as well as its goals. Finally, a description of how the dissertation is structured is presented.

1.1 Motivation

Fabry disease (FD) is a rare genetic disease that affects life quality and may lead to premature death. There are treatments for this disease, like Enzyme replacement therapy (ERT) [1] [2]. Although this treatment has shown good results, it may not be able to stabilize disease manifestations if it is already in an advanced stage, so it is essential to begin ERT as early as possible [1]. The mean delay between the symptom onset and diagnosis is around 14.7 and 15.1 years in male and female patients, respectively [3]. Therefore, it is extremely important to find biomarkers that could assist in an early diagnosis of Fabry disease.

Despite all the progress in recent years, FD remains misunderstood. There is evidence that specific cardiovascular, neurological, and cerebrovascular manifestations may be characteristics of this disease [4]. Hence, it is important to identify significant electrocardiogram (ECG) characteristics for diagnosing FD, and differentiate FD with white matter lesions (WMLs) from FD without WMLs patients.

In this dissertation, 24-hour ECG data (Holter) is used since Electrocardiogram (ECG) is a medical

exam that is usually performed to diagnose and follow patients with cardiac diseases, like FD [5]. The cohort of patients may have or not have white matter lesions. Patients with Hypertrophic Cardiomyopathy are also included in the study because this disease has similarities in cardiac manifestations to those found in FD, mainly left ventricular hypertrophy.

Statistics and machine learning models are developed and applied to extract and select important information from clinical data to understand better and achieve a more accurate disease prediction [6].

1.2 Goals

The main goal of this dissertation is to detect important electrocardiogram characteristics for the diagnosis of Fabry disease, using for that statistics and machine learning methods.

More precisely, the objectives of this dissertation are:

- Detect important ECG characteristics to differentiate FD patients with white matter lesions (WMLs) from FD patients without WMLs, and these two groups from patients with sarcomeric hypertrophic cardiomyopathy (HCM);
- Investigate which physical characteristics, like age and sex, are correlated with ECG characteristics and, therefore, could impact the classification results;
- Implement and evaluate the performance of different Machine Learning models on the differentiation between the groups of patients.

1.3 Dissertation Structure

This dissertation is divided into three Parts and, in turn, these Parts are divided into Chapters, then Sections and Subsections.

The first Part (Part I) is denoted “Introduction and State of Art”, which, as indicated by the name, includes an introduction where the motivation, goals and structure (current subsection) of this dissertation are presented. It also includes a Chapter relative to Fabry and Hypertrophic Cardiomyopathy diseases, in which, for each disease, a description, its phenotype and clinical manifestations, how it is diagnosed, and possible treatments, are presented. The value of electrocardiogram (ECG) characteristics as data for classification is also explored in this part, mainly through studies that have already used ECG characteristics to diagnose and distinguish diseases. Machine learning diagnosis value is also explained and several studies that use ECG-based machine learning models for classification problems are also presented.

In Part II the materials and methods are delineated. The study data is first presented, followed by the description of the adopted methodology to performed statistical analysis and a summarized explanation of the different implemented statistical methods. Last in this part, the adopted machine learning methodology and a brief explanation of the different methods and concepts incorporated in this methodology, are presented.

The last Part of this dissertation (Part III) incorporates the statistical results and machine learning results Chapters. Thus, the results obtained from the application of the statistical and machine learning methods, respectively, are presented and discussed in these Chapters. In this Part there is also a Chapter dedicated to the conclusion of the dissertation as well as its limitations and future work.

The Appendices can be found in the last pages of this dissertation. These include an appendix that is relative to the statistical analysis and another one that is relative to some results obtained using machine learning techniques. The content of the appendices is intermediate results, that is, results that were necessary to reach the final results and conclusions.

Chapter 2

Fabry disease and hypertrophic cardiomyopathy disease

Since the main goal of this study is to detect important electrocardiogram characteristics to support and assist in an early diagnosis of Fabry disease, it is essential to present a description of this disease, its phenotype and clinical manifestations, how to diagnose it, and possible treatments. Thus, this is the focus of the first section of this chapter. Fabry is not the only disease addressed in this chapter, Hypertrophic Cardiomyopathy is also described in the second section. This disease is included in this study due to the similarities in the cardiac manifestations of these two diseases, mainly left ventricular hypertrophy (LVH), which is normally (40-60% of cases) caused by Sarcomeric Hypertrophic Cardiomyopathy, however, in about 5-10% of cases, unexplained LVH can be caused by other non-genetic or rarer genetic disorders, such as Fabry [7].

2.1 Fabry Disease

This section includes a description of Fabry disease, its phenotype and clinical manifestations (with a description of White Matter Lesions), how it is diagnosed and the possible treatments.

2.1.1 Description

Fabry disease (FD), first recognized as a systemic vascular disease, is an X-chromosome-linked lysosomal storage disorder caused by mutations in the galactosidase alpha (GLA) gene, that result in a deficient or even absent activity of the enzyme alpha-galactosidase A (α -GAL A). This deficient activity leads to lysosomal accumulation of globotriaosylceramide (Gb3) and other related glycosphingolipids, since α -GAL A is responsible for catalyzing the hydrolysis of glycosphingolipids and participates in their degradation in the lysosome. Gb3 can accumulate in cells and organs, for example, body fluids, vascular endothelial cells, perithelial, smooth-muscle cells of blood vessels, cardiomyocytes, cardiac conduction tissue, and valvular fibroblasts. Fabry disease affects a small percentage of the population, and so, it is considered a rare disease [1] [4] [8] [9].

2.1.2 Phenotype and Clinical Manifestations

The mutations in the GLA can cause a practically null enzymatic activity or can lead to residual enzymatic activity (REA). The first case is related to severe and early onset classical phenotypes that develop in childhood or adolescence, like acroparesthesias (burning, tingling, or prickling sensations or numbness in the extremities), neuropathic pain, hypohydrosis (decreased sweating), heat, cold and exercise intolerance, cornea verticillata (whorl-like pattern of golden brown or gray deposits in the inferior interpalpebral portion of the cornea), angiokeratomas (vascular lesions in the skin and mucous membranes), gastrointestinal symptoms and proteinuria (high proteins concentration in the urine); in adulthood, patients also experience sensorineural deafness and cardiac, renal and cerebrovascular manifestations. The second case is associated with attenuated and late-onset phenotypes such as those indicated for adulthood and the phenotype may be influenced by the participation of an organ, such as the kidneys or the heart [4] [10].

Some of the cardiac manifestations in FD are left ventricular hypertrophy (LVH) (unexplained by abnormal cardiac loading conditions), which is the most common, palpitations and arrhythmias, exertional dyspnoea (person feels short of breath during exercise), small-vessel coronary disease (condition in which

the walls of the small arteries in the heart are not working properly), and even heart failure. Therefore, this disease leads to premature death, being cardiac complications one of the most common causes [1] [4] [9].

It has been proved that symptoms occur earlier in males than in females, including cardiac symptoms like left ventricular hypertrophy and that the disease manifests in a moderate way in women [1].

A study including Fabry patients with p.F113L mutation reported white matter lesions (WMLs) as a common manifestation that increases with age and becomes a universal finding above 70 years old (in both sexes). However, due to the higher prevalence of WMLs in young patients without other comorbidities, WMLs seem to be caused by FD, probably mainly by Gb3 deposits in microglial cells and astrocytes, and not by age or other conditions. A similar conclusion was found in a large cohort of p.N215S patients, differing only in the appearing age, which is early before 30 years old in the p.F113L mutation and before 40 years in this cohort [4].

White Matter Lesions

The exchange of information and communication between different areas of the gray matter is carried out by a network of nerve fibers denominated white matter (WM). The White matter is located below gray matter in the brain covering almost half of it and is superficial to gray matter in the spinal cord. The neural networks are formed by nerve fibers called axons that are the extensions of nerve cells (neurons). The axons are surrounded by a type of sheath, the myelin, that allows electrical impulses to transmit quickly and efficiently along the nerve cells. Since myelin is composed of protein and fatty substances it is responsible for the white color in the white matter [11] [12] [13].

White matter lesions (WMLs) are a consequence of cerebral small vessel disease (CSVD) on the brain parenchyma, along with lacunar infarcts, cerebral microbleeds and enlarged perivascular spaces. The term CSVD incorporates all the pathological processes of the cerebral small vessels and it is the most common cause of vascular cognitive impairment and dementia. White matter lesions are also called leukoaraiosis or white matter hyperintensities being the last one more of a descriptive expression used on magnetic resonance imaging (MRI) since these lesions are best seen as hyperintensities on T2 weighted and FLAIR (Fluid-attenuated inversion recovery) sequences of MRI. Non-vascular conditions may also lead to WMLs, for example, any change in chemical composition, damage, or ischemia of myelinated fibers. Dementia,

cognitive decline, abnormal gait, and urinary incontinence are all clinical manifestations of WMLs [11] [12] [13].

2.1.3 Diagnosis

Many of the non-cardiac symptoms mentioned before may not be evident in non-classic forms of FD and in some patients the cardiac symptoms may be similar to those observed in sarcomeric hypertrophic cardiomyopathy HCM, for these reasons it is difficult to diagnose Fabry disease [1] [7] [9].

To diagnose FD, first of all, an examination of the patient's medical and family history is essential. Also, the combination of unexplained LVH and non-cardiac signs and symptoms should raise suspicion of FD. If there is, in fact, suspicion of FD, then, in males, determination of α -GAL activity in plasma or peripheral leukocytes is an efficient way to diagnose the disease. On the other hand, due to random x-chromosome inactivation, females can have a normal to very low α -GAL activity, so finding a mutated GLA gene is crucial to diagnose FD in females [1] [7] [9] [10] [14].

2.1.4 Treatment

The therapies that are currently used in FD are Enzyme replacement therapy (ERT) with agalsidase-alpha or agalsidase beta and chaperone treatment [2] [10]. ERT is unable to eliminate neurophatic pain, only attenuate it and reduce left ventricular wall mass. However, conduction and valve abnormalities, as well as coronary artery disease keep progressing. Besides the improvement in function of cerebral vasculature, there is no clear reduction in the risk of stroke. This therapy provides a significant decrease in plasma Gb3 levels in males with classical FD, matching with the levels in females and non-classical patients. In fact, Enzyme replacement therapy presents higher efficiency in kidney function comparatively with other organ systems. However, ERT includes numerous disadvantages, the major one is the development of anti-drug antibodies (ADAs) that reduce the treatment efficiency. It is also a treatment for life, thus, requires infusion reactions and venous ports for long term access [2] [9].

The chaperone treatment uses chemical chaperones, which are small molecules that bind to mutant enzyme proteins, in the endoplasmic reticulum and assist in their correct folding, maturation, and trafficking to their functional site, thus increasing the residual activity. A pharmacological chaperone used in FD treatment is Migalastat (approved and sold as Galafold), it binds selectively and reversibly to active sites of amenable mutant forms of α -GAL. This bind stabilizes these forms of α -GAL in the endoplasmic reticulum and assist in their transport to lysosomes and then, when the environment is appropriate, Migalastat dissociates from α -Gal A leading to an increase in enzyme activity and, consequently, to catabolism of glycosphingolipids like globotriaosylceramide (Gb3). This pharmacological chaperone has been evaluated in ERT-naïve (patients who have never received ERT) and ERT-treated FD patients and the results were an increase of α -GAL activity, reduction of left ventricular mass index and kidney function preservation. Unfortunately, this therapy is only efficient in certain mutations of FD, actually just 30-50% of mutations are amenable to Migalastat use. The hypothesis that in-vitro mutation amenability correlates with in-vivo amenability may not correspond to the truth and this also complicates the use of Migalastat [2] [9] [15].

2.2 Hypertrophic Cardiomyopathy

This section includes a description of Hypertrophic Cardiomyopathy, its phenotype and clinical manifestations, how it is diagnosed and the possible treatments.

2.2.1 Description

Hypertrophic cardiomyopathy (HCM) is a heterogeneous myocardial disease, most often caused by autosomal dominant sarcomeric gene mutations, in this case called Sarcomeric hypertrophic cardiomyopathy, and it is one of the most common causes of sudden cardiac death, specially in young people [16].

Sarcomeric HCM is associated with mutations in genes encoding sarcomere or sarcomere-related

proteins that are critical for contractile function. This disease has been identified in at least 12 genes, the most common being the MYBPC3 and MYH7 genes. In most cases, HCM is inherited as an autosomal dominant trait, with offspring having a 50% chance of inheriting the same gene variant [17].

Sometimes the terms hypertrophic cardiomyopathy and sarcomeric cardiomyopathy are used as synonyms. However, this is not correct since that, in hypertrophic cardiomyopathy, approximately 10% of mutations are located in genes of non-sarcomeric proteins and in 25 to 30% the cause of the disease is unknown [18].

2.2.2 Phenotype and Clinical Manifestations

The clinical phenotypes of HCM and the severity of these phenotypes may differ among the different mutations of HCM. However, the clinical phenotypes of the most common HCM genes (MYBPC3 and MYH7) do not seem to differ significantly [18].

The phenotypic expression of HCM is manifest by left ventricular hypertrophy (LVH), which is the main manifestation, there are also three key pathological hallmarks and two important pathophysiologic features. The pathological hallmarks are myocyte hypertrophy, myofibrillar disarray, and interstitial fibrosis. As for the pathophysiologic features, these are impaired ventricular filling and dynamic left ventricular outflow tract (LVOT) obstruction [18] [19].

The most frequent symptoms of HCM arise from four major pathophysiologic conditions which are diastolic ventricular dysfunction, obstruction to left ventricular outflow tract (LVOT), imbalance between myocardial oxygen supply and demand, and cardiac arrhythmias. The first condition, diastolic ventricular dysfunction leads to exertional dyspnea, exercise intolerance, orthopnea, peripheral edema and heart failure with preserved ejection fraction (HFpEF), due to the increase of the left ventricular diastolic pressure on exertion. As for LVOT obstruction, when patients have severe obstruction usually they have elevated diastolic pressure and exertional dyspnea and some of them develop heart failure. The pathophysiologic condition of imbalance between myocardial oxygen supply and demand causes ischemic chest pain. There are several clinical manifestations of arrhythmias, the cardinal ones being palpitations, pre-syncope, and syncope [18] [19] [20].

In HCM, the expression of cardiac hypertrophy is uncommon in childhood, it normally develops during adolescence and rarely after the fifth decade. Thus, the expression of cardiac hypertrophy in HCM is considered to be age-dependent [18].

It is also important to notice that myocardial hypertrophy may also occur in patients with phenocopy conditions which may mimic HCM. One example is Fabry disease [18].

2.2.3 Diagnosis

The diagnosis of HCM is based on the presence of a left ventricular wall thickness of 15 mm or more, that is otherwise unexplained by loading conditions or other secondary causes, such as hypertension, aortic stenosis, or other phenocopied conditions. In relatives of individuals with HCM or people with positive genotype, a left ventricular wall thickness of 13 is enough to diagnose HCM [18] [20].

The presence, location, and mechanism of left ventricular obstruction should be identified with serial continuous-wave Doppler. Then, the peak instantaneous continuous-wave Doppler gradient value is analysed. A value equal to 30 mm or higher at rest indicates obstructive physiology and equal to 50 mm or higher, either at rest or with provocation, is considered the threshold for septal reduction therapy in symptomatic patients [18] [20].

Electrocardiography, genetic testing and CMRI are also important in the diagnosis of HCM [18] [20].

Some electrocardiographic findings are not specific to HCM, however, they should raise suspicion of HCM. Examples of these findings are signs of ventricular hypertrophy with increased precordial voltages, and non-specific ST segment and T-wave abnormalities, deep inferior and/or lateral Q-waves (suggestive of a hypertrophied septal depolarization), P-wave abnormalities (suggestive of left atrial enlargement), and T-wave inversions [18] [20] [21].

Although genetic testing is not required for the diagnosis of HCM, it should be done if another genetic condition is suspected and also to help identify first-degree family members at risk of developing HCM. A positive genetic test will support the diagnosis in a proband, however, a negative test will not exclude it. When a genetic test is not performed or if a causal variant is not identified in the proband, then, is recommended that an ongoing periodic clinical surveillance, with electrocardiography and echocardiography, is

applied to the relatives [18] [20].

As for cardiac magnetic resonance imaging (CMRI), it provides better morphologic and tissue characterization and volumetric assessment compared with echocardiography. CMRI is recommended for patients on which echocardiography was inconclusive for diagnose HCM or patients with known HCM for whom additional information would be helpful for treatment decisions [18] [20].

Electrocardiogram characteristics, patterns of inheritance, multiorgan involvement, and, obviously, genetic testing are useful tools to differentiate HCM from other genetically mediated phenocopies of HCM, like Fabry disease [18] [20].

2.2.4 Treatment

The treatment that is applied to symptomatic patients is pharmacological treatment. This treatment includes beta-blockade which relief ischemic chest discomfort and may attenuate LVOT obstruction induced by exercise and resulting dyspnea. Another agent may be added to a beta blocker, called disopyramide, which may reduce, even more, the symptoms in patients with LVOT obstruction. For patients that do not tolerate or respond to beta blockers, calcium channel blocker therapy may be beneficial [18] [22].

Septal reduction therapy (SRT) is indicated for patients that remain symptomatic even with pharmacologic treatment. There are two forms of SRT, surgical septal myectomy or alcoholic septal ablation. To note that, despite the good intermediate and long-term results of SRT, some patients continue to be symptomatic or develop heart failure after this therapy [18] [22].

There is also a possibility for patients with symptoms refractory to pharmacological treatment, which is catheter ablation, although it has not been studied in HCM cohorts [18] [22].

An important goal and desire is that future therapies can effectively address the pathophysiology of HCM since the current therapies focus only on the disease manifestations. There are trials investigating cardiac myosin inhibitors, specifically their effect on LVOT obstruction and diastolic heart failure that is associated with HCM[18] [22].

Chapter 3

Electrocardiogram and machine learning diagnostic value

This chapter is divided into 2 sections. The first section includes an explanation of the electrocardiogram (ECG), as well as the main electrocardiographic findings in patients with Fabry disease and patients with sarcomeric HCM. Several studies where ECG characteristics are used to diagnose and distinguish diseases, proving their diagnostic value as data for classification problems, are presented in this section. The second section is relative to machine learning diagnostic value, more specifically, how can machine learning improve the medical diagnosis and different studies that successfully apply classification machine learning algorithms to ECG characteristics.

3.1 Electrocardiogram characteristics as data for classification

To understand how electrocardiogram characteristics can be used as data for classification and their diagnostic value, it is important to first understand what is an electrocardiogram.

Electrocardiogram (ECG) is the record (*gram*) of the potential changes at skin surface that result from

the depolarization and repolarization of heart muscle (*cardio*) allowing the identification and location of pathology. This electrical activity (*electro*) is small and is detected by contact electrodes placed on different parts of patient's chest and limb [23] [24]. The ECG may be composed of 12 leads which provides views of the heart in both frontal and horizontal planes, thus obtaining records in different orientations and perspectives. There are two types of leads: the limb leads that view the heart in frontal plane and the precordial leads or chest leads that view the heart in horizontal plane. The first one is composed by the leads I, II and III, called the standard leads and the ones that form the Eithoven's triangle, and also by three augmented limb leads, aVR, aVL and aVF. The second one, the precordial leads are V1, V2, V3, V4, V5 and V6 [25] [26].

The vertical axis of an ECG record is voltage and the horizontal is time, therefore, measurements in horizontal axis indicate the overall heart rate, regularity, and the time intervals during electrical activation. In vertical axis, measurements indicate the voltage that is measured in the surface and that represent the "summation" of the electrical activation of all the cardiac cells [27].

Interpretation of ECG may detect many cardiac abnormalities, some only by a single ECG record and others by serial recording over time [27]. Records of heart's activity during 24 to 48 hours, or even more, are archived by a portable device called Holter monitor [28].

Electrocardiogram characteristics, such as intervals, amplitudes, and rates, can be obtained by interpreting the ECG.

There are already several electrocardiographic findings observed in Fabry disease patients. A short PR interval, signs of LVH, T wave inversion, bradycardia, and atrial fibrillation are examples of these electrocardiographic findings [8] [10]. Shortening of QRS width and increased QTc duration are also identified in some studies [29] [30].

As previously described, patients with HCM also exhibit electrocardiographic abnormalities, for example, signs of ventricular hypertrophy with increased precordial voltages, and non-specific ST segment and T-wave abnormalities, deep inferior and/or lateral Q-waves (suggestive of a hypertrophied septal depolarization), P-wave abnormalities (suggestive of left atrial enlargement), and T-wave inversions [18] [20] [21].

This proves that Fabry disease and HCM patients do, in fact, show electrocardiographic abnormalities. Also, there is a huge need to identify other causes of LVH besides sarcomeric HCM and electrocardiography

is one of the first steps when evaluating these patients. Therefore, electrocardiogram could be an important tool to distinguish and diagnose these diseases.

Some studies already tried to identify ECG characteristics to distinguish between FD and HCM, which is the case of Junqua et al. [5] and Namdar et al [31].

Junqua et al. [5] aimed to assess the diagnostic value of electrocardiographic scores for LVH in Fabry disease patients and also the diagnostic value for Fabry disease of a model combining electrocardiographic and echocardiographic criteria. Junqua et al. evaluated several ECG characteristics of 61 Fabry disease patients and 59 patients with sarcomeric HCM. These characteristics were heart rate (bpm), corrected PQ (ms), QRS duration (ms), corrected QT (ms), RBBB, LBBB, Left anterior hemiblock, Left posterior hemiblock, pre-excitation, pathologic Q wave, atrial fibrillation, and LVH indexes (Cornell voltage index (mm), Gubner index (mm), Sokolow—Lyon voltage index (mm), Romhilt—Estes score, Sokolow—Lyon product (mm.ms), and Cornell product (mm.ms)). A univariate analysis was performed, using Student's t-test for continuous variables and Fisher test or chi-squared test for qualitative variables. Among the characteristics, QRS duration was significantly higher and RBBB more frequent in the Fabry group. The LVH indexes, Gubner index and Sokolow—Lyon product, were also higher in this group (Fabry patients). As for the remaining ECG characteristics, Junqua et al. did not find significant differences between FD and HCM patients. A multivariable analysis, with the electrocardiogram and echocardiogram variables that presented a p-value ≤ 0.01 in the univariate analysis, was performed. Among the ECG variables, only RBBB and Sokolow—Lyon product were independently associated with Fabry disease. It is important to mention that Junqua et al. found that age was significantly different between FD and HCM groups, being Fabry patients older than HCM patients [5].

The aim of the study of Namdar et al. [31] was to investigate the value of common ECG parameters in the diagnosis of patients with Fabry disease (FD), amyloidosis, hypertensive heart disease (HHD), aortic stenosis (AS), and nonobstructive hypertrophic cardiomyopathy (HC). These ECG parameters include resting heart rate (beats/min), P-wave duration (ms), P wave corrected for heart rate (ms), PQ interval (ms), corrected PQ interval (ms), PQ interval < 120 ms, PQ interval minus P-wave duration in lead II (ms), QRS duration (ms), Sokolow—Lyon index (mV, left ventricular hypertrophy and right ventricular hypertrophy), maximum peak to end of T wave (ms), corrected QT mean interval (ms), and corrected QT dispersion (ms). It should be noted that, in this study, groups were matched for age and left ventricular muscle

mass index so that it would not influence ECG parameters. Namdar et al. compared the continuous variables with analysis of variance for repeated measurements and post hoc analysis with Fisher's probable least-significant differences test and/or Student's t-test. As for categorical variables, these were compared using Fisher's exact test. The study included 94 patients, being 17 diagnosed with FD, 20 with HHD, 17 with amyloidosis, 20 with AS, and 20 with HC. In the FD group, PQ interval, corrected PQ interval, and PQ interval minus P-wave duration in lead II variables were significantly shorter. These two last characteristics led to the highest diagnostic performance for the diagnosis of FD, with cut-off values of 114 ms and 40 ms, specificity of 90% and 99%, AUC of 0.90 and 0.94, respectively, and sensitivity of 82% on both. However, Namdar et al. realised that the combination of normal correct QT interval (i.e., < 440 ms) and short PQ interval minus P-wave duration in lead II (i.e., < 40 ms) gave a better diagnostic performance for the diagnosis of FD, with specificity of 99% and sensitivity of 100%. Should be noted that none of the HC patients fulfilled these two criteria [31].

As previously mentioned, WMLs are a manifestation in FD patients, and similar to the studies that aimed to distinguish FD from HCM patients, there is also studies that aimed to distinguish FD patients with WMLs from FD patients without WMLs, based on electrocardiogram characteristics, which is the case of Braga (2021) [32].

Braga (2021) [32] evaluated ECG characteristics of 114 FD patients, in which 61 have WMLs and 53 do not have WMLs. These ECG characteristics include heart rate variables (HR Min, HR Mean, HR Max), HRV variables (ASDNN 5, SDANN 5, SDNN, RMSSD), QT analysis variables (QT Min, QT Mean, QT Max, QTc Min, QTc Mean, QTc Max, $QTc \geq 450$). and a supraventricular ectopy variable (Longest R-R). Braga (2021) identified that 6 ECG characteristics, HR Max, QT Min, QT Mean, QT Max, QTc Mean, and $QTc \geq 450$, were significantly different between the two groups. Age was also significantly different between the groups, with FD with WMLs being older than FD without WMLs patients. These comparison of continuous variable was done with t-test and Mann Whitney U Test, for normally and non-normally distributed variables, respectively. Then, Braga (2021) performed a univariate logistic regression analysis to evaluate the association between each ECG variable and the presence of WMLs and realised that HR Max, QT Min, QT Mean, QTc Mean, and $QTc \geq 450$ variables presented a significant association with the presence of WMLs. However, when adjusted by age, none of the ECG variables presented a significant association with the presence of WMLs. Due to the significant difference in the age of the two groups, Braga (2021) divided the patients into age

classes, and repeated the univariate and multivariable analysis in the age class of [40,60[(the only class with similar number of patients in each group) that did not present significant differences on patients' age or sex. Two variables, of HRV, were significantly associated with the presence of WMLs, SDANN 5 and SDNN variables. When adjusted to age or sex, these two variables remained associated with the presence of WMLs [32].

Another relevant study, Galluzzi et al. [33], also uses ECG data, more specifically, HRV data of patients with and without WMLs. However, in this study, patients have mild cognitive impairment (MCI) instead of FD. Thus, the aim of this study was to evaluate the independent association of HRV with WMLs in patients with mild cognitive impairment (MCI). Galluzzi et al. used 24-hour ECG recordings of 82 MCI patients, in which 32 present WMLs, to calculate HRV variables. In time domain, standard deviation of the RR intervals, standard deviation of the 5-minute mean values of RRs for each 5-minute interval, average of standard deviations of RR for each 5-minute interval, and the RMSSD were calculated. As for frequency domain, low frequency (LF) 0.04 – 0.15 Hz, HF 0.15 – 0.40 Hz, and low high-frequency ratio were calculated. The differences between MCI with and without WMLs groups were assessed using t-test and Mann-Whitney for normally distributed and non-normally distributed continuous variables, respectively, and chi-square for categorical variables. Galluzzi et al. also evaluated the association between each HRV index and WMLs (ARWMC, age-related white matter changes, scale total score, from magnetic resonance imaging) through linear regression models, and the potential of HRV indices to predict the extent of WMLs with a stepwise multiple regression model. In this study, age was significantly different between the two groups, patients with WMLs being older than those without WMLs. Differences were also found in RMSSD and LF that were reduced in MCI patients with WMLs. Galluzzi et al. found that RMSSD, LF, and HF were inversely associated with the extent of WMLs in the unadjusted models, however, in the adjusted models, RMSSD was the only variable that remained inversely associated. As for the stepwise results, these showed that RMSSD and age were significant predictors of WMLs [33].

In conclusion, all the presented studies demonstrate that ECG characteristics are already used as data to diagnose and to distinguish diseases and/or manifestations, which is the case of Fabry disease and White Matter Lesions, and initially promising results have been achieved.

3.2 ECG-based machine learning algorithms for classification

Machine learning (ML) is a branch of computational algorithms that are intended to reproduce human intelligence by being able to detect meaningful patterns in data even if the datasets are complex and large [34].

A correct diagnosis of a disease or condition is a complex process since a lot of symptoms are non-specific and variable among patients, some diagnostic tests are not regularly done and are expensive, and it requires a huge human effort, as well as time. Also, physicians are usually susceptible to cognitive bias during the diagnosis stage, being more biased to diseases or conditions that they have already diagnosed in the past and, therefore, are more familiar with. Due to the prevalence of rare diseases, there is a higher probability that physicians have cognitive bias during the diagnosis stage of these diseases [35] [36] [37].

Machine learning is useful for medical diagnosis since with unbiased and balanced datasets, ML algorithms are able to attenuate the cognitive bias problem and, thus, produce higher accuracy. Machine learning has been applied to a variety of medical diagnosis problems, for example, breast cancer, diabetes, cancer tissues, and thyroid [35] [37].

Similar to what is proposed in this study, several studies have also applied classification supervised ML algorithms to ECG data, which is the case of Hijari et al. [6], Vigier et al. [38], Munla et al. [39], and Braga (2021) [32]. What is proposed by each of these studies, as well as the main methods and results, are now going to be briefly explained.

Hijari et al.[6] attempted to identify patients at risk of Long QT Syndrome (LQTS) symptoms, by training ML algorithms with input variables extracted from raw ECG data. The purpose was for a classifier to be able to output “symptoms expected” or “no symptoms expected” based on some measurements from an ECG. Hijari et al. considered that QT and RR intervals were the most relevant markers for their study and used 24-hour ECG records of 434 patients with the most common LQTS genotypes. Then, they implemented several classification algorithms, such as k-nearest neighbors, linear SVM, RBF SVM, random forest, and AdaBoost, using scikit-learn (Python library) [40]. Finally, they determined the classifiers’ accuracy by leaving 30% of the samples for testing and 70% for training, and doing the process of selection of training data, training and testing 50 times for each classifier, then calculating the average result. In conclusion,

the classifier that performed better, in the test phase, in the Hijari et al. study, was the RBF SVM classifier with about 70% accuracy, and the random forest classifier achieved the second-best accuracy [6].

Vigier et al. [38] aimed to create and evaluate a machine learning model able to discriminate between cancer patients and healthy controls based on 5-min ECG recordings. They selected 12 heart rate variability (HRV) features, which were then reduced to the top 5 (SDNN, RMSSD, pNN50%, HRV triangular index, and SD1) with recursive feature elimination (RFE). These features were then used as input to three different machine learning algorithms: linear discrimination analysis (LDA), random forest, and naive bayes. An ensemble model based on the stacking method, which comprised the predictions of the three base classifiers, was also implemented. The study included 77 cancer patients (breast, prostate, lung, colorectal, and pancreatic cancers) and 57 healthy controls, 40% of the dataset was used for testing and the remaining 60% for training, with a tenfold cross validation. Vigier et al. concluded that among the three base ML algorithms, random forest was the one that performed best during the training evaluation, with an accuracy of 85%. The ensemble model performed better than all base classifiers, with an accuracy of 0.929 in the training phase and 0.865 in the test phase [38].

A study, performed by Munla et al. [39], investigated stress level detection of a driver during a real word driving experiment, based on heart rate variability (HRV) analysis. Munla et al. performed different HRV analyses, in time, frequency, non-linear, and time-frequency. The features extracted from these analyses were used, separately, as input data to three ML algorithms, k-nearest neighbor (KNN), radial basis function (RBF) and support vector machine with linear and RBF kernels, in order to classify data into two physiological states of the driver, highly stressed or normal. The SVM with RBF kernel was the model that presented the best results, with a correct prediction rate of 83.33% both with time and non-linear parameters, and 66.66% with frequency, poicare (a non-linear analysis) and STFT (a time-frequency analysis method) parameters [39].

One of the objectives of a study performed by Braga (2021) was to evaluate the effectiveness of machine learning methods to discriminate Fabry disease (FD) patients with white matter lesions (WMLs) from Fabry disease patients without white matter lesions based on ECG data. For this, Braga (2021) used 48 Fabry disease patients, where 23 have WMLs and 25 do not have WMLs. After performing two feature selection methods, one being a filter method that selected the 10 most significant features not correlated (using Mann Whitney U Test and a $|\rho| < 0.80$) and the other recursive feature elimination (RFE), Braga

(2021) ended up with 4 electrocardiogram features, which were heart rate mean and maximum, QT maximum, and SDANN 5. Using these 4 features and SDANN 5 feature alone, five classification algorithms, logistic regression (LR), SVM linear Kernel, SVM RBF kernel, random forest (RF), and k-nearest neighbors (KNN), were applied, in order to identify the optimal combination of features and the classification model with the best performance. In the case of the 4-feature combination, the model that gave the highest validation accuracy, $79.72\% \pm 1.40\%$, was RF, followed by KNN with a validation accuracy equal to $77.75\% \pm 1.57\%$. However, using only the SDANN 5 feature, the model that gave the highest validation accuracy was SVM linear kernel with $74.53\% \pm 0.67\%$, followed by LR with $74.52\% \pm 0.72\%$. To note that to evaluate the models' performance a 5-fold cross validation was used [32].

All these studies are examples of the medical diagnosis power of machine learning, in this case, more specifically, when it is applied to ECG data, since they show several machine learning algorithms that lead to high classification performances (most above 70%).

Part II :

Materials and Methods

Chapter 4

Study data

This chapter introduces the sample of individuals included in this study and some of their demographic characteristics. The evaluated variables and their normal reference range of values according to age and/or sex are also presented in this chapter.

4.1 Study Sample

The study sample included Fabry disease (FD) patients and Sarcomeric Hypertrophic cardiomyopathy (HCM) patients from the Cardiology and Neurology Departments of Hospital Senhora da Oliveira in Guimarães city, Portugal.

This study retrospectively enrolled 114 patients with FD (GLA gene mutation c.337T>C (p.F113L)), in which 61 (53.51%) have WMLs and 53 (46.49%) do not, and also 39 patients with Sarcomeric HCM.

The patients' ages range between 19-91 years inclusive (Table 1). While the percentages of females in the FD with WMLs group and the FD without WMLs are 62,3% and 60,4% respectively, in the HCM group it is only 33,3%.

Table 1: Demographic characteristics of the study sample

	HCM (n=39)	FD with WMLs (n=61)	FD without WMLs (n=53)
Sex (Female number)	13	38	32
Age ([Minimum, Maximum] years)	[21, 87]	[20, 91]	[19, 71]

Since Holter variable measures vary across age and sex [41] and age is the strongest known risk factor for the onset of WMLs [42] [43], the patients were subgrouped into three age classes: age 19-39 years, age 40-59 years, and age > 59 years (Table 2). The number of FD patients with WMLs and HCM patients in age class 19–39 years is very low, with 8 and 4 patients, respectively. Thus, this age class was not included in this study. Furthermore, as there are only 3 FD patients without WMLs older than 59 years, this subgroup of patients was also excluded.

Table 2: Patients count per group and age class

Age Class	HCM Patients (Females)	FD with WMLs Patients (Females)	FD without WMLs Patients (Females)
[19,39]	4 (1)	8 (6)	25 (14)
[40,59]	16 (6)	23 (13)	25 (17)
[60,91]	19 (6)	30 (19)	3 (1)

4.2 24-Hour Holter

As mentioned in Section 3.1, an Electrocardiogram (ECG) is an exam that records the potential changes at skin surface that result from the depolarization and repolarization of the heart muscle, thus, allowing the identification and location of pathologies. Therefore, a single ECG or a recording over time (24 to 48 hours) achieved by a Holter monitor, are powerful exams since they can detect many cardiac abnormalities [23] [27].

Several variables were evaluated in this study, such as patients' sex, which is the only dichotomous variable, age and, since all patients were submitted to a 24h-ECG exam (Holter), there are also 15 electrocardiogram characteristics, which can be observed in Table 3, that analyse heart rate, heart rate variability, QT interval, and supraventricular ectopy.

Table 3: 24-Hour Holter variables

Heart rate	
HR Min	Minimum heart rate
HR Mean	Mean heart rate
HR Max	Maximum heart rate
Heart rate variability	
ASDNN 5	Average standard deviation of the averages of all normal-to-normal R-R intervals in all 5-min segments of the entire recording
SDANN 5	Standard deviation of all average normal sinus R-R intervals for all 5-minute segments in the 24-hour recordings
SDNN	Standard deviation of all normal sinus R-R intervals in the entire 24-hour recording
RMSSD	Root mean square of the mean of the squared differences of two consecutive RR intervals
QT analysis	
QT Min	Minimum QT interval
QT Mean	Mean QT interval
QT Max	Maximum QT interval
QTc Min	Minimum corrected QT interval
QTc Mean	Mean of corrected QT interval
QTc Max	Maximum corrected QT interval
QTc > 450	Prolonged corrected QT interval
Supraventricular ectopy	
Longest R-R	Maximum difference between two R-R peaks

The normal reference range values, that is, the range of values of a healthy patient, of heart rate and QT analysis ECG features, per age class, can be consulted in Table 4. Similarly, in Table 5, are displayed the normal reference range values of heart rate and heart rate variability ECG features, also, per age class. To note that the lower and upper age classes are different among these two tables, being, respectively, 16-19 and 80-89 in Table 4, and 10-19 and 80-99 in Table 5. Another important difference between the results presented in these tables is the way that data is presented. In Table 4 data are presented as median (2nd percentile, 98th percentile) and, in Table 5, are presented as mean $\pm$ standard deviation.

Table 4: Heart rate and QT analysis normal reference range of values.

ECG Features	Sex	Age range (years)							
		16-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
HR (bpm)	Female	72 (47, 105)	67 (48, 98)	66 (47, 95)	67 (47, 90)	69 (52, 94)	71 (53, 94)	72 (55, 98)	72 (50, 102)
	Male	73 (49, 107)	65 (45, 94)	65 (46, 95)	66 (47, 95)	67 (48, 94)	67 (48, 95)	67 (50, 99)	74 (40, 97)
QT (ms)	Female	390 (337, 455)	394 (340, 456)	400 (346, 460)	396 (350, 458)	398 (349, 458)	396 (351, 454)	394 (342, 454)	394 (332, 461)
	Male	378 (332, 452)	394 (342, 454)	396 (344, 454)	394 (342, 458)	396 (342, 458)	398 (346, 454)	398 (336, 458)	395 (334, 476)
QTc (ms)*	Female	429 (382, 473)	418 (374, 458)	419 (377, 464)	421 (379, 468)	427 (391, 472)	429 (391, 473)	432 (396, 476)	432 (393, 480)
	Male	416 (379, 460)	409 (364, 453)	413 (365, 458)	416 (372, 462)	418 (375, 463)	419 (379, 463)	421 (379, 478)	430 (388, 500)

Data are presented as median (2nd percentile, 98th percentile).

* Values according to Bazett's formula.

Table 5: Heart rate and heart rate variability normal reference range of values.

ECG Features	Age range (years)							
	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-99
HR (bpm)	80 $\pm$ 10	79 $\pm$ 10	78 $\pm$ 7	78 $\pm$ 7	76 $\pm$ 9	77 $\pm$ 9	72 $\pm$ 9	73 $\pm$ 10
ASDNN 5 (ms)	81 $\pm$ 20	72 $\pm$ 22	64 $\pm$ 15	60 $\pm$ 13	52 $\pm$ 15	42 $\pm$ 13	43 $\pm$ 11	37 $\pm$ 12
SDANN 5 (ms)*	159 $\pm$ 35	137 $\pm$ 43	130 $\pm$ 33	116 $\pm$ 31	106 $\pm$ 27	111 $\pm$ 31	114 $\pm$ 20	95 $\pm$ 24
SDNN (ms)*	176 $\pm$ 38	153 $\pm$ 44	143 $\pm$ 32	132 $\pm$ 30	121 $\pm$ 27	121 $\pm$ 32	124 $\pm$ 22	106 $\pm$ 23
RMSSD (ms)*	53 $\pm$ 17	43 $\pm$ 19	35 $\pm$ 11	31 $\pm$ 11	25 $\pm$ 9	22 $\pm$ 6	24 $\pm$ 7	21 $\pm$ 6

Data are presented as mean $\pm$ standard deviation.

Chapter 5

The implemented statistical methods

Statistical analysis is an important tool to describe data through summary charts and tables. It is also extremely useful to test hypotheses and draw conclusions about the data. In this study, a descriptive statistical analysis was first performed on all variables. Then, inferential statistics were used to compare the differences between groups. Logistic regression analysis was also performed to examine the association of independent variables (Holter variables and demographics variables) with one dichotomous dependent variable (presence or absence of a disease). Statistical analysis was performed using Python 3.7 utilizing a set of open source libraries including pandas [44], scipy [45], statsmodels [46], scikit-learn [40], and pingouin [47]. All the statistical methods used are going to be briefly explained in the following sections as well as the Python functions used in each case.

5.1 Descriptive statistics

To better understand the data, multiple analyses were implemented. First, a description of all variables of each group was performed, in which multiple values were accessed such as count, mean, standard deviation (std), minimum (min), 25th, 50th and 75th percentiles and also maximum (max). For this, the *describe()* function from the pandas library [44] was used.

Then, to analyse sex distribution, a count plot for each group was obtained using *countplot()* function

from seaborn, a Python data visualization library based on matplotlib.

Finally, another important analysis was performed, which was obtaining the histogram of all variables (excluding Sex variable) of each group, to see how each variable is distributed and to help decide, along with a normality test, if the variable has a normal distribution. To obtain the histogram, the *hist()* function from the matplotlib library was used.

5.2 Statistical inference: Hypothesis testing

On carrying out a hypothesis test, we start to formulate the hypothesis to be evaluated and possibly reject. Commonly, the null hypothesis (H_0) is defined by the null existence of differences between the results of compared groups, and H_0 , in turn, is the opposition to the alternative hypothesis (H_1) [48] [49].

When accepting or rejecting a hypothesis test, there is a certainty (not absolute) based on the significance level (α). This value is referred to as type I error, i.e., the probability of error on rejecting H_0 when H_0 is actually true in the population. In this work, we assume as level of significance $\alpha = 0.05$. In deciding whether to accept or reject H_0 , we obtain a p-value as a result of the contrast made. Since p-value (p) is a probability, it ranges between 0 and 1. This probability is defined as the minimum level of significance with which H_0 is rejected. If $p < 0.05$, we will have enough sample evidence to reject the contrasted null hypothesis in favor of the alternative hypothesis [48] [49].

5.2.1 Normality

The Shapiro-Wilk test is a normality test that analyzes a data sample and, by comparing its scores to a normally distributed set of scores with the same mean and standard deviation, quantifies how likely it is that the data was drawn from a Gaussian distribution [50] [51].

- Null hypothesis (H_0): Data was drawn from a normal distribution.
- Alternative hypothesis (H_1): Data was not drawn from a normal distribution.

This means that, if the test is non-significant, that is, if $p > 0.05$, then the distribution of the sample is not significantly different from a normal distribution. On the other hand, if the test is significant, with $p < 0.05$, then the distribution is significantly different from a normal distribution, in other words, it is not normally distributed.

This test was implemented in each variable of each group using the *shapiro()* function from the *scipy* library [45], *stats* module. Both the test statistic and the p-value for the hypothesis test were obtained to determine if variables were or were not drawn from a normal distribution.

5.2.2 Homogeneity of variances

Levene's test is used to test the homogeneity of variance in two or more groups by essentially applying a one-way ANOVA on the deviations, that is, the absolute value of the difference between each score and the mean of its group [49] [50].

- Null hypothesis (H0): The variances in different groups are equal.
- Alternative hypothesis (H1): The variances in different groups are significantly different.

If the test is non-significant ($p > 0.05$) then the variances in the different groups are equal and there is homogeneity of variances. However, if the test is significant ($p < 0.05$) then the variances in the different groups are significantly different and, therefore, the assumption of homogeneity of variances has been violated.

For Levene's test, the *levvene()* function was used, again from the *scipy.stats* module [45]. The homogeneity of variances was only analyzed in the variables which presented a normal distribution, so, the ones that gave a non-significant result ($p > 0.05$) in the Shapiro-Wilk test (presented previously). Both the test statistic and the p-value were acquired.

5.2.3 Differences between independent samples: quantitative data

One-way ANOVA Test

The one-way Analysis of variance (ANOVA) is a statistical test used to compare the means of two or more groups on one dependent variable to see if the groups' means are significantly different from each other. To achieve this, the test separates the total variability in the data into that which can be attributed to differences between the individuals from the different groups (the between-group variation), and to the random variation between the individuals within each group (the within-group variation) [49] [51].

This test requires that groups are defined by the levels of a single factor, that the variable is normally distributed in each group (Shapiro-Wilk test) and lastly, requires homogeneity of variances (Levene's test). One-way ANOVA can only be performed when there is independence of observations, that is, there is no relationship between the observations in each group or between the groups [49] [51].

- Null hypothesis (H0): All group means in the population are equal.
- Alternative hypothesis (H1): At least one group mean in the population differs from the others.

If the test is significant ($p < 0.05$) then at least one group mean in the population differs from the others. It should be noted that the test only observes if all samples are the same or not, it cannot evaluate which samples differ or by how much. In this case, the between-group variance will be larger than the within-group variance. Although, if the test is non-significant ($p > 0.05$) it means that all group means in the population are equal, so the between-group variance will be similar to the within-group variance.

Only the variables that fulfill the assumptions of normal distribution and homogeneity of variances (evaluated previously by the Shapiro-Wilk and Levene's tests, respectively) were submitted to the one-way ANOVA test. This test was performed by the `f_oneway()` function, once again from the `scipy.stats` module [45]. To compare the means between the three groups, test statistic and p-value were obtained.

Welch-ANOVA Test

As presented previously, the one-way ANOVA is not a suitable test when there is no homogeneity of variances. For this situation, Welch-ANOVA should be used since it does not require homogeneity of

variances. Welch test has the advantage of keeping type I error rate despite heteroscedastic variances [51] [52] [53].

- Null hypothesis (H0): All group means in the population are equal.
- Alternative hypothesis (H1): At least one group mean in the population differs from the others.

To apply the Welch-ANOVA test, a function from the pingouin library [47], the *welch_anova()* function, was used. And, as explained, this test was applied to the variables that have a normal distribution (evaluated previously by Shapiro test) but do not have homogeneity of variances (evaluated by Levene's test)

Kruskal-Wallis Test

The Kruskal-Wallis test is the nonparametric equivalent of the one-way ANOVA and so it is used to determine whether more than two independent samples have a different distribution [51].

- Null hypothesis (H0): Population median of all groups are equal.
- Alternative hypothesis (H1): Population median of all groups are not equal.

The rejection of the null hypothesis (significant test) and therefore the acceptance of the alternative hypothesis, means that at least one of the samples is different from the other samples. Like in one-way ANOVA, Kruskal-Wallis also does not identify where the difference occurs neither how many occur. We do not reject the null hypothesis when the test is non-significant and, so, we conclude that all sample distributions are equal.

Since this test is the nonparametric equivalent of the one-way ANOVA test, it was only performed on the variables which were not normally distributed (evaluated previously by Shapiro-Wilk test). One more time, the test statistic and p-value were obtained, but this time with the *kruskal()* function from the *scipy.stats* module [45].

Unpaired t-test

The unpaired t-test is also known as independent t-test and it is a statistical hypothesis test used to test whether the means of two groups are equal, when there is two independent (unrelated) groups and one numerical or ordinal variable of interest. This test makes some assumptions, such as that the variable of interest is normally distributed in each group and that there is homogeneity of variances (the variances of the two groups are equal) [49] [51].

- Null hypothesis (H0): The population means in the two groups are equal.
- Alternative hypothesis (H1): The population means in the two groups are not equal.

So, when we reject the null hypothesis (significant test, with $p < 0.05$) that indicates that there is sufficient evidence that the population means in the two groups are different, and so, the distributions are not equal. However, if we do not reject the null hypothesis, (non-significant test, with $p > 0.05$) then it indicates that the population means in the two groups are equal.

The Python function used for this test was `ttest_ind()` from the statsmodels library [46], stats.weightstats module. Due to test assumptions, this function was only applied to the variables that presented a normal distribution (verified by Shapiro-Wilk test) and homogeneity of variances (verified by Levene's test).

Mann Whitney U Test

The Mann-Whitney is the non-parametric equivalent to the unpaired t-test. This test was named for Henry Mann and Donald Whitney, although it is sometimes called the Wilcoxon-Mann-Whitney test, from Frank Wilcoxon, who also developed a variation of the test or Wilcoxon's rank sum test. This test is a statistical significance test used to determine whether two independent samples were drawn from a population with the same distribution [49] [51].

- Null hypothesis (H0): Sample distributions are equal.
- Alternative hypothesis (H1): Sample distributions are not equal.

So, if we do not reject the null hypothesis ($p > 0.05$) it means that both samples were drawn from a population with the same distribution. On the other hand, we conclude that sample distributions are not

equal when we reject the null hypothesis and that happens when the test is significant ($p < 0.05$). This test was applied through the *mannwhitneyu()* function from the stats module of the scipy library [45]. Since this test is the non-parametric equivalent to the unpaired t-test, it was only applied to those variables which were not normally distributed (evaluated previously by the Shapiro-Wilk test).

Post Hoc Tests

Post hoc tests, also referred to as a posteriori or unplanned tests, or even multiple comparison tests, are statistical tests that are applied after a significant F test. As explained, the previous tests, i.e, one-way ANOVA, Welch ANOVA and Kruskal-Wallis only indicate that there is a significant means difference between the groups, they do not indicate in which groups pair or pairs that difference occurs. To overcome this, post hoc tests allow comparisons between a pair of group means in order to specify which of the pairs of means are statistically significantly different from each other. This is almost similar to taking every pair of groups and performing a t-test on each pair, however, pairwise comparisons have an advantage, they control the familywise error by correcting the level of significance for each test, so that the overall Type I error rate across all comparisons remains at 0.05 [50] [54] [55].

The Tukey test uses pairwise post hoc testing to determine which of three or more sample means are significantly different after obtaining a significant ANOVA F-test. Thus, this test is an ANOVA post hoc test, which is usually known as Tukey's HSD (Honestly Significant Difference), however, the Tukey HSD is only used when the samples sizes for the groups are equal. This test was modified by Kramer, so that it could be used in cases where the samples sizes of the groups are not equal and it is called the Tukey-Kramer method [56] [57] [58]. Then, this test was only applied on the variables that gave a significant one-way ANOVA, through the *pairwise_tukey()* function from the pingouin library [47]. This function applies the pairwise Tukey HSD post hoc test, however, if the samples sizes are not equal, it automatically uses the Tukey-Kramer method.

The Games-Howell test is an extension of the Tukey-Kramer test and is appropriate in cases with unequal variances as well as unequal sample sizes. The test controls the type I error for the whole comparison and also maintains the default significance level even when the size of the sample is different. However, this test may be too liberal for small sample sizes and therefore should only be used in sample sizes greater than five [56] [59]. Since this test is indicated for cases with unequal variances, Games-

Howell is the post hoc test of Welch ANOVA, and, therefore, is applied only in the variables that had a significant result in Welch ANOVA test. The Python function used for this test was *pairwise_gameshowell()* from the pingouin library [47].

There is also Dunn's test which is a post hoc pairwise test for multiple comparisons of mean rank sums. This test is the appropriate procedure after a significant Kruskal-Wallis test [60]. For Dunn's test, the Python function used was *posthoc_dunn()* from the scikit-posthocs package [40].

5.2.4 Differences between independent samples: qualitative data

Chi-Squared Test

Chi-squared test, also known as Pearson's Chi-squared test, named for Karl Pearson, is a statistical hypothesis test to determine whether there is a relationship between two categorical variables (factors), more specifically, whether the proportions of individuals who possess a certain characteristic are the same in the independent groups. This data can be represented in a $r \times c$ contingency table, with r rows and c columns, in which for each cell we calculate the expected frequencies, then, determine whether the division of the groups, called the observed frequencies, matches the expected frequencies. The result is a test statistic that has a Chi-Squared distribution, denominated for the Greek lowercase letter chi (χ) [49] [50] [51].

- Null hypothesis (H_0): There is no association between the categories of one factor and the categories of the other factor in the population.
- Alternative hypothesis (H_1): The two factors are associated in the population.

So, if the test is significant ($p < 0.05$) we should reject the null hypothesis and accept the alternative one, that the two factors are associated in the population. On the other hand, if the test is non-significant ($p > 0.05$), we should not reject the null hypothesis, so, there is no association between the categories of one factor and the categories of the other factor in the population, that is, they are independent.

The Python function used for Chi-Squared test was *chi2_contingency()* from the *scipy* library [45], *stats* module. This function was applied to determine whether there was a relationship between Sex (Female or Male) and group, after the acquisition of the 3x2 contingency table with these proportions.

Fisher's Exact test

Fisher's exact test is normally used when there are two independent groups and we are interested in whether the proportions of individuals who have a characteristic are the same in both groups. This test should be applied in small samples (where the expected frequencies are small) since it was designed to overcome a problem with these samples which is that the sampling distribution of the chi-square statistic differ substantially from a chi-square distribution. In fact, Fisher's exact test is mostly a way of computing the exact probability of the chi-square statistic, instead of a test [49] [50].

This test can also be used as post hoc test by applying a 2x2 Fisher's exact test to each of the pairwise comparisons, but then correct for multiple comparisons using for example the Bonferroni correction (used to control the familywise error rate) [58].

- Null hypothesis (H0): the proportions of individuals with the characteristic are equal in the two groups in the population.
- Alternative hypothesis (H1): these population proportions are not equal.

Identically to other hypothesis tests, when the Fisher's exact test is non-significant ($p > 0.05$), we must not reject the null hypothesis, and in this case that means that the proportions of individuals with the characteristic are equal in the two groups in the population. On the other hand, if the test is significant ($p < 0.05$) we should reject the null hypothesis and accept the alternative one, which indicates that these population proportions are not equal.

This test was applied to evaluate whether the proportions of female and male individuals (sex variable) are the same in two groups. That way, the Python function *fisher_exact()* from the *scipy* library [45], *stats* module was used, after obtaining the 2x2 contingency table. When used as post hoc test, the function *multipletests()* from *statsmodel* library, *stats.multitest* module, was also applied. This function takes the pvalues obtained in the Fisher's exact test as argument and, the "method" argument equal to "bonferroni".

5.3 Logistic Regression Analysis

Logistic regression is used when the outcome/dependent variable (Y) is categorical and there is one or more predictor/explanatory/independent variables (X) that are continuous or categorical. When the outcome variable has only two possible categorical outcomes, this regression is known as binary logistic regression. When we analyze the relationship between one independent variable and one dependent variable, we often call it univariate analysis. When we analyze the relationship between two or more independent variables and one outcome variable (dependent) we call it a multivariable analysis. In the case of relationships of several predictors with two or more outcome/dependent variables at the same time, it is called a multivariate analysis [49] [50] [61] [62].

The function used to express the relationship between the predictor variables and the binary outcome variable is the logistic, this is why it is called logistic regression. The logistic regression equation enables the determination of which explanatory variables influence the outcome and, using an individual's values of the explanatory variables, evaluate the probability that the individual will have a particular outcome. A multivariable logistic regression model (two or more explanatory variables) is expressed as [49] [50] [61]:

$$\text{logit}(p) = a + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (5.1)$$

Where,

$$\text{logit}(p) = \ln \frac{p}{1-p} \quad (5.2)$$

In these equations:

- x_i is the i th explanatory variable ($i=1,2,3,\dots,k$);
- p is a binomial proportion;
- a is the estimated constant term;
- $b_1, b_2, \dots, b_k$ are the estimated logistic regression coefficients.

In a logistic regression model, the odds ratio (OR) is a measure of association between an explanatory variable and the outcome. The OR represents the odds that an outcome will occur given a particular explanatory variable, compared to the odds of the outcome occurring in the absence of that predictor variable. The OR is obtained by the exponential of a coefficient. For example, $\exp(b_1)$ is the estimated odds of an outcome for $(x_1 + 1)$ relative to the estimated odds of an outcome for x_1 , while adjusting for all other x 's in the equation. Therefore, using the $\exp(b_1)$ example, it means that when there is a one unit change in the x_1 and the other explanatory variables are held constant, the odds of $Y = 1$ change by $\exp(b_1)$ units. When the odds ratio value is above one this indicates an increased odds of having the outcome linked to $Y = 1$, on the other hand, if the odds ratio value is below one, then it indicates a decreased odds of having the outcome linked to $Y = 1$, as the particular explanatory variable increases by one unit [49] [61].

It is possible to obtain a Wald test statistic for each explanatory variable [49]:

- Null hypothesis (H_0): The relevant logistic regression coefficient is zero.
- Alternative hypothesis (H_1): The relevant logistic regression coefficient is different from zero.

It is also possible to evaluate all explanatory variables and their effect on the outcome, for example with chi-square for covariates [49]:

- Null hypothesis (H_0): All the logistic regression coefficients in the model are zero.
- Alternative hypothesis (H_1): At least one logistic regression coefficient is different from zero.

For each Holter variable, a univariate logistic regression analysis, assuming the Holter variable as the explanatory variable and the presence of one of the diseases as the outcome, was performed, using the `Logit()` function from the `statsmodels` library [46], `discrete.discrete_model` module.

As the association of a Holter variable with the presence of a disease could be affected by Sex and Age differences [63] [64], for each Holter variable, two multivariable logistic regression analyses were also performed: 1) Age and Holter variable as explanatory variables; and 2) Sex and Holter variable as explanatory variables. These logistic regression models have also been applied through the `Logit()` function.

Chapter 6

The implemented Machine Learning methods

This chapter incorporates a description of the different steps that were applied to our data in order to achieve the final classification machine learning models. The main steps are presented in the flowchart of Figure 1. As can be seen, the methodology is divided into two different phases: feature selection and classification.

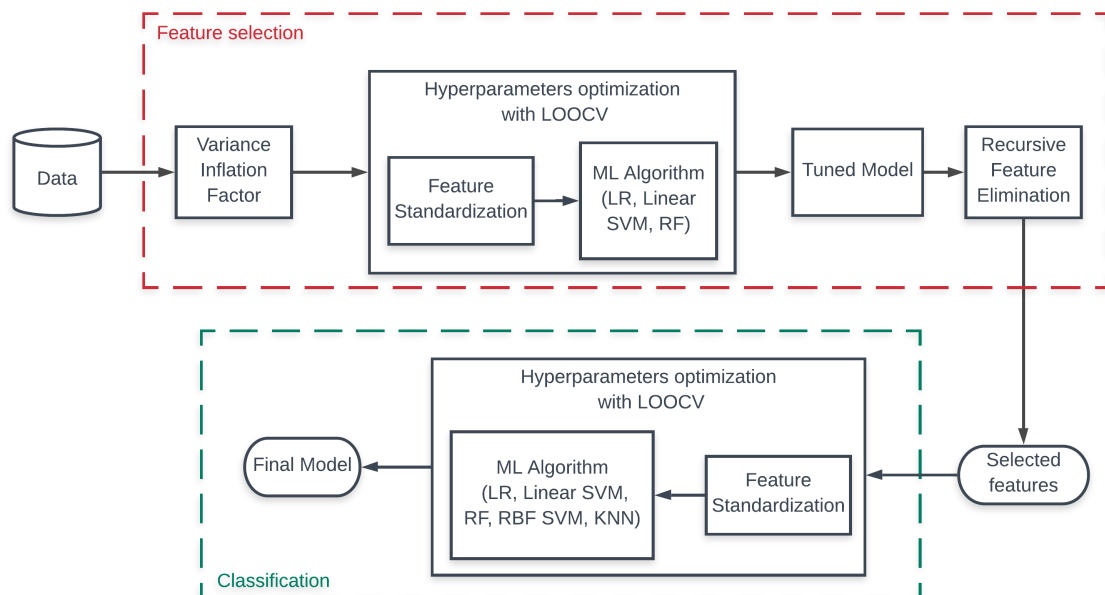


Figure 1: Methodology machine learning-based flowchart.

First, data were submitted to a filter supervised feature selection method. The goal was to determine if there was any multicollinearity between the independent variables in this study. This was done by determining the variance inflation factor (VIF) values and the highly correlated variables (which have a great VIF value) were excluded from the set through a stepwise procedure.

Then, for each of the 3 machine learning algorithms (logistic regression (LR), support vector machines with linear kernel (linear SVM), and random forest (RF)), the hyperparameters values/status that adjusted better to the data (higher accuracy), were discovered. This is called hyperparameter tuning and was done with a grid search strategy using the standardized data to train the models and leave-one-out cross-validation (LOOCV) to evaluate the performance of the models.

After the hyperparameters tuning each of the improved models was used in a second feature selection method. This time a wrapper feature selection method was performed, the recursive feature elimination (RFE), to discover which number and combination of features lead to better classification performances.

The just described procedures are part of the feature selection phase. Now that features are selected, the classification phase takes place.

Each of the feature combinations was used to train 5 different machine learning models: logistic regression (LR), SVM with linear kernel (linear SVM), SVM with radial basis function (RBF) kernel (RBF SVM), random forest (RF), and k-nearest neighbors (KNN). Since the data that is used in these models is not the original data, the hyperparameters tuning procedure, described previously, is also applied in the classification phase. This way, the final classification models are achieved.

The described methodology is applied to three distinct binary classification problems. In one problem the classes indicate whether an FD patient has WMLs (class 1) or not (class 0), the other whether a patient has FD with WMLs (class 1) or has HCM (class 0), and the last, whether a patient has FD without WMLs (class 1) or has HCM (class 0).

6.1 Cross-validation

When we are building a machine learning model, we are interested in evaluating its performance, mainly, in discovering how well the model performs on unseen data.

Overfitting is the most common problem in applied machine learning and it leads to models with poor performance. Overfitting happens when a model is too adjusted to the training data, that is, when the model actually learns too much of the detail and noise of the training data. This results in a model with poor performance because the model interprets the learned noise as concepts and, with new data, these concepts do not exist, so, the model's performance on the new data is negatively affected. This is a huge problem because this is the performance that we are most interested in, not the training performance [65] [66].

To overcome this problem we can use cross-validation to estimate the model's performance on unseen data. In cross-validation, the training dataset is split into train set and validation set, in different ways (different observations in each set, each time) using some strategy and then, multiple iterations of each model on these different splits are built [65] [66].

There are several cross-validation strategies that differ only in the way that the initial train set is divided into the train set and validation set. In this study leave-one-out cross-validation (LOOCV) is the used strategy since several studies defend its use to evaluate the performance of a classification model when the sample size is small, which is our case. In LOOCV, the validation set includes only an observation (in our case, a patient), and the remaining observations, $n-1$ (being n the sample size), are part of the train set, thus, this strategy performs n iterations of each model [65] [66] [67] [68].

The scikit-learn [40] library has a *LeaveOneOut* class that was used in this study to apply the leave-one-out cross-validation. As previously mentioned, this cross-validation strategy is incorporated in the hyperparameter tuning process, which will now be explained.

6.2 Hyperparameter Tuning

First, it is essential to understand what hyperparameters are. Hyperparameters are meta parameters associated with a machine learning algorithm. Since these hyperparameters are not derived from the training dataset, they are set before this phase and before the building process. They are extremely important for tuning the performance of learning algorithms [66].

Cross-validation, that was previously explained, can be used for tuning the algorithms by training the models with the train data and evaluating their performance with the validation data [66].

The search of the parameter space for an optimal model can be done through two major ways: systemic and random. Grid search is a systemic method and the simplest method for hyperparameter optimization. First, we specify a grid of hyperparameters values that we want to attempt and we optimize to get the best parameter combinations. Then, a model is built for each of these combinations of hyperparameters values, using cross-validation, and the best combination in the whole grid is reported. The model using the best combination is the output of this search method. Grid search is a simple method, however, a significant disadvantage of this method is the fact that the user has to manually supply the hyperparameters values and he may or may not choose and supply the most optimal hyperparameters values [66].

In this study, a pipeline was created for each of the machine learning algorithms. This pipeline included a standardization step, applied through *StandardScaler* class from scikit-learn [40], and the machine learning model. After the creation of the pipeline, hyperparameters were tuned using a grid search strategy. Since we wanted to evaluate the model's classification performance, cross-validation is incorporated in the grid search, and performance metrics were calculated to enable the selection of the best hyperparameter value/status.

Each hyperparameter was evaluated separately, so, first, the hyperparameter's values are ranged in a large scale (when numeric), and then, according to which hyperparameter value gave the highest performance, a more precise search around that same value was performed. Then, the value that led to the highest accuracy was selected and held while the following hyperparameters were evaluated in the same way.

For the logistic regression model, the only hyperparameter that was tuned was C (inverse of regulariza-

tion strength). The used solver (the algorithm used in the optimization problem) was “newton-cg” and all the remaining hyperparameters values/status were the default. In the SVM model, since we already had decided to use a linear kernel, this parameter was set to “linear” and the only evaluated hyperparameter was C (regularization parameter). In the case of the random forest model, several hyperparameters were evaluated: bootstrap, n_estimators (the number of trees in the forest), max_depth (the maximum depth of the tree), max_leaf_nodes, min_samples_leaf (the minimum number of samples required to be at a leaf node), and min_samples_split (the minimum number of samples required to split an internal node). In the classification phase, two different machine learning models were added to the models evaluated in the feature selection phase, which were the SVM with RBF kernel and the KNN. For this SVM model, the kernel was set to “rbf”, and besides C, gamma parameter was also evaluated. As for the KNN model, the parameters n_neighbors (number of neighbors to use by default for k-neighbors queries) and weights (weight function used in prediction) were evaluated. The remaining parameters of the models were used with the default value/status.

The grid search method was applied using the *GridSearchCV* class from the scikit-learn [40]. The *refit* parameter of this class was set to “Accuracy” and, as previously mentioned, the cross-validation method (cv parameter) was *LeaveOneOut*, and the classification model was passed within a pipeline.

6.3 Feature Selection

Feature selection is the procedure of excluding, from a predictive model, the predictors (independent variables) that are not relevant in the prediction of the target variable (independent variable), and, that, overall, lead to higher training times and system memory use, and can even degrade the model’s performance. However, this selection should not be applied to a point that, in fact, compromises the model’s performance [69] [70].

Feature selection can be divided into two groups, supervised selection, and unsupervised selection, according to whether the dependent variable is used in the selection or not, respectively. The supervised selection methods can also be classified into three classes: intrinsic or implicit, filter, and wrapper meth-

ods. In the intrinsic methods, feature selection is done automatically by the algorithms during training. In turn, the filter methods select subsets of features based on the scores obtained when evaluating their relationship with the target variable through statistical techniques. As for the wrapper methods, they are iterative methods that create several models with different subsets of features and then select the features that yield the best model's performance, which is evaluated through a performance metric [69] [70].

The wrapper methods can either take a greedy or non-greedy approach. A greedy search means that a path (feature combination) is chosen according to the direction that, at the time, seems the best to achieve the best immediate benefit. On the other hand, a non-greedy search has the ability to re-evaluate the chosen path and change the direction to one that appears to have a potential benefit in the next steps, even if it was not a favorable path initially. The greedy approach methods may fail when there are important interactions between the predictors, like collinearity and multicollinearity, being only one of the predictors significant in the presence of the other(s) [69] [70].

Considering the importance of feature selection, in this study, two different feature selection methods were applied, a filter method, based on variance inflation factor (VIF) and also one of the most used wrapper methods, the recursive feature elimination (RFE). Both methods will be briefly described in this section.

6.3.1 Variance Inflation Factor

We are dealing with multicollinearity when there is a correlation between two or more explanatory variables (predictors). Since these variables are correlated to each other, they can, likewise, be correlated with the target variable [71] [72].

Multicollinearity can be estimated through the variance inflation factor (VIF), which assesses how much the variance of an estimated regression coefficient increases when predictors are correlated. The VIF can be calculated according to Equation 6.1.

$$VIF = \frac{1}{1 - R^2} \quad (6.1)$$

As can be observed, VIF is dependent on R^2 , which is the coefficient of determination, obtained

from a multiple linear regression model whose target variable is one of the explanatory variables and the explanatory variables (predictors) are the remaining explanatory variables [71] [72].

A VIF equal to 1 means that there is no multicollinearity among predictors, however, if VIF is greater than 5 we are in the presence of multicollinearity. In this case, a solution may be to remove the highly correlated predictors, which strength of multicollinearity in the model is responsible for making the other predictors redundant [71] [72].

Therefore, in this study, a function, in Python, was implemented to determine the VIF for each variable. The arguments of this function were the data frame with all variables data and an array with variables designations, that is, the data frame's columns' names. Each variable was considered as the target variable, and the remaining as explanatory variables, then, a linear regression model was applied using the *LinearRegression()* from the scikit-learn library [40], and R^2 value was obtained. Thereafter, VIF value was determined using the equation presented above. This procedure was applied for each variable, so, each variable had a VIF value. After obtaining all VIF values, the variable with higher VIF (>5), was eliminated and VIF values were once more determined and, as previously, the variable with higher VIF was eliminated. This procedure of obtaining the VIF values and excluding the variable with higher VIF was done until there were only variables with VIF values under 5 so that relevant multicollinearity among predictors does not exist.

6.3.2 Recursive Feature Elimination

As previously mentioned, recursive feature elimination (RFE) is a wrapper feature elimination method, and, therefore, a different machine learning (ML) algorithm is wrapped by RFE and used to help select features [69] [70].

First, a model, the desired machine learning model, is built with all features, then, features are ranked by importance and the least important ones are removed, after that, the model is re-built and importance scores are, once again, determined. This procedure is repeated until the desired number of features, which can be chosen, as well as each subset size, is reached. The optimal subset of features is then used to train the final model [69] [70].

Although multicollinearity (highly correlated features) does not affect predictive performance in tree-based models, it can negatively affect the importance scores. Thus, when using RFE with random forest, for example, filtering highly correlated features before the application of recursive feature elimination is suggested [69] [70].

This feature selection method was implemented in Python (scikit-learn library [40] mainly) using the *RFE* class, which received the number of features to select and the desired ML model. RFE estimator was then incorporated in a pipeline (*Pipeline* class), along with a scaler (*StandardScaler* class) and a machine learning model (*RandomForest* class, for example). Next, in this pipeline, the explanatory variables and the target variable were used as arguments in the *cross_val_score()* function, where leave-one-out cross-validation is used (*LeaveOneOut* class), to obtain the model's accuracy (also passed as argument). A “for” cycle was implemented so that the number of features to select ranged from 1 to the maximum number of features and the procedure just described was applied in each of these situations. Finally, a plot with the scores in function of the number of features was obtained, as well as the maximum score and the respective subset of features that led to that score.

6.4 Machine Learning Classification Algorithms

Machine Learning methods can be separated into different groups, according to how they learn: supervised, unsupervised, semi-supervised, and reinforcement learning [65] [66].

Most practical machine learning uses supervised learning. In this case, an algorithm is used to learn the mapping function from the input variables (X) to the respective output variable (Y), that is, the “f” in the term $Y=f(X)$. The goal is to achieve such a good mapping function approximation, that, in the future, the model would be able to predict an output (Y) for any new input data sample (X). It is called supervised learning since the model already knows the output labels/responses of the data samples in the training phase [65] [66].

However, in some cases, we do not have the output labels for our data samples. In these situations, we can still use machine learning methods, the ones with unsupervised learning that enables the extraction

of useful patterns and insights from our data. These methods are called unsupervised learning since there are no output variables, so the algorithm just tries to learn relationships and patterns of the data without any help, as it happens in supervised learning [65] [66].

There are also semi-supervised methods, which fall between the supervised and unsupervised learning methods, and, therefore, a mixture of both techniques can be applied. These methods are used when we have a large amount of input data (X) and only some of it is labeled (Y) [65] [66].

In the reinforcement learning methods, an agent is trained for a period of time by interacting with a specific environment and improving its performance in relation to the actions that performs on the environment. To start, the agent has a set of strategies or policies for interacting with the environment. Then, by observing the environment state, the agent takes an action based on a policy and gets a reward for this action, that could be beneficial or a penalty, and, considering the reward, the agent, if needed, updates its current strategies and policies. This iterative process continues until the agent had learned enough about the environment and gets the desired rewards [65] [66].

This study included patients that have already been diagnosed, so we had the output variable (Y), and, therefore, we will focus on the supervised learning methods from now on.

Supervised learning methods can be further divided into two types of problems, classification problems and regression problems. The main difference between these two is the type of the output/response variable, when this variable is categorical (for example, disease or no disease), in this case also known as classes, we are dealing with a classification problem. In this type of problem, the purpose is to predict the categorical response variable, from the input data, based on what the model has learned in the training phase. Some classification algorithms are support vector machines, logistic regression, neural networks, ensembles like gradient boosting and random forest, and K-nearest neighbors (KNN). However, in the regression problems, the output variable is continuous (for example weight) and the purpose of these problems is to estimate the output value for a given unknown input data. This is possible because, in the training phase, the model has learned specific relationships and associations between the features (explanatory/independent variables) and the respective continuous numeric output values (response/independent variable). Some examples of regression problems are linear regression, polynomial regression, and non-linear regression [65] [66].

We have already mentioned that we had the output variable (y) for our input data (X), but to be more

specific, this output variable is categorical, in fact, is dichotomous, since it is either having or not having a condition/disease. So, as explained, this is a classification problem, and, therefore, the models applied to our data are logistic regression, support vector machine, random forest, and also K-nearest neighbors.

6.4.1 Logistic Regression

Logistic regression was already described in Section 5.3 where it was used to express the relationship between the predictor variables and the binary outcome variable. This time we are using logistic regression as a predictive model. However, the logistic regression function is the same in both situations, the only difference is that when using logistic regression as a predictive model, we want to calculate the probability of the default class (the first class), and, therefore, we use Equation 6.2, which is equivalent to Equation 5.2 (introduced in Section 5.3) [65] [66].

$$p(x) = \frac{e^{a+b_1x_1+b_2x_2+\dots+b_kx_k}}{1 + e^{a+b_1x_1+b_2x_2+\dots+b_kx_k}} \quad (6.2)$$

The logistic regression coefficients ($b_1, b_2, \dots, b_k$) are estimated from the training data using maximum-likelihood estimation (common learning algorithm). After estimating these coefficients, they are used in the equation, as well as the input vector x ($x_1, x_2, \dots, x_k$), to calculate the probability ($p(x)$) [65] [66].

Then, the probability prediction is transformed into a binary value/ class (0 or 1) [65] [66]:

$$prediction = \begin{cases} 0, & \text{if } p(x) < 0.5 \\ 1, & \text{if } p(x) \geq 0.5 \end{cases} \quad (6.3)$$

It is important to notice that, when there are highly correlated inputs, the likelihood estimation process can fail to converge and the logistic regression model can overfit. Therefore, it is important to remove highly correlated inputs before applying the logistic regression [65] [66].

In this study, the logistic regression model was created using the *LogisticRegression* scikit-learn's class [40].

6.4.2 Support Vector Machines

The support vector machine (SVM) model is a supervised machine learning algorithm that finds a hyperplane, in a multidimensional space, that is able to separate the training data by labels/class. For example, if we have two known labels and two input variables (features), then, this forms a two-dimensional space. In this case, the selected hyperplane, to separate the points in the space by their class, is a line. The perpendicular distance between the line and the closest points is known as the margin. The maximal-margin hyperplane is the name given to the optimal hyperplane, the one that separates the classes with the largest margin. The closest points are, in fact, the only relevant points to define the hyperplane and to construct the classifier. These points are called the support vectors. Figure 2 illustrates a possible two-dimensional space situation with the respective hyperplane, support vectors, and margin [65] [73].

After discovering the hyperplane, the equation that defines it can be used to make predictions, by inserting the input values into the equation. The label of the new input variable is indicated by whether the point is above or below the hyperplane (easier to understand when the hyperplane is a line) [65] [73].

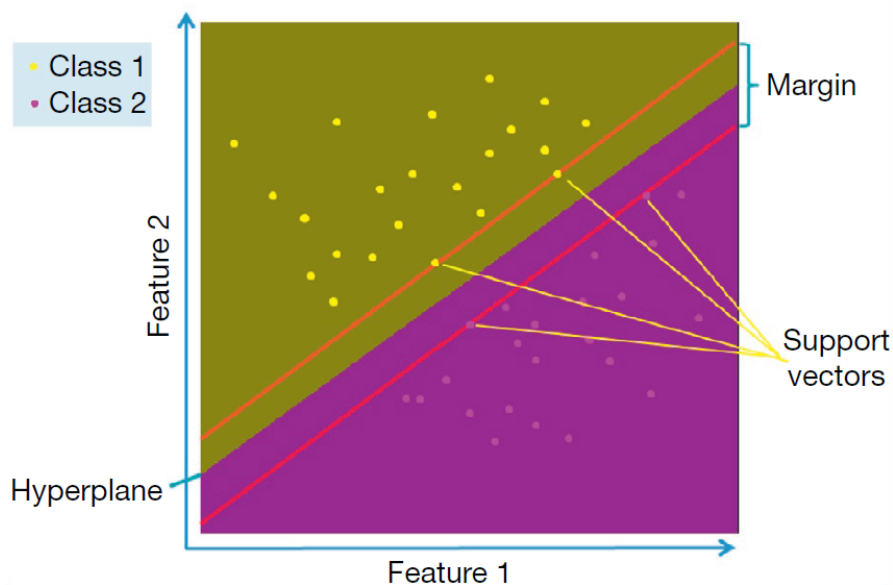


Figure 2: Example of support machine classification results [73].

However, the maximal-margin classifier is a hypothetical classifier, since real data may not be perfectly

separated by a hyperplane. So, in real situations, the margin is not maximized, and some points of the training data are allowed to violate the separating line. The amount of violation of the margin that is allowed is defined by the parameter C , being that larger C values indicate that more violations of the hyperplane are allowed. Thus, C parameter also influences the number of support vectors used by the model [65] [73].

In practice, the SVM algorithm is implemented using a kernel. This kernel can be linear, polynomial or radial. In this study, linear and radial kernels were used. The linear kernel uses the dot-product and defines the similarity between new data and the support vectors. The radial kernel can create complex regions and, for this kernel, a gamma parameter should be specified to the algorithm [65] [73].

We obtained SVM classifier using the *SVC* class from scikit-learn library [40] and set the kernel parameter to “linear” or “rbf” according to the desired model.

6.4.3 Random Forest

Random Forest is an ensemble method based on bagging (bootstrap aggregating) where independent classification or regression un-pruned decision trees are aggregated (using majority vote for classification or averaging for regression) to produce a more accurate classification [65] [74] [75].

Therefore, first, let's understand what are decision trees and how they work. Decision trees, or CART (classification and regression trees), as indicated by the name, are inspired in the ordinary tree structure, which is composed of a root and nodes (the positions where the branches divide), branches, and leaves. Similarly, a decision tree is formed from nodes that are represented by circles and the branches that are represented by the segments that connect the nodes. The decision tree starts in the root node, moves downward and ends in the leaf node and normally is drawn from left to right. Each internal node (a node that is not a leaf node) can have two or more branches. A node represents a certain characteristic and the branches represent a range of values. The described decision tree structure can be observed in Figure 3 [65] [74] [75].

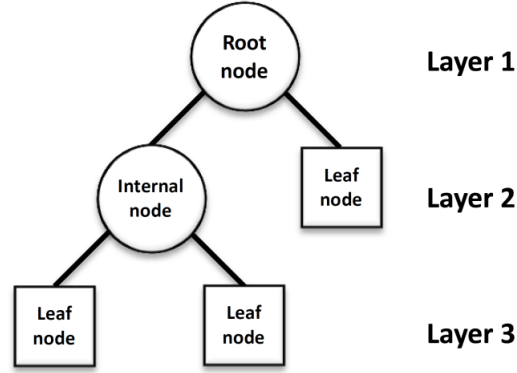


Figure 3: Decision tree structure [76].

Predictions can be made by visiting the tree and performing the tests included in the root and internal nodes, and, when a leaf node is reached, the predicted label corresponds to the reported class [65] [74] [75].

To build a decision tree, a greedy approach (the very best split point is chosen in each time) is used to divide the input space, called recursive binary splitting. In this procedure all the values are lined up and then, using a cost function, different split points are tried and tested. The split that presents the lowest cost is the chosen one. For classifications problems, the cost function used is the Gini, which indicates how pure the leaf nodes are. Gini cost (G) is calculated as shown in Equation 6.4, where $\hat{p}_k$ is the fraction of training instances with class K in the rectangle of interest [65] [74] [75].

$$G = \sum_{k=1}^n \hat{p}_k (1 - \hat{p}_k) \quad (6.4)$$

Thus, a node that has all the training instances of the same class (perfect class purity), has a $G = 0$. On the other hand, G reaches its maximum value ($G = 0.5$) when there is a 50-50 split of classes ($\hat{p}_k = \frac{1}{2}$) [65] [74] [75].

The recursive binary splitting requires a stopping procedure to know when it should stop splitting. This stopping procedure can be limiting the minimum level of impurity to split a node or the minimum number of samples per node. In the second procedure, if the count of training instances is less than the minimum, then, the split is not accepted and the respective node is taken as a final leaf node [65] [74] [75].

However, rigorous stopping criteria can lead to underfitting trees. To solve this problem, pruning

criteria is used instead. So, better performances can be achieved by training an unrestricted tree and then pruning it using for that, for example, the weakest link pruning procedure on which a learning parameter (α) is used to weigh whether nodes can be removed based on the size of the sub-tree [65] [74] [75].

Although decision trees present desirable properties and success, their discrete nature involves high variance in predictions. As previously explained, random forest is an ensemble method, and these type of methods allow the construction of models that produce better predictions by aggregating the predictions of a set of individual classifiers. Random forest is based on bagging which is a procedure that reduces the variance for algorithms that have high variance, like decision trees [65] [74] [75].

As previously explained, decision trees are greedy. When dealing with multiple decision trees, these trees can have structural similarities and, so, highly correlated predictions. The ensemble methods are more accurate if the predictions from the sub-trees are weakly correlated, ideally, uncorrelated. Thus, in random forest, the sub-trees learn in a way that their predictions have less correlation. In decision trees (or CART) the learning algorithm, when selecting a split point, is allowed to look to all variables and all their values to select the most optimal split-point. However, in random forest, the learning algorithm search is limited to a random sample of features. The number of features that can be searched at each split point must be specified as a parameter to the algorithm [65] [74] [75].

In Python, random forest model was implemented through the *RandomForestClassifier* class, from the scikit-learn library [40].

6.4.4 K-Nearest Neighbors

In the K-nearest neighbors (KNN) algorithm, no learning is in fact necessary, all the procedure of KNN is made when a prediction is required. The only model in KNN is storing the entire training dataset. So, KNN stores the training dataset and makes a prediction for a new data point by searching, in the entire dataset, for the K most similar instances, called the neighbors. A distance measure is used to find out which K instances, in the training dataset, are most similar to the new input. Euclidean distance is the most common distance measure for real-valued input variables and it is the measure used in this study. However, there are other distance measures like Hamming distance and Manhattan distance [65].

The Euclidean distance is calculated as the square root of the sum of the squared differences between a point a and point b across all input attributes i, as presented in Equation 6.5 [65].

$$EuclideanDistance(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (6.5)$$

For classification problems, the class of the new input point can be predicted by determining, from among the K most similar instances, the class with higher frequency. Thus, for a new input instance, we determine the K most similar instances and, then, for this set, we calculate the normalized frequency of samples that belong to each class, as exemplified in Equation 6.6 [65].

$$p(class = 0) = \frac{count(class = 0)}{count(class = 0) + count(class = 1)} \quad (6.6)$$

This algorithm is more indicated for lower dimensional data, so it can benefit from feature selection to reduce the dimensions of the input feature space [65].

The KNN version for classification problems was achieved with the *KNeighborsClassifier()* class from the scikit-learn library [40].

6.5 Performance Metrics

To evaluate a model's performance some performance metrics are measured in the validation phase, such as accuracy, sensitivity, specificity, and area under curve (AUC).

To better comprehend these metrics, first, we should understand what is a confusion matrix. So, a confusion matrix is not a metric but is a way to evaluate a classification model since its representation can be used to delineate several metrics. This representation is obtained by comparing the predicted labels, that were given by the classifier, with the true labels, that are already known. A typical confusion matrix for a binary classification problem is presented in Figure 4, where p designates the positive class and n designates the negative class and where the columns represent the predicted labels and the rows the true labels. To note that the confusion matrix can also be represented with the predicted labels in the rows and

the true labels in the columns. As can be seen, the comparison between predicted and true labels leads to four different terms: true negative, false negative, false positive, and, true positive [66].

- True Positive (TP): Total counts of when the true label is positive and the predicted label is positive.
- True Negative (TN): Total counts of when the true label is negative and the predicted label is negative.
- False Positive (FP): Total counts of when the true label is negative and the predicted label is positive.
- False Negative (FN): Total counts of when the true label is positive and the predicted label is negative.

		PREDICTED LABELS	
		n' (Predicted)	p' (Predicted)
TRUE LABELS	n (True)	True Negative <i>(Number of instances of negative class 'n' correctly predicted)</i>	False Positive <i>(Number of instances of negative class 'n' incorrectly predicted as the positive class 'p')</i>
	p (True)	False Negative <i>(Number of instances of positive class 'p' incorrectly predicted as the negative class 'n')</i>	True Positive <i>(Number of instances of positive class 'p' correctly predicted)</i>

Figure 4: Confusion Matrix structure [66].

Thus, from the confusion matrix, it is possible to obtain the desired metrics. First, the accuracy, this metric is one of the most popular to evaluate a classifier performance and it is defined as the proportion of correct predictions of the model, since, as illustrated in Equation 6.7, it is the number of correct predictions (true) over the total amount of predictions (equal to the sample size) [66].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.7)$$

Sensitivity, also known as recall, is a measure to identify the percentage of relevant data points. Sensitivity is calculated as presented in Equation 6.8 and it is defined as the number of cases of the positive class that were correctly predicted [66].

$$Sensitivity = \frac{TP}{TP + FN} \quad (6.8)$$

Specificity, in turn, is defined as the number of cases of the negative class that were correctly predicted and is measured as presented in Equation 6.9 [77].

$$Specificity = \frac{TN}{FP + TN} \quad (6.9)$$

The receiver operating characteristic (ROC) curve is also an important visual tool to interpret classification models that is obtained by plotting the true positive rate (TPR), which is equal to sensitivity, versus the false positive rate (FPR), which is equal to 1 - specificity. So, each prediction result, obtained from the confusion matrix, takes one point in the ROC space, which is between (0,0) and (1,1). An illustration of a ROC curve is presented in Figure 5. In this figure, we can see a diagonal line, which corresponds to the ROC curve of a classifier that does a random guess. An ideal model would result in a ROC curve at the (0,1) point, however, this is not realistic, so, as close is the ROC curve to this point, which is situated in the top left corner of the graph, the better is the classifier [66] [77].

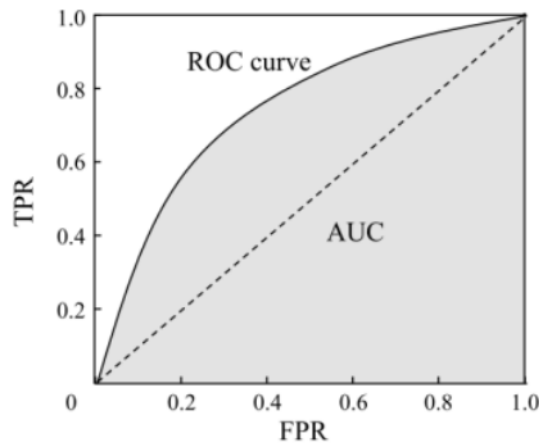


Figure 5: Example of ROC curve and AUC [78].

In order to compare models, a metric is calculated from the ROC curve, which is called area under curve (AUC). The ROC curve illustrated in Figure 5 has the respective area under curve, which is the gray-colored area. This metric can be calculated by integrating the areas under the steps of the ROC curve. So, if ROC curve is obtained by connecting the points $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, where $x_1 = 0$ and

$x_m = 1$, then, AUC is estimated as presented in Equation 6.10. Since this metric represents the area under the ROC curve, its value will always be between 0 and 1. An ideal classifier would result in a unit AUC value, and, due to the diagonal ROC curve, a realistic classifier has an AUC value above 0.5 (area under the diagonal). Thus, with the AUC it is possible to compare models, and generally, the best model is the one that presents the higher AUC value [66] [77].

$$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i) \cdot (y_i + y_{i+1}) \quad (6.10)$$

In this study, these metrics (accuracy, sensitivity, specificity, and AUC) were calculated during the hyperparameter tuning process to enable the selection of the best hyperparameter value/status, however, the decision relied on the accuracy, and, in case of ties, on the AUC.

These performance metrics were obtained, in Python and using scikit-learn [40], in different ways. The accuracy was obtained by accessing the mean of the validation results produced by the *accuracy_score()* function. The AUC was obtained applying the *auc()* function to the FPR and TPR values given by the *roc_curve()* function. As for the sensitivity and specificity, first, the TP, TN, FP, and FN values were obtained for each cross-validation iteration by comparing the true and predicted labels, and then, these two metrics were determined using the formulas previously presented in Equation 6.8 and 6.9, respectively.

Part III :

**Results, Discussion, Conclusion, and Future
Work**

Chapter 7

Statistical Results

As indicated by the name, in this chapter are the results obtained by the application of the statistical methods (previously described). These results are organized by age class, first the 18-91 years class (all patients), then the 40-59 years (inclusive), and last the 60-91 years class.

7.1 Results from patients over 18 years old

The descriptive results of each variable per group (HCM, FD with WMLs and FD without WMLs) and comparative results are shown in Table 6. Note that data with a normal distribution is presented as mean $\pm$ standard deviation, and data with a non-normal distribution is presented as median (quartile 1 - quartile 3). The normality and homogeneity of variance results, obtained through the Shapiro test and Levene's test, respectively, can be consulted in Table 22 of Appendix A.

Group differences were found for Sex, Age, HR Min, HR Mean, HR Max, QT Min, QT Mean, QT Max, QTc Mean, QTc Max, QTc > 450 and Longest R-R variables ($p < 0.0195$).

The percentage of females in the HCM group was slightly lower compared to the two FD groups ($p < 0.0366$). It is also observable that FD patients without WMLs were slightly younger ($p = 0.0010$). In comparison to HCM patients, both FD patients, with and without WMLs, presented higher HR Min and HR Mean ($p < 0.0419$); and lower QT Min, QT Mean, QT Max, and Longest R-R ($p < 0.0130$). FD patients

without WMLs presented higher HR Max ($p = 0.0010$); and lower QT Min, QT Mean, QT Max, QTc Mean, and QTc >450 ($p < 0.0225$) compared to both HCM patients and FD patients with WMLs. Another group difference was found between HCM and FD without WMLs groups, relative to QTc Max Holter variable, which was higher in HCM patients ($p = 0.0051$).

Table 6: Variables summary and comparison of means/medians between HCM, FD with WMLs and FD without WMLs groups

Variable	HCM (n=39)	FD w/ WMLs (n=61)	FD w/o WMLs (n=53)	P-Value -All groups	HCM vs FD w/ WMLs	HCM vs FD w/o WMLs	FD w/ WMLs vs FD w/o WMLs
Sex (Female number (%))	13 (33.33)	38 (62.30)	32 (60.38)	p = 0.0097^a	p = 0.0221^e	p = 0.0366^e	p = 1.0000 ^e
Age	59.00 ± 14.68	56.77 ± 15.77	40.74 ± 13.04	p < 0.0001^b	p = 0.7323 ^f	p = 0.0010^f	p = 0.0010^f
HR Min	46.08 ± 7.96	49.90 ± 5.78	49.51 ± 6.78	p = 0.0152^b	p = 0.0157^f	p = 0.0419^f	p = 0.9000 ^f
HR Mean	67.67 ± 10.85	74.41 ± 7.08	76.21 ± 8.74	p = 0.0005^c	p = 0.0018^g	p = 0.0010^g	p = 0.4580 ^g
HR Max	114.31 ± 24.18	121.98 ± 12.32	134.25 ± 18.98	p < 0.0001^c	p = 0.1600 ^g	p = 0.0010^g	p = 0.0010^g
ASDNN 5	55.20 (44.00 - 65.80)	54.10 (44.60 - 70.40)	59.10 (48.60 - 73.80)	p = 0.5082 ^d	-	-	-
SDANN 5	116.52 ± 33.68	121.99 ± 30.71	121.10 (105.40 - 132.00)	p = 0.6955 ^d	-	-	-
SDNN	135.60 (106.60 - 170.50)	139.81 ± 33.00	136.70 (118.60 - 154.50)	p = 0.9837 ^d	-	-	-
RMSSD	42.10 (32.45 - 66.85)	39.10 (29.70 - 62.20)	37.20 (26.70 - 51.90)	p = 0.4334 ^d	-	-	-
QT Min	325.67 ± 34.83	307.64 ± 26.75	292.70 ± 28.74	p < 0.0001^b	p = 0.0087^f	p = 0.0010^f	p = 0.0203^f
QT Mean	424.00 (399.50 - 458.00)	402.36 ± 31.96	380.00 (367.00 - 394.00)	p < 0.0001^d	p = 0.0005^h	p < 0.0001^h	p = 0.0043^h
QT Max	525.00 (482.50 - 614.00)	481.00 (437.00 - 508.00)	449.00 (426.00 - 479.00)	p < 0.0001^d	p = 0.0003^h	p < 0.0001^h	p = 0.0225^h
QTc Min	369.62 ± 36.12	381.00 (354.00 - 396.00)	371.02 ± 30.81	p = 0.6241 ^d	-	-	-
QTc Mean	451.82 ± 31.80	434.00 (423.00 - 460.00)	427.00 (416.00 - 441.00)	p = 0.0003^d	p = 0.1161 ^h	p = 0.0001^h	p = 0.0076^h
QTc Max	569.00 (527.00 - 658.50)	545.00 (511.00 - 587.00)	528.00 (497.00 - 571.00)	p = 0.0195^d	p = 0.0798 ^h	p = 0.0051^h	p = 0.2161 ^h
QTc > 450	39.00 (6.50 - 81.50)	16.00 (4.00 - 63.00)	6.00 (2.00 - 18.00)	p = 0.0005^d	p = 0.0963 ^h	p = 0.0001^h	p = 0.0122^h
Longest R-R	1.80 (1.50 - 2.15)	1.50 (1.40 - 1.80)	1.50 (1.40 - 1.80)	p = 0.0170^d	p = 0.0092^h	p = 0.0130^h	p = 0.9560 ^h

If data have a normal distribution, they are represented as mean ± standard deviation. In case of a non-normal distribution they are represented as median (quartile1 — quartile3). In every comparison of means test α was considered 0.05.

^a Chi-Squared Test

^b One-Way ANOVA Test.

^c Welch - ANOVA Test.

^d Kruskal-Wallis H Test.

^e Fisher Exact T-Test

^f Tukey HSD Test as Post hoc.

^g Games-Howell Test as Post hoc.

^h Dunn's Test as Post hoc.

Part of the binary logistic regression results to predict group membership (HCM vs FD without WMLs, HCM vs FD with WMLs, and FD without WMLs vs FD with WMLs) are presented in Tables 7, 8 and 9, respectively. These tables only show the logistic regression results in which the Holter variables was statistically significant at the significance level of 0.05. The remaining results can be consulted in Appendix A, Tables 23, 24 and 25, respectively.

In Table 7, the two possible outcomes were defined as: having HCM ($Y = 0$) or having FD without WMLs ($Y = 1$). In this situation, most of the Holter variables (10 out of 15) were significantly associated with group membership ($p < 0.0330$). However, when the model was adjusted to age, that is, when age was included as an explanatory variable, age was statistically significant in all models ($p < 0.0013$) and, in some cases, the Holter variable, including HR Max, QTc Max and Longest R-R, was no longer significant. When adjusted to sex, this variable was not significant and the Holter variable remained significantly associated with group classification in the models with HR Mean, QT Min or QT Mean as the Holter variable ($p < 0.0010$). This suggests that a higher Heart Rate (HR Min, HR Mean and HR Max values) may lead to a higher odds of having FD without WMLs rather than HCM ($OR \geq 1.05$, $1.01 \leq 95\%CI \leq 1.15$), and, contrarily, a higher QT analysis (QT Min, QT Mean, QT Max, QTc Mean, QTc > 450 values) or higher Longest R-R could mean lower odds of having FD without WMLs in comparison to HCM ($OR \leq 0.99$, $0.16 \leq 95\%CI \leq 1.00$).

The logistic regression results for when the dependent variable has the outcomes: having HCM ($Y = 0$) or having FD with WMLs ($Y = 1$), are presented in Table 8. From the univariate logistic regression analysis, 7 Holter variables were significantly associated with group classification, namely HR Min, HR Mean, HR Max, QT Min, QT Mean, QT Max and QTc Max ($p < 0.0441$). When adjusted to age, we can see that age was not statistically significant in all models, and only HR Max variable was no longer significant ($p = 0.0591$). On the other hand, when the model was adjusted to sex, this variable was statistically significant in all models ($p < 0.0448$) and the Holter variables remained significant ($p < 0.0392$). Therefore, the increase of heart rate Holter variables (HR Min, HR Mean and HR Max) leads to increased odds of having FD with WMLs rather than HCM ($OR > 1.02$, $1.00 \leq 95\% CI \leq 1.16$). On the other hand, an increase of some of the QT analysis variables (QT Min, QT Mean, QT Max and QTc Max) leads to decreased odds of having FD with WMLs ($OR < 0.99$, $0.96 \leq 95\% CI \leq 1.00$).

In Table 9 are the logistic regression results for the outcomes: having FD without WMLs ($Y = 0$) or having FD with WMLs ($Y = 1$). Only 5 of the 15 Holter variables were significantly associated with group

classification, which were HR Min, QT Min, QT Mean, QTc Mean and, QTc > 450 ($p < 0.0071$). In the multivariable logistic regression model adjusted to age, it is noticeable that the age variable was statistically significant in all models and that all the Holter variables were no longer statistically associated with group membership. Contrarily, when the model was adjusted to sex, this variable was not statistically significant in all models and all the Holter variables remained statistically significant ($p < 0.0071$). In this case, higher QT analysis values (QT Min, QT Mean, QTc Mean and QTc > 450) may lead to higher odds of having WMLs within FD patients ($OR = 1.02$, $1.01 = 95\% CI \leq 1.04$), and, higher HR Max may lead to lower odds of having WMLs ($OR = 0.95$, $95\% CI = 0.93 - 0.98$).

Table 7: Statistically significant logistic regression results of HCM and FD without WMLs groups

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to Age		Multivariable Logistic Regression - Adjusted to Sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	1.07 (1.01 - 1.13)	0.0330	1.09 (1.01 - 1.17)	0.0205	1.05 (0.99 - 1.12)	0.1303
Age	-	-	0.91 (0.87 - 0.95)	< 0.0001	-	-
Sex	-	-	-	-	2.52 (1.02 - 6.20)	0.0448
HR Mean	1.09 (1.04 - 1.15)	0.0003	1.07 (1.02 - 1.13)	0.0104	1.09 (1.03 - 1.14)	0.0010
Age	-	-	0.92 (0.89 - 0.96)	< 0.0001	-	-
Sex	-	-	-	-	2.37 (0.93 - 6.04)	0.0693
HR Max	1.05 (1.02 - 1.07)	0.0002	1.02 (0.99 - 1.05)	0.1292	1.04 (1.02 - 1.07)	0.0004
Age	-	-	0.92 (0.88 - 0.96)	0.0003	-	-
Sex	-	-	-	-	2.63 (1.03 - 6.76)	0.0441
QT Min	0.97 (0.95 - 0.98)	0.0001	0.97 (0.96 - 0.99)	0.0079	0.97 (0.95 - 0.99)	0.0005
Age	-	-	0.92 (0.89 - 0.96)	0.0002	-	-
Sex	-	-	-	-	1.65 (0.61 - 4.44)	0.3257
QT Mean	0.96 (0.94 - 0.98)	< 0.0001	0.97 (0.95 - 0.99)	0.0008	0.96 (0.94 - 0.98)	< 0.0001
Age	-	-	0.93 (0.89 - 0.97)	0.0013	-	-
Sex	-	-	-	-	2.80 (0.96 - 8.19)	0.0604
QT Max	0.99 (0.98 - 0.99)	0.0003	0.99 (0.98 - 1.00)	0.0118	0.99 (0.98 - 0.99)	0.0002
Age	-	-	0.93 (0.89 - 0.97)	0.0004	-	-
Sex	-	-	-	-	4.03 (1.42 - 11.42)	0.0089
QTc Mean	0.97 (0.95 - 0.99)	0.0005	0.98 (0.96 - 0.99)	0.0123	0.97 (0.95 - 0.98)	0.0003
Age	-	-	0.91 (0.88 - 0.96)	0.0001	-	-
Sex	-	-	-	-	3.88 (1.46 - 10.30)	0.0064
QTc Max	0.99 (0.99 - 1.00)	0.0082	1.00 (0.99 - 1.00)	0.2512	0.99 (0.98 - 1.00)	0.0022
Age	-	-	0.92 (0.88 - 0.95)	< 0.0001	-	-
Sex	-	-	-	-	5.42 (1.91 - 15.38)	0.0015
QTc > 450	0.97 (0.96 - 0.99)	0.0002	0.98 (0.96 - 1.00)	0.0105	0.97 (0.96 - 0.99)	0.0003
Age	-	-	0.92 (0.88 - 0.96)	0.0001	-	-
Sex	-	-	-	-	3.16 (1.21 - 8.21)	0.0184
Longest R-R	0.37 (0.16 - 0.86)	0.0212	0.45 (0.17 - 1.21)	0.1125	0.30 (0.13 - 0.72)	0.0071
Age	-	-	0.91 (0.88 - 0.95)	< 0.0001	-	-
Sex	-	-	-	-	4.25 (1.64 - 11.03)	0.0029

Table 8: Statistically significant logistic regression results of HCM and FD with WMLs groups

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	1.09 (1.02 - 1.16)	0.0097	1.09 (1.02 - 1.17)	0.0087	1.08 (1.01 - 1.16)	0.0222
Age	-	-	0.99 (0.96 - 1.02)	0.3878	-	-
Sex	-	-	-	-	2.97 (1.24 - 7.09)	0.0142
HR Mean	1.09 (1.04 - 1.15)	0.0009	1.09 (1.04 - 1.15)	0.0011	1.09 (1.03 - 1.15)	0.0013
Age	-	-	1.00 (0.97 - 1.03)	0.9738	-	-
Sex	-	-	-	-	3.32 (1.35 - 8.20)	0.0092
HR Max	1.02 (1.00 - 1.05)	0.0441	1.03 (1.00 - 1.05)	0.0591	1.03 (1.00 - 1.05)	0.0355
Age	-	-	1.00 (0.97 - 1.03)	0.9148	-	-
Sex	-	-	-	-	3.53 (1.48 - 8.42)	0.0045
QT Min	0.98 (0.97 - 0.99)	0.0068	0.98 (0.97 - 1.00)	0.0086	0.98 (0.97 - 1.00)	0.0392
Age	-	-	1.00 (0.97 - 1.03)	0.8283	-	-
Sex	-	-	-	-	2.57 (1.06 - 6.23)	0.0363
QT Mean	0.98 (0.96 - 0.99)	0.0004	0.98 (0.96 - 0.99)	0.0005	0.98 (0.96 - 0.99)	0.0004
Age	-	-	1.01 (0.98 - 1.04)	0.6271	-	-
Sex	-	-	-	-	3.69 (1.45 - 9.38)	0.0061
QT Max	0.99 (0.99 - 1.00)	0.0013	0.99 (0.99 - 1.00)	0.0015	0.99 (0.98 - 1.00)	0.0006
Age	-	-	1.00 (0.98 - 1.04)	0.7507	-	-
Sex	-	-	-	-	4.81 (1.79 - 12.95)	0.0019
QTc Max	0.99 (0.99 - 1.00)	0.0270	0.99 (0.99 - 1.00)	0.0340	0.99 (0.99 - 1.00)	0.0058
Age	-	-	1.00 (0.97 - 1.03)	0.8252	-	-
Sex	-	-	-	-	5.02 (1.90 - 13.26)	0.0012

Table 9: Statistically significant logistic regression Results of FD without WMLs and FD with WMLs groups

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Max	0.95 (0.93 - 0.98)	0.0003	0.98 (0.95 - 1.01)	0.2405	0.95 (0.92 - 0.97)	0.0002
Age	-	-	1.06 (1.03 - 1.10)	0.0004	-	-
Sex	-	-	-	-	1.69 (0.72 - 3.96)	0.2251
QT Min	1.02 (1.01 - 1.03)	0.0068	1.02 (1.00 - 1.03)	0.0683	1.03 (1.01 - 1.04)	0.0025
Age	-	-	1.07 (1.04 - 1.10)	< 0.0001	-	-
Sex	-	-	-	-	1.99 (0.82 - 4.85)	0.1283
QT Mean	1.02 (1.01 - 1.03)	0.0052	1.01 (0.99 - 1.02)	0.4081	1.02 (1.01 - 1.04)	0.0038
Age	-	-	1.07 (1.04 - 1.10)	< 0.0001	-	-
Sex	-	-	-	-	1.46 (0.64 - 3.31)	0.3643
QTc Mean	1.02 (1.01 - 1.04)	0.0071	1.01 (0.99 - 1.03)	0.3947	1.02 (1.01 - 1.04)	0.0071
Age	-	-	1.07 (1.04 - 1.10)	< 0.0001	-	-
Sex	-	-	-	-	1.11 (0.51 - 2.44)	0.7916
QTc > 450	1.02 (1.01 - 1.03)	0.0065	1.00 (0.99 - 1.02)	0.7636	1.02 (1.01 - 1.03)	0.0054
Age	-	-	1.07 (1.04 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.33 (0.60 - 2.97)	0.4869

7.2 Results from patients aged 40 to 59 years (inclusive)

The descriptive results of each Holter variable per group (HCM, FD with WMLs and FD Without WMLs) and comparative results, for this age class, are presented in Table 10. Once again, data with a normal distribution is presented as mean $\pm$ standard deviation and data with a non-normal distribution is presented as median (quartile 1 - quartile 3). The normality and the homogeneity of variance test results can be consulted in Table 26 of Appendix A.

We can see that, in contrast to the results obtained from the patients over 18 years old, for this subgroup of patients, sex and age variables were not significantly different between groups ($p > 0.0601$). However, group differences were found in some of the Holter variables, such as, HR Min, HR Mean, SDANN 5, SDNN, QT Mean, QT Max and, QTc > 450 ($p < 0.0414$).

In the patients aged 40 to 59 years (inclusive), HCM patients presented higher QT Mean and QT Max compared with FD patients, both with and without WMLs ($p < 0.0048$) and also a higher QTc > 450 compared to FD with WMLs patients ($p = 0.0131$). The FD without WMLs patients presented a lower SDNN compared with HCM and FD with WMLs patients ($p < 0.0171$). The SDANN 5 variable means in the FD with WMLs and FD without WMLs groups were different ($p = 0.0117$), with a lower value in the FD patients without WMLs. Two differences between the HCM and FD without WMLs groups were found, in the HR Min and HR Max variables ($p < 0.0059$), which were both higher in the FD without WMLs patients.

Table 10: Variables summary and comparison of means/medians between HCM, FD with WMLs and FD without WMLs patients aged 40 to 59 (inclusive)

Variable	HCM (n=16)	FD w/ WMLs (n=23)	FD w/o WMLs (n=25)	P-Value -All groups	HCM vs FD w/ WMLs	HCM vs FD w/o WMLs	FD w/ WMLs vs FD w/o WMLs
Sex (Female number (%))	6 (37.50)	13 (56.52)	17 (68.00)	p = 0.1581 ^a	-	-	-
Age	54.50 (49.50 - 58.00)	50.61 ± 5.17	49.00 ± 5.67	p = 0.0601 ^d	-	-	-
HR Min	45.75 ± 7.57	49.22 ± 5.94	52.24 ± 5.77	p = 0.0083 ^b	p = 0.2188 ^e	p = 0.0059 ^e	p = 0.2306 ^e
HR Mean	68.25 ± 9.85	74.43 ± 6.27	77.48 ± 10.10	p = 0.0072 ^b	p = 0.0885 ^e	p = 0.0051 ^e	p = 0.4641 ^e
HR Max	118.00 ± 20.79	125.65 ± 10.06	128.84 ± 16.92	p = 0.1114 ^b	-	-	-
ASDNN 5	67.09 ± 30.22	52.30 (48.05 - 65.10)	50.60 ± 12.15	p = 0.0889 ^d	-	-	-
SDANN 5	124.36 ± 39.24	133.18 ± 33.67	106.11 ± 23.05	p = 0.0141 ^b	p = 0.6546 ^e	p = 0.1765 ^e	p = 0.0117 ^e
SDNN	153.49 ± 43.43	145.43 ± 34.90	121.30 (111.10 - 137.20)	p = 0.0129 ^d	p = 0.6617 ^f	p = 0.0094 ^f	p = 0.0171 ^f
RMSSD	39.80 (34.00 - 96.58)	33.90 (28.20 - 40.25)	31.81 ± 12.08	p = 0.0694 ^d	-	-	-
QT Min	313.44 ± 23.35	306.39 ± 21.52	295.76 ± 28.50	p = 0.0809 ^b	-	-	-
QT Mean	412.00 (400.25 - 443.00)	392.61 ± 26.28	385.00 (367.00 - 408.00)	p = 0.0019 ^d	p = 0.0048 ^f	p = 0.0007 ^f	p = 0.5520 ^f
QT Max	525.50 (491.75 - 633.25)	462.13 ± 40.12	455.00 (424.00 - 489.00)	p = 0.0016 ^d	p = 0.0011 ^f	p = 0.0017 ^f	p = 0.8424 ^f
QTc Min	365.88 ± 27.51	382.00 (376.00 - 399.50)	373.24 ± 37.84	p = 0.1874 ^d	-	-	-
QTc Mean	447.25 ± 23.31	431.74 ± 17.35	430.00 (425.00 - 444.00)	p = 0.0823 ^d	-	-	-
QTc Max	602.25 ± 96.65	533.48 ± 54.43	552.92 ± 60.02	p = 0.0472 ^c	-	-	-
QTc > 450	29.50 (6.75 - 70.50)	7.00 (2.50 - 23.50)	9.00 (2.00 - 36.00)	p = 0.0414 ^d	p = 0.0131 ^f	p = 0.0633 ^f	p = 0.4607 ^f
Longest R-R	1.80 (1.40 - 2.33)	1.50 (1.35 - 1.75)	1.50 (1.30 - 1.60)	p = 0.1504 ^d	-	-	-

If data have a normal distribution, they are represented as mean ± standard deviation. In case of a non-normal distribution they are represented as median (quartile1 — quartile3). In every comparison of means test α was considered 0.05.

^a Chi-Squared Test

^b One-Way ANOVA Test.

^c Welch - ANOVA Test.

^d Kruskal-Wallis H Test.

^e Tukey HSD Test as Post hoc.

^f Dunn's Test as Post hoc.

Similar to what was presented for patients aged above 18 years old, in Tables 11, 12 and, 13 are the binary logistic results that were significantly associated with group classification (HCM vs FD without WMLs, HCM vs FD with WMLs, and FD without WMLs vs FD with WMLs) of the patients aged 40 to 59 years. The results that were not significantly associated are presented in Tables 27, 28 and 29 of Appendix A.

Table 11 is relative to the following two outcomes: having HCM ($Y = 0$) or having FD without WMLs ($Y = 1$) logistic regression results. We can see that 6 out of 15 Holter variables were significantly associated with group membership, which were, HR Min, HR Mean, ASDNN 5, SDNN, QT Mean and QT Max ($p < 0.0425$). When the model was adjusted to age, this variable was statistically significant only in the model with SDNN as Holter variable, and, in turn, this variable remained significantly associated with group membership ($p = 0.0134$). Only two other Holter variables remained significantly associated with group membership when adjusted to age, HR Min and QT Mean variables ($p < 0.0297$). When adjusted to sex, this variable was not significant in all models and the 6 Holter variables remained significantly associated with group membership ($p < 0.0389$). From these results we can see that a higher HR Min and HR Mean lead to higher odds of having FD without WMLs rather than HCM ($OR \geq 1.10$, $1.02 \leq 95\% CI \leq 1.30$), on the other hand, an increase in ASDNN 5, SDNN, QT Mean and QT Max lead to decreased odds of having FD without WMLs ($OR \leq 0.99$, $0.92 \leq 95\% CI \leq 1.00$).

Six variables, HR Mean, QT Mean, QT Max, QTc Mean, QTc Max and QTc > 450, were significantly associated with group membership ($p < 0.0331$) when the two possible outcomes were: having HCM ($Y = 0$) or having FD with WMLs ($Y = 1$), as can be seen in Table 12. In the models adjusted to age and to sex, both variables were not significant in all models and only one Holter variable was no longer significant, which was HR Mean variable when adjusted to age ($p = 0.0602$). Only the increase of the HR Mean variable leads to increased odds of having FD with WMLs rather than HCM ($OR = 1.11$, $95\% CI = 1.01 - 1.21$), when the remaining 5 Holter variables of QT analysis (QT Mean, QT Max, QTc Mean, QTc Max and QTc > 450) increase the odds of having FD with WMLs decrease ($OR \leq 0.99$, $0.93 \leq 95\% CI \leq 1.00$).

The results for when the dependent variable has the outcomes: having FD without WMLs ($Y = 0$) or having FD with WMLs ($Y = 1$), are presented in Table 13. In this case, only 2 out of 15 Holter variables were significantly associated with group membership, SDANN 5 and SDNN ($p < 0.0130$), and these variables remained significant when the model was adjusted both to age and sex (which were not significant in all

models). A higher heart rate variability (SDNN and SDANN 5 values) lead to higher odds of having WMLs within FD patients (OR = 1.03, 95% CI = 1.01 - 1.06 for SDNN and OR = 1.04, 95% CI = 1.01 – 1.07 for SDANN5).

Table 11: Statistically significant logistic regression results of HCM and FD without WMLs patients aged 40 to 59 (inclusive)

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	1.16 (1.04 - 1.30)	0.0088	1.14 (1.02 - 1.27)	0.0224	1.17 (1.04 - 1.31)	0.0113
Age	-	-	0.90 (0.79 - 1.03)	0.1210	-	-
Sex	-	-	-	-	3.57 (0.81 - 15.68)	0.0914
HR Mean	1.10 (1.02 - 1.18)	0.0126	1.08 (1.00 - 1.16)	0.0588	1.09 (1.01 - 1.17)	0.0191
Age	-	-	0.92 (0.80 - 1.05)	0.2314	-	-
Sex	-	-	-	-	3.11 (0.74 - 13.04)	0.1202
ASDNN 5	0.96 (0.92 - 1.00)	0.0425	0.96 (0.92 - 1.00)	0.0565	0.95 (0.91 - 1.00)	0.0389
Age	-	-	0.88 (0.77 - 1.00)	0.0498	-	-
Sex	-	-	-	-	4.49 (1.02 - 19.68)	0.0464
SDNN	0.97 (0.94 - 0.99)	0.0134	0.96 (0.94 - 0.99)	0.0134	0.96 (0.94 - 0.99)	0.0176
Age	-	-	0.86 (0.74 - 0.99)	0.0373	-	-
Sex	-	-	-	-	3.87 (0.84 - 17.71)	0.0815
QT Mean	0.97 (0.94 - 0.99)	0.0116	0.97 (0.94 - 1.00)	0.0297	0.97 (0.94 - 0.99)	0.0137
Age	-	-	0.92 (0.80 - 1.05)	0.2268	-	-
Sex	-	-	-	-	3.86 (0.86 - 17.40)	0.0784
QT Max	0.99 (0.98 - 1.00)	0.0162	0.99 (0.98 - 1.00)	0.0519	0.99 (0.98 - 1.00)	0.0182
Age	-	-	0.91 (0.80 - 1.04)	0.1693	-	-
Sex	-	-	-	-	4.10 (0.94 - 17.95)	0.0613

Table 12: Statistically significant logistic regression results of HCM and FD with WMLs patients aged 40 to 59 (inclusive)

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Mean	1.11 (1.01 - 1.21)	0.0331	1.10 (1.00 - 1.21)	0.0602	1.11 (1.01 - 1.22)	0.0304
Age	-	-	0.94 (0.82 - 1.08)	0.4036	-	-
Sex	-	-	-	-	2.50 (0.59 - 10.39)	0.2173
QT Mean	0.96 (0.94 - 0.99)	0.0125	0.96 (0.94 - 0.99)	0.0171	0.96 (0.93 - 0.99)	0.0112
Age	-	-	0.94 (0.81 - 1.08)	0.3702	-	-
Sex	-	-	-	-	3.16 (0.66 - 15.04)	0.1488
QT Max	0.97 (0.95 - 0.99)	0.0085	0.97 (0.95 - 0.99)	0.0093	0.97 (0.95 - 0.99)	0.0069
Age	-	-	0.94 (0.81 - 1.10)	0.4611	-	-
Sex	-	-	-	-	4.64 (0.73 - 29.62)	0.1045
QTc Mean	0.96 (0.93 - 1.00)	0.0329	0.96 (0.92 - 1.00)	0.0403	0.96 (0.93 - 1.00)	0.0300
Age	-	-	0.91 (0.79 - 1.05)	0.1906	-	-
Sex	-	-	-	-	2.47 (0.60 - 10.22)	0.2114
QTc Max	0.99 (0.98 - 1.00)	0.0173	0.99 (0.98 - 1.00)	0.0292	0.98 (0.97 - 1.00)	0.0101
Age	-	-	0.95 (0.82 - 1.09)	0.4466	-	-
Sex	-	-	-	-	5.28 (0.93 - 30.10)	0.0608
QTc > 450	0.97 (0.95 - 1.00)	0.0252	0.97 (0.95 - 1.00)	0.0312	0.97 (0.95 - 1.00)	0.0297
Age	-	-	0.91 (0.79 - 1.05)	0.2014	-	-
Sex	-	-	-	-	2.03 (0.50 - 8.27)	0.3241

Table 13: Statistically significant logistic regression results of FD without WMLs and FD with WMLs patients aged 40 to 59 (inclusive)

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
SDANN 5	1.04 (1.01 - 1.07)	0.0080	1.05 (1.01 - 1.08)	0.0046	1.04 (1.01 - 1.07)	0.0064
Age	-	-	1.13 (0.99 - 1.28)	0.0733	-	-
Sex	-	-	-	-	0.42 (0.11 - 1.66)	0.2169
SDNN	1.03 (1.01 - 1.06)	0.0130	1.04 (1.01 - 1.07)	0.0077	1.03 (1.01 - 1.06)	0.0119
Age	-	-	1.11 (0.98 - 1.26)	0.0867	-	-
Sex	-	-	-	-	0.52 (0.14 - 1.91)	0.3235

7.3 Results from patients over 59 years old

The variables summary and the comparison of means/medians between the two groups, for patients with age > 59 , are presented in Table 14. The normality and homogeneity of variances test results can be consulted in Table 30 of Appendix A.

Group differences were found in the sex variable, HR Min, HR Mean and HR Max Holter variables ($p < 0.0421$) which were higher in the FD patients with WMLs. In contrast, QT Min and QT Mean Holter variables ($p < 0.0182$) were lower in these patients.

The logistic regression analysis results to predict group membership are displayed in Table 15 for results that were significantly associated and in Table 31 of Appendix A for results that were not associated. In Table 15 we can see that 4 Holter variables were significantly associated with group classification, HR Min, HR Mean, QT Min and QT Mean ($p < 0.0444$). When the model was adjusted to age, this variable was not significantly associated with group membership and all 4 Holter variables remained associated ($p < 0.0402$). However, when adjusted to sex, only two Holter variables remained significantly associated, which were HR Mean and QT Min ($p < 0.0422$), and the sex variable was not associated. Therefore, the higher the heart rate variables (HR Min and HR Mean values) are, the higher are the odds of having FD with WMLs instead of HCM (OR = 1.10, 95% CI = 1.00 - 1.20 for HR Min and OR = 1.09, 95% CI = 1.02 - 1.17 for HR Mean). The increase of either of the two QT analysis variables (QT Min and QT Mean), in contrast, leads to increased odds of having HCM rather than FD with WMLs (OR = 0.98, 95% CI = 0.96 - 0.99 for QT Min and OR = 0.98, 95% CI = 0.96 - 1.00 for QT Mean).

Table 14: Variables summary and comparison of means/medians for HCM and FD with WMLs patients aged above 59

Variable	HCM (n=19)	FD w/ WMLs (n=30)	P-Value
Sex (Female number (%))	6 (31.58)	19 (63.33)	p = 0.0421 ^a
Age	67.00 (62.00 - 77.00)	69.00 (63.25 - 75.00)	p = 0.9590 ^b
HR Min	45.84 ± 8.98	50.57 ± 6.00	p = 0.0319 ^c
HR Mean	65.84 ± 11.69	73.37 ± 7.81	p = 0.0095 ^c
HR Max	102.00 (94.50 - 111.50)	116.07 ± 11.17	p = 0.0065 ^b
ASDNN 5	48.60 (38.15 - 58.10)	48.80 (40.05 - 68.43)	p = 0.5866 ^b
SDANN 5	109.39 ± 25.41	114.76 ± 28.66	p = 0.5082 ^c
SDNN	121.70 (100.75 - 154.65)	134.05 ± 34.09	p = 0.8535 ^b
RMSSD	46.20 (28.75 - 57.20)	44.05 (32.45 - 86.75)	p = 0.7272 ^b
QT Min	340.63 ± 38.60	301.00 (285.75 - 339.75)	p = 0.0059 ^b
QT Mean	442.53 ± 40.63	417.00 ± 32.08	p = 0.0182 ^c
QT Max	533.00 (485.00 - 576.00)	503.50 (474.00 - 533.50)	p = 0.1788 ^b
QTc Min	372.05 ± 44.34	363.23 ± 43.10	p = 0.4934 ^c
QTc Mean	453.32 ± 33.14	455.10 ± 32.70	p = 0.8539 ^c
QTc Max	569.00 (527.00 - 626.50)	559.50 (545.00 - 624.00)	p = 0.9427 ^b
QTc > 450	41.00 (10.00 - 81.50)	62.50 (17.25 - 80.00)	p = 0.6965 ^b
Longest R-R	1.70 (1.65 - 2.00)	1.60 (1.43 - 1.80)	p = 0.0573 ^b

If data have a normal distribution they are represented as mean ± standard deviation. In case of a non-normal distribution they are represented as median (quartile1 – quartile3). In every comparison of means test α is considered 0.05.

^a Fisher's Exact Test

^b Mann Whitney U Test

^c Student's T-Test

Table 15: Statistically significant logistic regression results of HCM and FD with WMLs groups with ages above 59

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	1.10 (1.00 - 1.20)	0.0444	1.10 (1.00 - 1.21)	0.0402	1.08 (0.98- 1.18)	0.1068
Age	-	-	0.98 (0.91 - 1.06)	0.5904	-	-
Sex	-	-	-	-	2.94 (0.82 - 10.56)	0.0985
HR Mean	1.09 (1.02 - 1.17)	0.0163	1.09 (1.02 - 1.17)	0.0145	1.08 (1.01 - 1.16)	0.0288
Age	-	-	0.97 (0.90 - 1.06)	0.5309	-	-
Sex	-	-	-	-	3.30 (0.91 - 12.02)	0.0703
QT Min	0.98 (0.96 - 0.99)	0.0090	0.97 (0.96 - 0.99)	0.0067	0.98 (0.96 - 1.00)	0.0422
Age	-	-	0.96 (0.89 - 1.05)	0.3806	-	-
Sex	-	-	-	-	2.18 (0.56 - 8.49)	0.2589
QT Mean	0.98 (0.96 - 1.00)	0.0258	0.98 (0.96 - 1.00)	0.0205	0.98 (0.96 - 1.00)	0.0470
Age	-	-	0.97 (0.89 - 1.05)	0.3959	-	-
Sex	-	-	-	-	3.28 (0.91 - 11.80)	0.0684

7.4 Discussion

Considering all patients (aged above 18 years) significant differences in age and sex between groups were found. FD patients with WMLs are older than FD patients without WMLs and HCM patients are also older than FD without WMLs patients (Table 6). The age difference found in FD with and without WMLs groups have also been identified in the study of Braga (2021) [32]. Galluzzi et al. [33] also identified that MCI patients with WMLs were older than the ones without WMLs. In fact, WMLs are quite common in the elderly general population [43] [79].

Several ECG characteristics give a significant association with the dependent variable (group class) in

the univariate logistic regression analysis (Tables 7, 8, and 9), however, some of these associations appear to be influenced by age, since that, when adjusted to age, some of the ECG characteristics do not remain associated with the dependent variable. This happens mainly in the analysis of groups with significant age differences, that is, HCM and FD without WMLs, and FD with and without WMLs (Tables 7 and 9). This situation was also reported in the study of Braga (2021) for FD with WMLs and FD without WMLs groups. As people age, heart functions tend to decline [63] which reflects on different values on the ECG variables.

After dividing the patients into age classes, age and sex are no longer significant between groups in the class of 40 to 59 years (including) (Table 10). In the class above 59 years, where only HCM and FD with WMLs groups are taken into account, age is not significant, however, sex is significantly different between these groups (Table 14).

When analyzing the association between each ECG characteristic and HCM and FD without WMLs groups, on patients of the age class from 40 to 59 years old, several ECG characteristics (HR Min, HR Mean, ASDNN 5, SDNN, QT Mean, QT Max) are associated with the dependent variable (HCM or FD). These variables do not seem to be from a specific type, since there are 2 of HR, 2 of HRV and 2 of QT analysis.

As for logistic regression analysis results of patients aged between 40 and 59 (including) with HCM and FD with WMLs, the ECG characteristics that are significantly associated with the dependent variable (HCM or FD) are HR Mean, QT Mean, QT Max, QTc Mean, QTc Max, QTc > 450. An interesting finding is that the associated variables are mainly QT analysis variables (5 out of the 6 associated variables). Namdar et al. [31], by using a combined method that includes the QTc interval, to diagnose Fabry disease, demonstrate that this variable can have an important role in the diagnosis of Fabry disease.

When the same analysis is performed for FD with and without WMLs groups (patients aged between 40 and 59 as well), only 2 variables are significantly associated with the dependent variable (presence or absence of WMLs), SDANN 5 e SDNN variables, which are both HRV variables, suggesting that this type of variables (HRV variables) could be the most important to distinguish between patients with and without WMLs. Braga (2021) [32] also performed this analysis and obtained the same results. Galluzzi et al. identified that RMSSD variable was associated with the extent of WMLs in patients with MCI. Although the variable is different from the ones obtained in the current study, it is also an HRV variable.

Another interesting finding is the fact that Longest R-R variable is not significantly associated with the

dependent variable in any of the three logistic regression analysis performed on patients aged between 40 and 59 years. This suggests that this variable is not important to differentiate any of the diseases or manifestations that are considered in this study.

The logistic regression results of patients aged above 59 suggest that HRV are not the most important variables to distinguish between HCM and FD with WMLs groups, since the variables that are significantly associated with the dependent variable (HCM or FD) are HR Min, HR Mean (HR variables), QT Min, and QT Mean (QT analysis variables).

Chapter 8

Machine Learning Results

This chapter presents the machine learning results for patients aged 40 to 59 years (inclusive) and patients aged above 60, that were obtained from the application of the different machine learning techniques outlined in Figure 1 and described in Chapter 6.

8.1 Results from patients aged 40 to 59 years (inclusive)

As explained, a feature selection procedure based on VIF values was implemented to eliminate features that presented high correlation with other features (multicollinearity). For the group of patients aged between 40 and 59 years old (both ages included), 6 features were eliminated, thus, we ended up with 9 features, which were: HR Mean, HR Max, SDANN 5, RMSSD, QT Min, QTc Min, QTc Max, QTc > 450, and Longest R-R. These variables are in bold type on Table 16, table where the VIF values obtained in each iteration of the process are displayed. In this table is also observable that each iteration has a value in red color. This value is the highest VIF value obtained in the iteration and the associated variable is the one that is eliminated.

Table 16: VIF values for patients aged 40 to 59 (inclusive)

Features	Iteration						
	1°	2°	3°	4°	5°	6°	7°
HR Min	9.18	8.78	8.78	8.41	8.34	-	-
HR Mean	33.96	32.29	6.13	5.98	5.50	4.73	4.21
HR Max	6.66	4.11	4.04	4.03	3.89	3.76	3.54
ASDNN 5	8.99	8.48	8.19	7.86	7.81	6.45	-
SDANN 5	45.00	3.43	3.42	3.41	3.41	1.87	1.59
SDNN	87.57	-	-	-	-	-	-
RMSSD	20.76	6.00	5.91	5.63	5.59	5.56	1.95
QT Min	3.31	3.26	3.13	3.12	3.11	3.11	2.99
QT Mean	52.96	52.41	-	-	-	-	-
QT Max	10.32	10.17	8.70	8.65	-	-	-
QTc Min	3.99	3.73	3.73	2.99	2.93	2.90	2.86
QTc Mean	33.45	33.00	9.41	-	-	-	-
QTc Max	7.56	7.32	7.02	6.70	2.72	2.49	2.44
QTc > 450	7.74	7.40	7.02	2.03	1.90	1.89	1.89
Longest R-R	3.67	2.44	2.44	2.43	2.33	2.21	2.21

The hyperparameters values/status that resulted from the hyperparameter tuning performed for this set of variables of the HCM and FD without WMLs patients aged between 40 and 59 years can be consulted in Table 32 of Appendix B. To note that the hyperparameters that were not evaluated during the hyperparameter tuning process took the default values/status. The accuracy values obtained by each of these models, to predict if a patient has FD, can also be consulted, this time in Table 33 of the same Appendix.

The application of RFE method using logistic regression as classifier led to the highest accuracy with 6 features, that were HR Mean, SDANN 5, RMSSD, QTMin, QTc Max, and QTc > 450, being this accuracy equal to 0.7805, as can be seen in Figure 6a. On the other hand, with linear SVM model, which results

are displayed in Figure 6b, the highest accuracy, 0.8049, was detected with only 4 features, which were HR Mean, SDANN 5, RMSSD, and QTc > 450. These features are a subgroup of the 6 features selected by logistic regression. With random forest model as classifier, all 9 features was the combination that gave a better performance, with an accuracy of 0.7561.

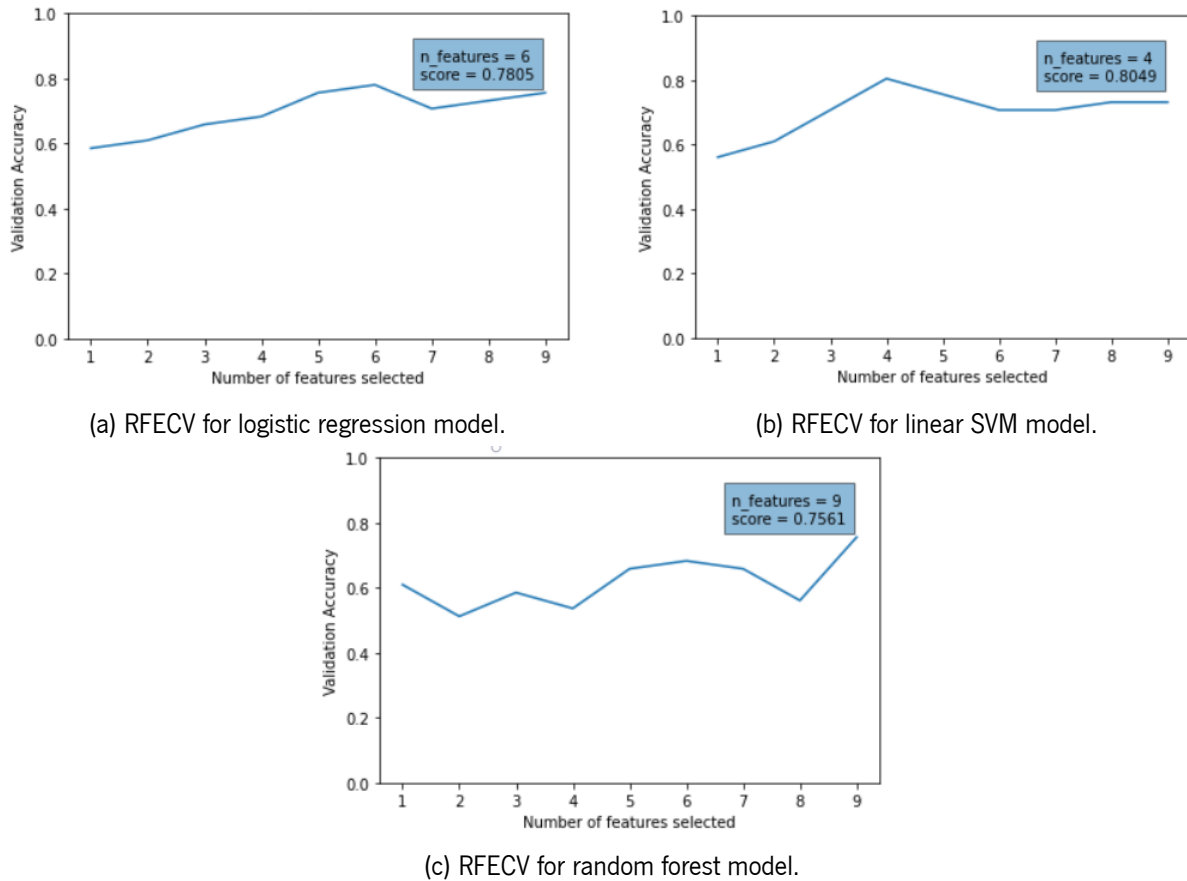


Figure 6: RFECV results for the different machine learning models of the HCM vs FD without WMLs patients aged 40 to 59 (inclusive).

The classification results, obtained by testing the different combinations of features in the 5 machine learning models, can be observed in Table 17. For each feature combination (each line of the table), the classifier that predicted with the highest accuracy (in case of equal accuracies among classifiers, the AUC is the tie-breaker metric) is in bold type. The model that better predicted if a patient had FD was the RBF SVM model with 4 features. In this case the accuracy of the model is 0.8049, which means that, in 41 predictions (number of HCM and FD without WMLs patients aged between 40 and 59), the model correctly predicted 33 (value of TN + TP), and the AUC is 0.8675. The specificity of this model is relatively low,

with a value of 0.5625, however, the sensitivity, which is equal to 0.9600, is very high. This indicates that the model is better at predicting if a patient has FD (positive class) rather than predicting if a patient has HCM (negative class). From this table it is also observable that regardless of the combination of features, RBF SVM algorithm always presents the highest accuracy. This suggests that this is the best algorithm for this classification problem. Also, the 4-feature combination led to the best results, being this the best combination to predict if a patient has FD or HCM.

Table 34 of Appendix B shows the hyperparameters values/status of these machine learning models.

Table 17: Models's classification performance metrics per number of features for HCM and FD without WMLs patients aged 40 to 59 (inclusive)

Number of features (Features)	ML Model	Metrics			
		Accuracy	Sensitivity	Specificity	AUC
4 features (HR Mean, SDANN 5, RMSSD, and QTc > 450)	LR	0.7805	0.8800	0.6250	0.8150
	Linear SVM	0.8049	0.9600	0.5625	0.8650
	RBF SVM	0.8049	0.9600	0.5625	0.8675
	RF	0.7805	0.8400	0.6875	0.7825
	KNN	0.8049	0.9600	0.5625	0.8175
6 features (HR Mean, SDANN 5, RMSSD, QT Min, QTc Max, and QTc > 450)	LR	0.7805	0.8400	0.6875	0.7850
	Linear SVM	0.7805	0.8800	0.6250	0.8250
	RBF SVM	0.8049	0.8400	0.7500	0.7950
	RF	0.7561	0.8000	0.6875	0.8150
	KNN	0.7561	0.9200	0.5000	0.7612
9 features (All)	LR	0.7561	0.8400	0.6250	0.7625
	Linear SVM	0.7317	0.8400	0.5625	0.7800
	RBF SVM	0.7805	0.8400	0.6875	0.7650
	RF	0.7561	0.6800	0.8750	0.7575
	KNN	0.7561	0.8400	0.6250	0.7700

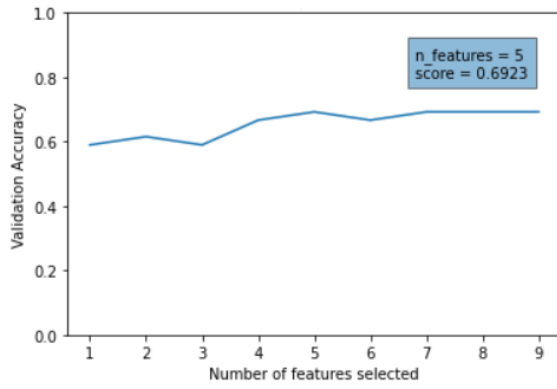
ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

In relation to HCM patients and FD with WMLs patients groups, the variables selected from the feature

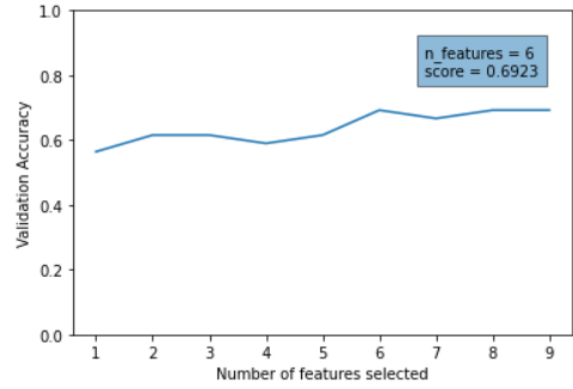
selection method based in the VIF values is the same, since this was applied to all the patients aged between 40 and 59 years old.

As for the hyperparameters values/status obtained in the hyperparameter tuning, these are different, of course, and can be consulted in Table 32 of Appendix B. The accuracy values obtained by these models at predicting if a patient has FD are displayed in Table 33, also in Appendix B.

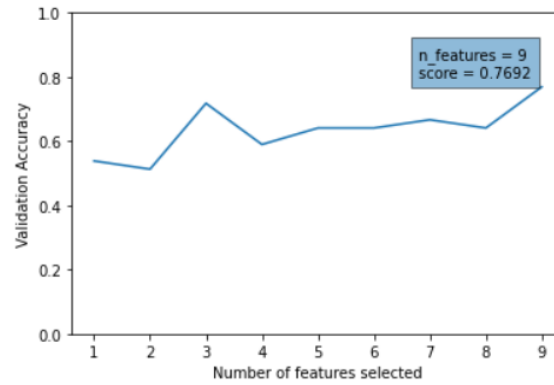
Applying the RFE method with logistic regression as classifier led to the results in Figure 7a, where the maximum accuracy value is 0.6923 for 5 features, being these features the HR Mean, RMSSD, QTc Min, QTc Max, and QTc > 450. Using linear SVM as classifier, the maximum accuracy value is also 0.6923, however, with this model, this maximum was obtained with 6 features, as shown in Figure 7b, which were HR Mean, HR Max, RMSSD, QTc Min, QTc Max, and QTc > 450. It is noticeable that these are the same 5 features selected by logistic regression plus HR Max feature. The results for when random forest algorithm is the classifier are in Figure 7c, in which the best accuracy is obtained using all features (9 features) and is equal to 0.7692.



(a) RFECV for logistic regression model.



(b) RFECV for linear SVM model.



(c) RFECV for random forest model.

Figure 7: RFECV results for the different machine learning models of the HCM vs FD with WMLs patients aged 40 to 59 (inclusive).

Considering that the 3 models that were wrapped by RFE gave three different combinations of features, these combinations were all applied to the classification stage, that is, applied to the 5 machine learning algorithms, just like what was done previously. The performance metrics obtained in each case are shown in Table 18. By analysing this table we see that the best model, in general, is using 6 features and KNN model, where an accuracy of 0.8461 and an AUC of 0.8370 are obtained. This accuracy value means that the model correctly predicted 33 patients in a total of 39 (number of HCM and FD with WMLs patients aged between 40 and 59). Also, the 6-feature combination seems to be the best combination to use to predict if a patient has FD (comparing with HCM patients), since it is the combination with the best classification performance metrics.

The hyperparameters values/status of each of these machine learning models can be consulted in Table 35 of Appendix B.

Table 18: Models's classification performance metrics per number of features for HCM and FD with WMLs patients aged 40 to 59 (inclusive)

Number of features (Features)	ML Model	Metrics			
		Accuracy	Sensitivity	Specificity	AUC
5 features (HR Mean, RMSSD, QTc Min, QTc Max, and QTc > 450)	LR	0.7179	0.9130	0.4375	0.7473
	Linear SVM	0.7179	0.8696	0.5000	0.6630
	RBF SVM	0.8462	0.9130	0.7500	0.7446
	RF	0.7179	0.8261	0.5625	0.7826
	KNN	0.7692	0.8696	0.6250	0.7962
6 features (HR Mean, HR Max, RMSSD, QTc Min, QTc Max, and QTc > 450)	LR	0.7179	0.8696	0.5000	0.7554
	Linear SVM	0.7179	0.8696	0.5000	0.7446
	RBF SVM	0.8205	0.8696	0.7500	0.8370
	RF	0.7949	0.8261	0.7500	0.8397
	KNN	0.8462	0.8696	0.8125	0.8370
9 features (All)	LR	0.6923	0.9130	0.3750	0.6902
	Linear SVM	0.6923	0.8696	0.4375	0.6739
	RBF SVM	0.7949	0.8696	0.6875	0.7663
	RF	0.7692	0.8261	0.6875	0.7663
	KNN	0.7179	0.9130	0.4375	0.7201

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

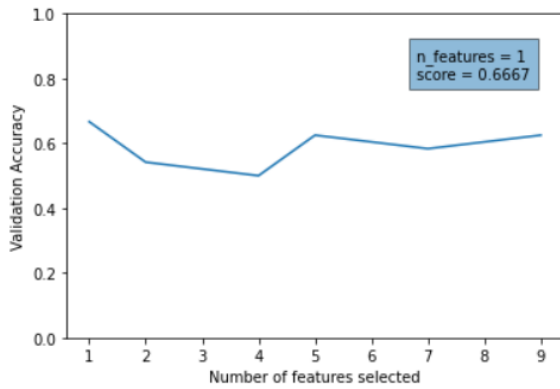
The third prediction is to predict if a patient has WMLs using FD patients, with and without WMLs.

For this prediction, since it was, once again, with patients aged between 40 and 59 (inclusive), we are dealing with 9 features (after the VIF step) which are HR Mean, HR Max, SDANN 5, RMSSD, QT Min, QTc Min, QTc Max, QTc > 450, and Longest R-R.

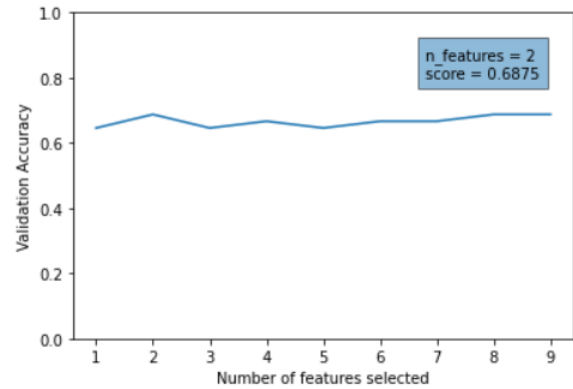
Once again, the hyperparameters values/status of each machine learning model and the accuracy values obtained by these models at predicting the presence of WMLs, can be consulted in the Appendix B, Tables 32 and 33, respectively.

These models were then wrapped by RFE method for a second feature selection. With logistic regres-

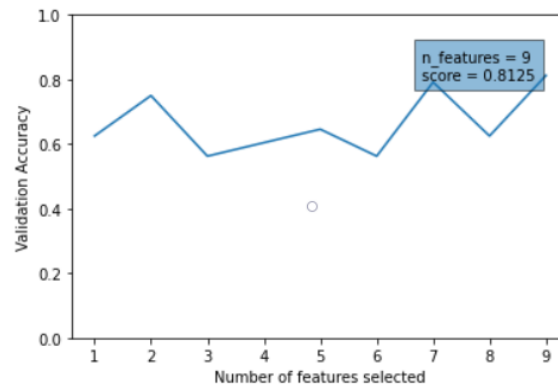
sion model, the best accuracy, 0.6667, was obtained with only one variable, as can be observed in Figure 8a, being this variable the SDANN 5 variable. With linear SVM as classifier, the RFE method identified that the best accuracy was 0.6875 and it was obtained with 2 features, as presented in Figure 8b, which were SDANN 5 and QTc Min features. However, with random forest model, the best accuracy, with a value equal to 0.8125, occurred with all the 9 variables, as shown in Figure 8c.



(a) RFECV for logistic regression model.



(b) RFECV for linear SVM model.



(c) RFECV for random forest model.

Figure 8: RFECV results for the different machine learning models of the FD with WMLs vs FD without WMLs patients aged 40 to 59 (inclusive).

These different combinations of features were then used to train the 5 machine learning algorithms, as previously explained, and the results obtained for each of these combinations for the FD patients, with and without WMLs, are presented in Table 19. We quickly realize that the best model is with all features (9 features) and with random forest as classifier, where the accuracy is equal to 0.8125. This value means that in 48 predictions (number of FD patients with and without WMLs aged between 40 and 59), the model correctly predicted 39. As for both 1 feature and 2 features situations, the RBF SVM classifier show

a better performance, with accuracy of 0.729 and 0.750, respectively.

Looking at the accuracy of all classifiers, in the 9 features case, we see that, except for random forest, all the classifiers have a weaker performance compared with the performance obtained with both 1 feature and 2 features cases. This suggests that, although the highest accuracy is in the 9-feature combination, if we are interested in selecting the best feature combination among classifiers, this may not be the best combination to predict if a FD patient has WMLs. Therefore, the 2-feature combination, which includes SDANN 5 and QTc Min variables, is the best combination, since it presents higher accuracy, sensitivity and AUC than the 1 feature combination.

The values/status of the hyperparameters, of each model, can be consulted in Table 36 of Appendix B.

Table 19: Models's classification performance metrics per number of features for FD with WMLs and FD without WMLs patients aged 40 to 59 (inclusive)

Number of features (Features)	ML Model	Metrics			
		Accuracy	Sensitivity	Specificity	AUC
1 feature (SDANN 5)	LR	0.6667	0.6087	0.7200	0.6939
	Linear SVM	0.6875	0.4783	0.8800	0.6783
	RBF SVM	0.7292	0.4783	0.9600	0.6087
	RF	0.7292	0.4783	0.9600	0.5757
	KNN	0.7292	0.4783	0.9600	0.5670
2 features (SDANN 5 and QTc Min)	LR	0.6667	0.5652	0.7600	0.6765
	Linear SVM	0.7083	0.5217	0.8800	0.6870
	RBF SVM	0.7500	0.5217	0.9600	0.6504
	RF	0.7083	0.4348	0.9600	0.5878
	KNN	0.7292	0.4348	1.0000	0.5948
9 features (All)	LR	0.6250	0.6087	0.6400	0.6000
	Linear SVM	0.6875	0.5652	0.8000	0.5696
	RBF SVM	0.6875	0.6087	0.7600	0.6122
	RF	0.8125	0.7391	0.8800	0.8348
	KNN	0.6667	0.4348	0.8800	0.6861

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

8.2 Results from patients over 59 years old

For the patients aged above 59 years old, the first feature selection procedure, based on VIF values, identified 5 features with multicollinearity. Therefore, these 5 features were excluded and 10 remained, which were HR Min, HR Max, ASDNN 5, SDANN 5, QT Min, QT Mean, QTc Min, QTc Max, QTc > 450, and Longest R-R features. The VIF values obtained in each iteration of the process are displayed in Table 20

and, in this table, the held features are in bold type and the highest value per iteration is in red color.

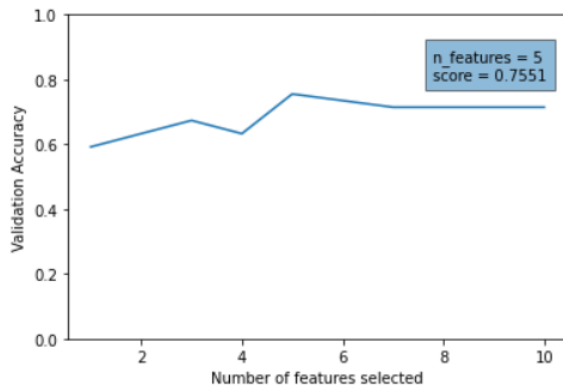
Table 20: VIF values for patients aged above 59

Features	Iteration					
	1°	2°	3°	4°	5°	6°
HR Min	4.55	4.49	3.74	3.65	3.47	3.13
HR Mean	18.94	9.33	9.28	9.08	-	-
HR Max	3.19	3.18	3.06	3.06	2.95	2.86
ASDNN 5	6.07	5.90	5.37	5.36	5.23	2.00
SDANN 5	9.90	9.48	2.11	1.98	1.95	1.64
SDNN	30.90	30.11	-	-	-	-
RMSSD	30.12	28.23	6.16	6.14	6.06	-
QT Min	5.40	4.35	4.27	4.27	3.77	3.76
QT Mean	25.07	6.97	6.84	6.46	3.00	2.95
QT Max	11.24	11.17	11.16	-	-	-
QTc Min	4.40	3.97	3.85	3.82	3.52	3.52
QTc Mean	31.69	-	-	-	-	-
QTc Max	9.01	8.32	8.32	1.79	1.73	1.72
QTc > 450	9.19	5.32	5.26	5.26	2.05	2.03
Longest R-R	1.73	1.69	1.65	1.64	1.62	1.37

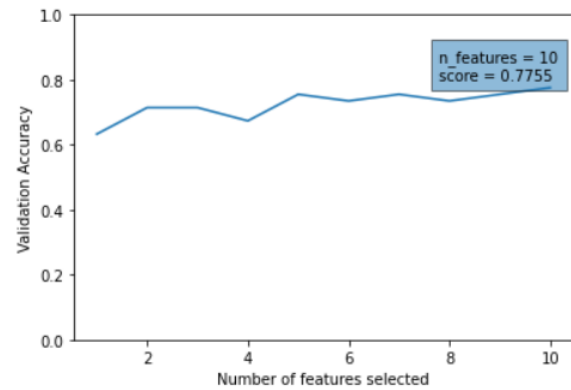
In Table 37 and Table 38 of Appendix B, are the hyperparameters values/status of each machine learning model obtained in the first hyperparameter tuning and the respective accuracy values obtained at predicting if a patient has FD, respectively.

The highest accuracy value obtained with logistic regression as classifier, which results are in Figure 9a, was 0.7551, and it occurred with 5 features: HR Min, SDANN 5, QT Min, QT Mean, and QTc > 450. In the case of linear SVM algorithm, the RFE method identified that the highest accuracy is obtained with all the 10 features, and this accuracy value is 0.7755, as shown in Figure 9b. A 9-feature combination, including HR Min, HR Max, ASDNN 5, SDANN 5, QT Min, QT Mean, QTc Min, QTc > 450, and Longest

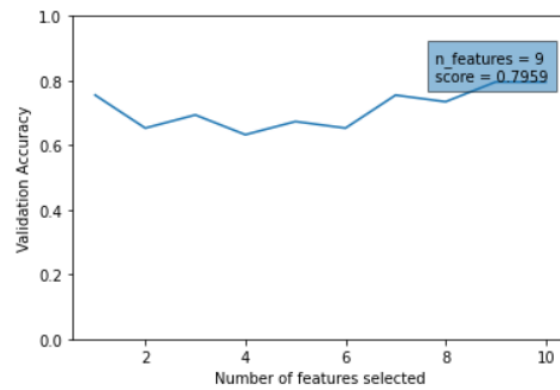
R-R features, is the combination that led to the highest accuracy, with a 0.7959 value, for random forest algorithm. The RFE results of this model can be observed in Figure 9c.



(a) RFECV for logistic regression model.



(b) RFECV for linear SVM model.



(c) RFECV for random forest model.

Figure 9: RFECV results for the different machine learning models of the HCM vs FD with WMLs patients aged above 59.

The hyperparameters values/status for each feature combination and machine learning algorithm conjugation, obtained in the hyperparameter tuning step, that was performed after the RFE, are displayed in Table 39 of Appendix B. The respective obtained classification performance metrics are presented in Table 21.

Random forest is the algorithm that shows better performance for both 9 features and 10 features. For 5 features the algorithm with the highest accuracy is the RBF SVM. However, the feature combination and algorithm conjugation that has the highest accuracy, 0.8163, is the one with 9 features (all variables except QTc Max) and random forest algorithm as classifier. Since the model predicted 49 situations (number of HCM and FD with WMLs patients aged above 9 years), the accuracy value obtained means that, in the 49

predictions, 40 were correctly predicted. Also, this 9-feature combination seems to be the best combination for this classification problem.

Table 21: Models's classification performance metrics per number of features for HCM and FD with WMLs patients aged above 59

Number of features (Features)	ML Model	Metrics			
		Accuracy	Sensitivity	Specificity	AUC
5 features (HR Min, SDANN 5, QT Min, QT Mean, and QTc > 450)	LR	0.7755	0.9000	0.5789	0.7333
	Linear SVM	0.7755	0.9000	0.5789	0.7702
	RBF SVM	0.7959	0.9333	0.5789	0.7351
	RF	0.7755	0.8333	0.6842	0.7193
	KNN	0.7143	0.9333	0.3684	0.6965
9 features (HR Min, HR Max, ASDNN 5, SDANN 5, QT Min, QT Mean, QTc Min, QTc > 450, and Longest R-R)	LR	0.7347	0.8333	0.5789	0.6825
	Linear SVM	0.7755	0.8667	0.6316	0.7404
	RBF SVM	0.7959	0.9000	0.6316	0.7509
	RF	0.8163	0.9000	0.6842	0.8368
	KNN	0.7143	1.0000	0.2632	0.6377
10 features (All)	LR	0.7143	0.8000	0.5789	0.6579
	Linear SVM	0.7755	0.8667	0.6316	0.7193
	RBF SVM	0.7755	0.8667	0.6316	0.7351
	RF	0.7959	0.9667	0.5263	0.7825
	KNN	0.6939	0.8333	0.4737	0.6561

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

8.3 Discussion

Turns out that several ECG characteristics are highly correlated with each other, that is, multicollinearity occurs. For patients aged between 40 and 59 years (inclusive), multicollinearity occurred in six variables

(VIF > 5). These variables do not belong to a specific type, in fact, one is from HR (HR Min), two are from HRV (ASDNN 5 and SDNN), and three are from QT analysis (QT Mean, QT Max and QTc Mean). As for the patients aged above 59 years old, multicollinearity occurred in five variables, one from HR (HR Mean), two from HRV (SDNN and RMSSD), and two from QT analysis (QT Max and QTc Mean).

SVM with RBF kernel was the model that showed the best classification performance in the case of the patients aged between 40 and 59 of groups HCM and FD without WMLs (accuracy of 0.8049). Also, this is the algorithm that leads to better classification performances in almost every feature combination. In the studies Hijari et al. [6] and Munla et al. [39], RBF SVM was also the model that led to better classification performances, at classifying if symptoms were or were not expected based on some ECG measurements, and classifying stress level into highly stressed or normal based on HRV analysis, respectively. This model's performance was obtained when using a 4-feature combination (HR Mean, SDANN 5, RMSSD, and QTc > 450). From these features, the only that was expected, taking into consideration the statistical results, was HR Mean, since, this variable was significantly associated with the dependent variable in the logistic regression analysis results of these patients. The remaining variables, that is, SDANN 5, RMSSD, and QTc > 450, were not significantly associated with the dependent variable. The ones that were significantly associated were all excluded during the filter feature selection method, based on VIF values, and, therefore, were not even used as input data to the machine learning models. However, it is worth remembering that the excluded variables by the iterative process based on VIF values can be predicted by other independent variables in the remaining subgroup of variables.

As for the classification of HCM and FD with WMLs patients aged between 40 and 59 years, the best model was KNN with a 6-feature combination (HR Mean, HR Max, RMSSD, QTc Min, QTc Max, and QTc > 450) (accuracy of 0.8462). This is probably the most discrepant result. However, RBF SVM was the second best model with this feature combination, and the best with the other combinations, proving, once again, its value to classification problems based on ECG characteristics. Evaluating the feature combinations, some of the variables (HR Mean, QTc Max and QTc > 450) were expected, since they were significantly associated with the dependent variable (HCM or FD) in the logistic regression results. The remaining variables (HR Max, RMSSD, and QTc Min) were not significantly associated with the dependent variable, however, as explained for HCM and FD without WMLs groups, the other variables that were significantly associated with the dependent variable in the logistic regression results, were not even used as input

variables for the machine learning models, since they (QT Mean, QT Max, and QTc Mean) were eliminated in the first feature selection method, which is based on the VIF value.

When the classification is the presence or absence of WMLs in FD patients aged between 40 and 59 years, random forest algorithm with all 9 features (HR Mean, HR Max, SDANN 5, RMSSD, QT Min, QTc Min, QTc Max, QTc > 450, and Longest R-R) was the model that had better classification performance (accuracy of 0.8125). Random forest was also the best model in the study of Vigier et al. [38], where the aim is to identify if patients have cancer or not. Although RBF SVM model was the best model to predict if symptoms were or were not expected in patients with LQTS, in the study of Hijari et al. [6], random forest was the second best model for this classification. However, as previously mentioned, if we are interested in selecting the best feature combination to this classification problem, then, all model, except RF, had a better classification performance with the 2-feature combination (SDANN 5 and QTc Min), being RBF SVM the best model (accuracy of 0.75). One of the variables, SDANN 5 variable, was excepted, since it was significantly associated with the dependent variable (with WMLs or without WMLs). SDNN variable was also significantly associated, however, it was eliminated during the feature selection method based on VIF value. In the study of Braga (2021), the same classification problem was evaluated, however, the results were slightly different. Using a 4-feature combination (HR Mean, HR Max, QT Max, and SDANN 5) the best algorithm, in Braga (2021) was RF, and using only SDANN 5 feature, the best algorithm was linear SVM. Despite the differences between the results obtained in the current study and in the study of Braga (2021), both studies agree that SDANN 5 feature is important to differentiate between patients with and without WMLs.

In the case of patients aged above 59 years, the group of variables that were highly correlated with each other is different. This group includes five features (HR Mean, SDNN, RMSSD, QT Max, and QTc Mean), that, similar to what was observed in patients between 40 and 59 years, are from all variable types.

For patients above 59 years from HCM and FD with WMLs groups, the best model to this classification problem is random forest with a 9-feature combination (HR Min, HR Max, ASDNN 5, SDANN 5, QT Min, QT Mean, QTc Min, QTc > 450, and Longest R-R) (accuracy of 0.8163). Three of these variables (HR Min, QT Min, and QT Mean) were significantly associated with the dependent variable (HCM or FD), and, therefore, being part of the feature combination that leads to a better classification performance was

expected. There was also another variable that was significantly associated with the dependent variable, in the logistic regression results of this case, which was HR Mean, however, as previously mentioned, this variable was eliminated during the first feature selection method (VIF value > 5). In the study of Vigier et al. [38], random forest was also the best model to classify cancer and health patients, based on HRV features .

There are some other interesting findings in the machine learning results obtained in the current study. One is the fact that, when performing the RFE feature selection method, random forest model always selects a 9-feature combination, which is all possible features in case of patients aged between 40 and 59 years (including) and one feature less than the possible features in case of patients aged above 59 years. Another finding is that logistic regression and linear SVM models are never the model with the best classification performance, in any classification problem or feature combination.

An important finding is that sensitivity values are higher than specificity values in the classification problems that include the HCM group. Since the positive class is having FD, it means that the models are better at predicting that a patient has FD rather than predicting that a patient has HCM. By contrast, when the classification problem is the presence or absence of WMLs among FD patients, sensitivity values are lower than specificity values, which means that models are better at predicting that a patient does not have WMLs rather than predicting that a patient has WMLs (having WMLs is the positive class). High specificity when screening a disease is a good thing since it means that there are fewer false positives, and therefore, fewer people without the disease screen positive, this way avoiding unnecessary diagnostic procedures. These differences may be due to the different amount of patients in each class, since, looking at patients aged between 40 and 59 years (including), the amount of patients with HCM is lower than the amount of patients with FD, either with or without WMLs (positive class). The amount of patients without WMLs (negative class) is also superior to the amount of patients with WMLs. This difference is also present in patients aged above 59 years, where the amount of HCM patients is inferior to the amount of FD with WMLs patients.

Chapter 9

Conclusion, limitations and future work

To finalize this dissertation, in this chapter, the general conclusions, the limitations, and the work that could be done in future similar research are portrayed.

9.1 General conclusions

The main goal of this dissertation was to detect important ECG characteristics for the diagnosis of Fabry disease, through statistics and machine learning methods. This was inspired by the power of machine learning to identify meaningful patterns in data, and the fact that this could be helpful to improve the diagnosis of Fabry disease, mainly at reducing the existing delay between the onset of symptoms and the diagnosis of the disease.

First of all, the objective was to detect, on ECG data, important biomarkers to differentiate between the groups in interest in this study (FD patients without WMLs and HCM patients, FD patients with WMLs and HCM patients, and FD patients with WMLs and FD patients without WMLs). Next, investigate if patients' age and sex could be correlated with the ECG data, and lastly, implement different machine learning algorithms on the ECG data to differentiate the groups and evaluate their classification performance.

The goal of detecting biomarkers in the ECG data was successfully achieved, since some ECG characteristics were identified to be relevant to differentiate the groups, which is the case of SDANN 5 variable,

to differentiate between FD patients with and without WMLs.

Age and sex were, in fact, significantly different between some of the groups, and age was identified to be correlated with some of the ECG characteristics, since some of these were no longer significantly associated with the dependent variable (group class) in the logistic regression analysis adjusted to age. An example of this situation is HR Max variable on the logistic regression analysis of HCM and FD without WMLs groups.

As for machine learning, five algorithms were successfully implemented and evaluated for each classification problem. The best model to differentiate HCM patients and FD without WMLs patients was RBF SVM with an accuracy of 0.8049, using HR Mean, SDANN 5, RMSSD, and QTc > 450 features. The best model to distinguish between HCM patients and FD with WMLs patients was KNN with HR Mean, HR Max, RMSSD, QTc Min, QTc Max, and QTc > 450 features, with an accuracy of 0.8462. To distinguish patients with and without WMLs among FD patients, the best model was random forest with HR Mean, HR Max, SDANN 5, RMSSD, QT Min, QTc Min, QTc Max, QTc > 450, and Longest R-R features, where accuracy was 0.8125. These models were obtained with patients aged between 40 and 59 years (including), where no significant age differences between groups were found.

As to distinguish between HCM and FD with WMLs groups, with patients aged above 59 years, the best model was random forest with HR Min, HR Max, ASDNN 5, SDANN 5, QT Min, QT Mean, QTc Min, QTc > 450, and Longest R-R, which yielded an accuracy of 0.8163.

The accomplished results have an important contribution to better understanding how ECG data can be used to diagnose Fabry disease as well as the machine learning models that are better suited to the different classification problems evaluated in the dissertation, and these are crucial steps to further develop a system that would be able to assist in the diagnosis of Fabry disease.

9.2 Limitations and future work

The main limitation of this dissertation was the small number of patients, in all the groups, HCM, FD without WMLs, and FD with WMLs. This limitation affects the reliability of the results, since the models

were trained with this small amount of data, and the general performance of the models depends on the size of the input dataset. Applying machine learning to small datasets increases the possibility of overfitting [80]. This small dataset also influenced the test dataset, which was not defined in order to maximize the size of the training dataset.

Therefore, the results obtained in this dissertation should be confirmed with a larger dataset, including not only a train set but also a test set, which is of extreme importance to develop a decision support system to diagnose Fabry disease.

Also, although some variables had been eliminated in the filter feature selection method, due to high correlation with other variables, the group of features obtained in the logistic regression analysis should be tested as input data to machine learning algorithms. The performance of these models (where the input data includes only the features that were significantly associated with the dependent variable in the logistic regression results) can then be compared with the performance of the models that were obtained in this dissertation.

The present dissertation implemented and evaluated the performance of five different Machine Learning models, i.e. Logistic regression, SVM with linear kernel, random forest, SVM with RBF kernel, and KNN. However, other machine learning algorithms should also be implemented and evaluated in future research. Deep learning algorithms should also be considered.

Another potential improvement to the classifiers' performances could result from a more complete dataset, with not only ECG, age, and sex variables, as the dataset used in the current dissertation, but also other dichotomous variables associated with the presence or absence of medical conditions, like hypertension, anxiety, diabetes, depression, obesity, among others. Some of these variables had shown to be correlated with the presence of WMLs [42] [81] [82] and with risk factors for heart disease [83] [84].

Healthy patients should also be considered in the diagnosis of Fabry disease, thus, further studies should include a healthy control group, this way, differentiating FD patients with and without WMLs from healthy individuals.

Bibliography

- [1] R. Giugliani, D.-M. Niu, U. Ramaswami, *et al.*, “A 15-Year Perspective of the Fabry Outcome Survey,” en, *Journal of Inborn Errors of Metabolism and Screening*, vol. 4, p. 232 640 981 666 629, Jul. 2016, ISSN: 2326-4098, 2326-4594. DOI: 10 . 1177/2326409816666298. [Online]. Available: <http://journals.sagepub.com/doi/10.1177/2326409816666298>.
- [2] S. Kant and M. G. Atta, “Therapeutic advances in Fabry disease: The future awaits,” en, *Biomedicine & Pharmacotherapy*, vol. 131, p. 110 779, Nov. 2020, ISSN: 07533322. DOI: 10 . 1016 / j . biopha . 2020 . 110779. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0753332220309720>.
- [3] R. Reisin, A. Perrin, and P. García-Pavía, “Time delays in the diagnosis and treatment of Fabry disease,” en, *International Journal of Clinical Practice*, vol. 71, no. 1, e12914, Jan. 2017, ISSN: 13685031. DOI: 10 . 1111 / ijcp . 12914. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1111/ijcp.12914>.
- [4] O. Azevedo, M. F. Gago, G. Miltenberger-Miltenyi, *et al.*, “Natural history of the late-onset phenotype of Fabry disease due to the p.F113L mutation,” en, *Molecular Genetics and Metabolism Reports*, vol. 22, p. 100 565, Mar. 2020, ISSN: 22144269. DOI: 10 . 1016 / j . ymgmr . 2020 . 100565. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2214426919302113>.
- [5] N. Junqua, D. Legallois, S. Segard, *et al.*, “The value of electrocardiography and echocardiography in distinguishing Fabry disease from sarcomeric hypertrophic cardiomyopathy,” en, *Archives of Cardiovascular Diseases*, vol. 113, no. 8-9, pp. 542–550, Aug. 2020, ISSN: 18752136. DOI:

-
- 10.1016/j.acvd.2020.04.008. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S187521362030156X>.
- [6] S. Hijazi, A. Page, B. Kantarci, and T. Soyata, "Machine Learning in Cardiac Health Monitoring and Decision Support," en, *Computer*, vol. 49, no. 11, pp. 38–48, Nov. 2016, ISSN: 0018-9162. DOI: 10.1109/MC.2016.339. [Online]. Available: <http://ieeexplore.ieee.org/document/7742297/>.
- [7] D. Brito, N. Cardim, L. R. Lopes, *et al.*, "Awareness of Fabry disease in cardiology: A gap to be filled," en, *Revista Portuguesa de Cardiologia (English Edition)*, vol. 37, no. 6, pp. 457–466, Jun. 2018, ISSN: 21742049. DOI: 10.1016/j.repce.2018.03.015. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2174204918301995>.
- [8] M. Namdar, "Electrocardiographic Changes and Arrhythmia in Fabry Disease," en, *Frontiers in Cardiovascular Medicine*, vol. 3, Mar. 2016, ISSN: 2297-055X. DOI: 10.3389/fcvm.2016.00007. [Online]. Available: <http://journal.frontiersin.org/Article/10.3389/fcvm.2016.00007/abstract>.
- [9] A. Hagège, P. Réant, G. Habib, *et al.*, "Fabry disease in cardiology practice: Literature review and expert point of view," en, *Archives of Cardiovascular Diseases*, vol. 112, no. 4, pp. 278–287, Apr. 2019, ISSN: 18752136. DOI: 10.1016/j.acvd.2019.01.002. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1875213619300397>.
- [10] P. Aguiar, M. F. Gago, M. G. Marques, *et al.*, "Fabry disease in adults," English, 2020.
- [11] R. Sharma, S. Sekhon, and M. Cascella, "White Matter Lesions," eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2021. [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK562167/>.
- [12] J. Lin, D. Wang, L. Lan, and Y. Fan, "Multiple Factors Involved in the Pathogenesis of White Matter Lesions," en, *BioMed Research International*, vol. 2017, pp. 1–9, 2017, ISSN: 2314-6133, 2314-6141. DOI: 10.1155/2017/9372050. [Online]. Available: <https://www.hindawi.com/journals/bmri/2017/9372050/>.

-
- [13] H. Jokinen, N. Gonçalves, R. Vigário, *et al.*, “Early-Stage White Matter Lesions Detected by Multi-spectral MRI Segmentation Predict Progressive Cognitive Decline,” *Frontiers in Neuroscience*, vol. 9, Dec. 2015, ISSN: 1662-453X. DOI: 10.3389/fnins.2015.00455. [Online]. Available: <http://journal.frontiersin.org/Article/10.3389/fnins.2015.00455/abstract>.
- [14] D. L. F. Martins, “Fabry disease associated with GLA p.Phe113Leu variant for a common ancestor in Portuguese and Italians, and use of linked markers to estimate the age of the mutation,” pt, Ph.D. dissertation.
- [15] *Migalastat*. [Online]. Available: <https://go.drugbank.com/drugs/DB05018>.
- [16] C. V. Tuohy, S. Kaul, H. K. Song, B. Nazer, and S. B. Heitner, “Hypertrophic cardiomyopathy: The future of treatment,” en, *European Journal of Heart Failure*, vol. 22, no. 2, pp. 228–240, Feb. 2020, ISSN: 1388-9842, 1879-0844. DOI: 10.1002/ejhf.1715. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/ejhf.1715>.
- [17] J. van der Velden, C. Y. Ho, J. C. Tardiff, I. Olivetto, B. C. Knollmann, and L. Carrier, “Research priorities in sarcomeric cardiomyopathies,” en, *Cardiovascular Research*, vol. 105, no. 4, pp. 449–456, Apr. 2015, ISSN: 0008-6363, 1755-3245. DOI: 10.1093/cvr/cvv019. [Online]. Available: <https://academic.oup.com/cardiovascres/article-lookup/doi/10.1093/cvr/cvv019>.
- [18] A. J. Marian and E. Braunwald, “Hypertrophic Cardiomyopathy: Genetics, Pathogenesis, Clinical Manifestations, Diagnosis, and Therapy,” en, *Circulation Research*, vol. 121, no. 7, pp. 749–770, Sep. 2017, ISSN: 0009-7330, 1524-4571. DOI: 10.1161/CIRCRESAHA.117.311059. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.117.311059>.
- [19] G. Hasenfuss, “Sarcomeric cardiomyopathies: From bedside to bench and back,” en, *Cardiovascular Research*, vol. 105, no. 4, pp. 395–396, Apr. 2015, ISSN: 0008-6363, 1755-3245. DOI: 10.1093/cvr/cvv035. [Online]. Available: <https://academic.oup.com/cardiovascres/article-lookup/doi/10.1093/cvr/cvv035>.
- [20] A. T. Owens and N. Reza, “Diagnosis of Hypertrophic Cardiomyopathy: What Every Cardiologist Needs to Know,” en, p. 5,

-
- [21] A. Weissler-Snir, "How NOT to miss Hypertrophic Cardiomyopathy?" en, p. 28,
- [22] D. Jacoby, M. H. Beasley, and K. B. Churchwell, "Treatment of Hypertrophic Cardiomyopathy: What Every Cardiologist Needs to Know," en, p. 6,
- [23] J. R. Levick, *An Introduction to Cardiovascular Physiology*. English. Oxford: Elsevier Science, 2014, OCLC: 1058490774, ISBN: 978-1-4831-8384-8. [Online]. Available: <http://qut.ebilib.com.au/patron/FullRecord.aspx?p=1635221>.
- [24] D. L. P. updated, *Understanding an ECG | ECG Interpretation | Geeky Medics*, en-GB, Section: Cardiology. [Online]. Available: <https://geekymedics.com/understanding-an-ecg/>.
- [25] *Basic Electrocardiography*, English. 2020, OCLC: 1199793340, ISBN: 978-3-030-32886-3. [Online]. Available: <https://doi.org/10.1007/978-3-030-32886-3>.
- [26] B. Aehlert, *ECGs made easy*, Sixth edition. Phoenix, Arizona: Southwest EMS Education, Inc, 2018, ISBN: 978-0-323-40130-2.
- [27] G. S. Wagner and D. G. Strauss, *Marriott's practical electrocardiography [12th ed.* English. Wolters Kluwer/Lippincott Williams & Wilkins, 2014, OCLC: 1198438888, ISBN: 978-1-4511-4625-7.
- [28] *Holter Monitor*, en. [Online]. Available: <https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/holter-monitor>.
- [29] M. Namdar, J. Steffel, M. Vidovic, *et al.*, "Electrocardiographic changes in early recognition of Fabry disease," en, *Heart*, vol. 97, no. 6, pp. 485–490, Mar. 2011, ISSN: 1355-6037. DOI: 10.1136/hrt.2010.211789. [Online]. Available: <https://heart.bmj.com/lookup/doi/10.1136/hrt.2010.211789>.
- [30] S. Nordin, R. Kozor, S. Baig, *et al.*, "Cardiac Phenotype of Prehypertrophic Fabry Disease," en, *Circulation: Cardiovascular Imaging*, vol. 11, no. 6, Jun. 2018, ISSN: 1941-9651, 1942-0080. DOI: 10.1161/CIRCIMAGING.117.007168. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.117.007168>.

-
- [31] M. Namdar, J. Steffel, S. Jetzer, *et al.*, "Value of Electrocardiogram in the Differentiation of Hypertensive Heart Disease, Hypertrophic Cardiomyopathy, Aortic Stenosis, Amyloidosis, and Fabry Disease," en, *The American Journal of Cardiology*, vol. 109, no. 4, pp. 587–593, Feb. 2012, ISSN: 00029149. DOI: 10.1016/j.amjcard.2011.09.052. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0002914911030505>.
- [32] J. A. F. Braga, "Machine learning for differential diagnosis of white matter lesions in Fabry Disease patients based on gait and cardiac characteristics," en, p. 116,
- [33] S. Galluzzi, F. Nicosia, C. Geroldi, *et al.*, "Cardiac Autonomic Dysfunction Is Associated With White Matter Lesions in Patients With Mild Cognitive Impairment," en, *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 64A, no. 12, pp. 1312–1315, Dec. 2009. DOI: 10.1093/gerona/ glp105. [Online]. Available: <https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/ glp105>.
- [34] I. El Naqa and M. J. Murphy, "What Is Machine Learning?" en, in *Machine Learning in Radiation Oncology*, I. El Naqa, R. Li, and M. J. Murphy, Eds., Cham: Springer International Publishing, 2015, pp. 3–11, ISBN: 978-3-319-18304-6. DOI: 10.1007/978-3-319-18305-3_1. [Online]. Available: http://link.springer.com/10.1007/978-3-319-18305-3_1.
- [35] M. Gagliano, J. V. Pham, B. Tang, H. Kashif, and J. Ban, "Applications of Machine Learning in Medical Diagnosis," en, p. 6,
- [36] J. Schaefer, M. Lehne, J. Schepers, F. Prasser, and S. Thun, "The use of machine learning in rare diseases: A scoping review," en, *Orphanet Journal of Rare Diseases*, vol. 15, no. 1, p. 145, Dec. 2020, ISSN: 1750-1172. DOI: 10.1186/s13023-020-01424-6. [Online]. Available: <https://ojrd.biomedcentral.com/articles/10.1186/s13023-020-01424-6>.
- [37] K. Arun Bhavsar, A. Abugabah, J. Singla, A. Ali AlZubi, A. Kashif Bashir, and Nikita, "A Comprehensive Review on Medical Diagnosis Using Machine Learning," en, *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1997–2014, 2021, ISSN: 1546-2226. DOI: 10.32604/cmc.2021.014943. [Online]. Available: <https://www.techscience.com/cmc/v67n2/41354>.

-
- [38] M. Vigier, B. Vigier, E. Andritsch, and A. R. Schwerdtfeger, "Cancer classification using machine learning and HRV analysis: Preliminary evidence from a pilot study," en, *Scientific Reports*, vol. 11, no. 1, p. 22 292, Dec. 2021, ISSN: 2045-2322. DOI: 10 . 1038 / s41598 – 021 – 01779 – 1. [Online]. Available: <https://www.nature.com/articles/s41598-021-01779-1>.
- [39] N. Munla, M. Khalil, A. Shahin, and A. Mourad, "Driver stress level detection using HRV analysis," en, in *2015 International Conference on Advances in Biomedical Engineering (ICABME)*, Beirut, Lebanon: IEEE, Sep. 2015, pp. 61–64, ISBN: 978-1-4673-6516-1. DOI: 10 . 1109 / ICABME . 2015 . 7323251. [Online]. Available: <http://ieeexplore.ieee.org/document/7323251/>.
- [40] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," en, *MACHINE LEARNING IN PYTHON*, p. 6,
- [41] G. R. Geovanini, E. R. Vasques, R. De Oliveira Alvim, et al., "Age and Sex Differences in Heart Rate Variability and Vagal Specific Patterns – Baependi Heart Study," en, *Global Heart*, vol. 15, no. 1, p. 71, Oct. 2020, ISSN: 2211-8179, 2211-8160. DOI: 10 . 5334 / gh . 873. [Online]. Available: <https://globalheartjournal.com/article/10.5334/gh.873/>.
- [42] A. R. Moura, F. Ferreira, E. Bicho, et al., "The association between potential risk factors and cerebral white matter lesions in Fabry Disease patients," en, *Journal of Statistics on Health Decision*, 15–17 Páginas, Jul. 2021, Artwork Size: 15-17 Páginas Publisher: Journal of Statistics on Health Decision. DOI: 10 . 34624 / JSHD . V3I1 . 24751. [Online]. Available: <https://proa.ua.pt/index.php/jshd/article/view/24751>.
- [43] S. Körver, M. Vergouwe, C. E. Hollak, I. N. van Schaik, and M. Langeveld, "Development and clinical consequences of white matter lesions in Fabry disease: A systematic review," en, *Molecular Genetics and Metabolism*, vol. 125, no. 3, pp. 205–216, Nov. 2018, ISSN: 10967192. DOI: 10 . 1016 / j . ymgme . 2018 . 08 . 014. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1096719218304736>.
- [44] W. McKinney, "Data Structures for Statistical Computing in Python," en, Austin, Texas, 2010, pp. 56–61. DOI: 10 . 25080 / Majora – 92bf1922 – 00a. [Online]. Available: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.

-
- [45] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, “SciPy 1.0: Fundamental algorithms for scientific computing in Python,” en, *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020. DOI: 10.1038/s41592-019-0686-2. [Online]. Available: <http://www.nature.com/articles/s41592-019-0686-2>.
- [46] S. Seabold and J. Perktold, “Statsmodels: Econometric and Statistical Modeling with Python,” en, Austin, Texas, 2010, pp. 92–96. DOI: 10.25080/Majora-92bf1922-011. [Online]. Available: <https://conference.scipy.org/proceedings/scipy2010/seabold.html>.
- [47] R. Vallat, “Pingouin: Statistics in Python,” en, *Journal of Open Source Software*, vol. 3, no. 31, p. 1026, Nov. 2018, ISSN: 2475-9066. DOI: 10.21105/joss.01026. [Online]. Available: <http://joss.theoj.org/papers/10.21105/joss.01026>.
- [48] S. Silvey, *Statistical Inference*, en, 1st ed. Routledge, Oct. 2017, ISBN: 978-0-203-73864-1. DOI: 10.1201/9780203738641. [Online]. Available: <https://www.taylorfrancis.com/books/9781351414500>.
- [49] A. Petrie and C. Sabin, *Medical statistics at a glance*, 2nd ed. Malden, Mass: Blackwell, 2005, ISBN: 978-1-4051-2780-6.
- [50] A. P. Field, *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll*, English. 2013, OCLC: 1029430457, ISBN: 1-4462-7457-8.
- [51] J. Brownlee, *Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python*, en.
- [52] K. Moder, “Alternatives to F-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study),” en, p. 12,
- [53] D. Rasch, K. D. Kubinger, and K. Moder, “The two-sample t test: Pre-testing its assumptions does not pay off,” en, *Statistical Papers*, vol. 52, no. 1, pp. 219–231, Feb. 2011, ISSN: 0932-5026, 1613-9798. DOI: 10.1007/s00362-009-0224-x. [Online]. Available: <http://link.springer.com/10.1007/s00362-009-0224-x>.
- [54] D. Cramer and D. Howitt, *The Sage dictionary of statistics: a practical resource for students in the social sciences*. London ; Thousand Oaks, CA: Sage Publications, 2004, ISBN: 978-0-7619-4137-8.

-
- [55] M. Allen, *The SAGE encyclopedia of communication research methods*, English. 2017, OCLC: 983736255, ISBN: 978-1-4833-8142-8.
- [56] S. Lee and D. K. Lee, "What is the proper way to apply the multiple comparison test?" en, *Korean Journal of Anesthesiology*, vol. 71, no. 5, pp. 353–360, Oct. 2018, ISSN: 2005-6419, 2005-7563. DOI: 10.4097/kja.d.18.00242. [Online]. Available: <http://ekja.org/journal/view.php?doi=10.4097/kja.d.18.00242>.
- [57] N. J. Salkind and K. Rasmussen, Eds., *Encyclopedia of measurement and statistics*. Thousand Oaks, Calif: SAGE Publications, 2007, ISBN: 978-1-4129-1611-0.
- [58] J. H. McDonald, *Handbook of Biological Statistics*, 3rd ed. Baltimore, Maryland: Sparky House Publishing.
- [59] J. E. De Muth, *Basic statistics and pharmaceutical statistical applications*, 2nd ed, ser. Biostatistics 16. Boca Raton: Chapman & Hall/CRC, 2006, OCLC: ocm64771022, ISBN: 978-0-8493-3799-4.
- [60] A. Dinno, "Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test," en, *The Stata Journal: Promoting communications on statistics and Stata*, vol. 15, no. 1, pp. 292–300, Apr. 2015. DOI: 10.1177/1536867X1501500117. [Online]. Available: <http://journals.sagepub.com/doi/10.1177/1536867X1501500117>.
- [61] P. Laake and M. W. Fagerland, "Statistical Inference," en, in *Research in Medical and Biological Sciences*, Elsevier, 2015, pp. 379–430, ISBN: 978-0-12-799943-2. DOI: 10.1016/B978-0-12-799943-2.00011-2. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B9780127999432000112>.
- [62] S. Gilliver and N. Valveny, "How to interpret and report the results from multivariable analyses," *European Medical Writers Association (EMWA)*, vol. 25, no. 3, pp. 37–42, 2016.
- [63] J. B. Strait and E. G. Lakatta, "Aging-Associated Cardiovascular Changes and Their Relationship to Heart Failure," en, *Heart Failure Clinics*, vol. 8, no. 1, pp. 143–164, Jan. 2012, ISSN: 15517136. DOI: 10.1016/j.hfc.2011.08.011. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1551713611001012>.

-
- [64] K. Umetani, D. H. Singer, R. McCraty, and M. Atkinson, "Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender Over Nine Decades," en, *Journal of the American College of Cardiology*, vol. 31, no. 3, pp. 593–601, Mar. 1998, ISSN: 07351097. DOI: 10.1016/S0735-1097(97)00554-8. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0735109797005548>.
- [65] J. Brownlee, *Master Machine Learning Algorithms: : Discover How They Work and Implement Them From Scratch*, en. Machine Learning Mastery, 2016.
- [66] D. Sarkar, R. Bali, and T. Sharma, *Practical Machine Learning with Python*, en. Berkeley, CA: Apress, 2018, ISBN: 978-1-4842-3207-1. DOI: 10.1007/978-1-4842-3207-1. [Online]. Available: <http://link.springer.com/10.1007/978-1-4842-3207-1>.
- [67] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," en, *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015, ISSN: 00313203. DOI: 10.1016/j.patcog.2015.03.009. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0031320315000989>.
- [68] L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, and L. Chiari, "A Clinical Application of Feature Selection: Quantitative Evaluation of the Locomotor Function," en, in *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, A. Fred, J. L. G. Dietz, K. Liu, and J. Filipe, Eds., vol. 272, Series Title: Communications in Computer and Information Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 151–157, ISBN: 978-3-642-29764-9. DOI: 10.1007/978-3-642-29764-9_10. [Online]. Available: http://link.springer.com/10.1007/978-3-642-29764-9_10.
- [69] J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. Machine Learning Mastery, 2020.
- [70] M. Kuhn and K. Johnson, *Feature engineering and selection: a practical approach for predictive models*, eng, ser. Chapman & Hall/CRC data science series. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2020, ISBN: 978-1-351-60946-3.

-
- [71] M. O. Akinwande, H. G. Dikko, and A. Samson, "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis," en, *Open Journal of Statistics*, vol. 05, no. 07, pp. 754–767, 2015, ISSN: 2161-718X, 2161-7198. DOI: 10.4236/ojs.2015.57075. [Online]. Available: <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ojs.2015.57075>.
- [72] J. H. Kim, "Multicollinearity and misleading statistical results," en, *Korean Journal of Anesthesiology*, vol. 72, no. 6, pp. 558–569, Dec. 2019, ISSN: 2005-6419, 2005-7563. DOI: 10.4097/kja.19087. [Online]. Available: <http://ekja.org/journal/view.php?doi=10.4097/kja.19087>.
- [73] J. Dukart, "Basic Concepts of Image Classification Algorithms Applied to Study Neurodegenerative Diseases," en, in *Brain Mapping*, Elsevier, 2015, pp. 641–646, ISBN: 978-0-12-397316-0. DOI: 10.1016/B978-0-12-397025-1.00072-5. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B9780123970251000725>.
- [74] M. Fratello and R. Tagliaferri, "Decision Trees and Random Forests," en, in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 374–383, ISBN: 978-0-12-811432-2. DOI: 10.1016/B978-0-12-809633-8.20337-3. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B9780128096338203373>.
- [75] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees," en, vol. 9, no. 5, p. 8, 2012.
- [76] M.-H. Chiu, Y.-R. Yu, H. L. Liaw, and L. Chun-Hao, "THE USE OF FACIAL MICRO-EXPRESSION STATE AND TREE-Forest MODEL FOR PREDICTING CONCEPTUAL-CONFLICT BASED CONCEPTUAL CHANGE," en, p. 10,
- [77] T. Fawcett, "An introduction to ROC analysis," en, *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006, ISSN: 01678655. DOI: 10.1016/j.patrec.2005.10.010. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X>.
- [78] Z.-H. Zhou, *Machine Learning*, English. 2021, ISBN: 9789811519673 OCLC: 1268185842.

-
- [79] J. Alber, S. Alladi, H.-J. Bae, *et al.*, “White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities,” en, *Alzheimer’s & Dementia: Translational Research & Clinical Interventions*, vol. 5, no. 1, pp. 107–117, Jan. 2019, ISSN: 2352-8737, 2352-8737. DOI: 10.1016/j.trci.2019.02.001. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1016/j.trci.2019.02.001> (visited on 03/16/2022).
- [80] S. J. Al’Aref, G. Singh, and L. Baskaran, Eds., *Machine learning in cardiovascular medicine*, 1st ed. Waltham: Elsevier, 2020, ISBN: 978-0-12-820273-9.
- [81] F.-E. de Leeuw, J. C. de Groot, M. Oudkerk, *et al.*, “Hypertension and cerebral white matter lesions in a prospective cohort study,” en, *Brain*, vol. 125, no. 4, pp. 765–772, Mar. 2002, ISSN: 1460-2156, 0006-8950. DOI: 10.1093/brain/awf077. [Online]. Available: <https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awf077>.
- [82] M. Johansson, E. Stomrud, O. Lindberg, *et al.*, “Apathy and anxiety are early markers of Alzheimer’s disease,” en, *Neurobiology of Aging*, vol. 85, pp. 74–82, Jan. 2020, ISSN: 01974580. DOI: 10.1016/j.neurobiolaging.2019.10.008. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0197458019303719>.
- [83] L. O. Lee, K. J. Grimm, A. Spiro, and L. D. Kubzansky, “Neuroticism, Worry, and Cardiometabolic Risk Trajectories: Findings From a 40-Year Study of Men,” en, *Journal of the American Heart Association*, vol. 11, no. 3, Feb. 2022, ISSN: 2047-9980. DOI: 10.1161/JAHA.121.022006. [Online]. Available: <https://www.ahajournals.org/doi/10.1161/JAHA.121.022006>.
- [84] I. Csige, D. Ujvárosy, Z. Szabó, *et al.*, “The Impact of Obesity on the Cardiovascular System,” en, *Journal of Diabetes Research*, vol. 2018, pp. 1–12, Nov. 2018, ISSN: 2314-6745, 2314-6753. DOI: 10.1155/2018/3407306. [Online]. Available: <https://www.hindawi.com/journals/jdr/2018/3407306/>.

Appendices

Appendix A

Statistical Analysis Appendix

Table 22: Normality of HCM, FD with WMLs and FD without WMLs groups and homogeneity of variance results

Variables	Normality - pvalues			Homogeneity of variances - pvalues
	HCM	FD with WMLs	FD without WMLs	
Age	p = 0.4713	p = 0.0591	p = 0.2180	p = 0.5073
HR Min	p = 0.6111	p = 0.1857	p = 0.6791	p = 0.1385
HR Mean	p = 0.8310	p = 0.5717	p = 0.7133	p = 0.0091
HR Max	p = 0.5331	p = 0.6100	p = 0.8439	p = 0.0001
ASDNN 5	p = 0.0001	p = 0.0003	p = 0.0019	-
SDANN 5	p = 0.5509	p = 0.1660	p = 0.0002	-
SDNN	p = 0.0002	p = 0.1267	p = 0.0007	-
RMSSD	p < 0.0001	p < 0.0001	p < 0.0001	-
QT Min	p = 0.3504	p = 0.2642	p = 0.6387	p = 0.2243
QT Mean	p = 0.0390	p = 0.1012	p = 0.0002	-
QT Max	p < 0.0001	p < 0.0001	p < 0.0001	-
QTc Min	p = 0.5694	p = 0.0030	p = 0.2402	-
QTc Mean	p = 0.5896	p = 0.0097	p = 0.0011	-
QTc Max	p = 0.0017	p = 0.0003	p < 0.0001	-
QTc > 450	p = 0.0003	p < 0.0001	p < 0.0001	-
Longest R-R	p < 0.0001	p < 0.0001	p = 0.0002	-

Table 23: Statistically non-significant logistic regression results of HCM and FD without WMLs groups

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
ASDNN 5	1.00 (0.98 - 1.02)	0.9359	0.98 (0.96 - 1.00)	0.0921	1.00 (0.98 - 1.02)	0.8666
Age	-	-	0.90 (0.86 - 0.94)	< 0.0001	-	-
Sex	-	-	-	-	3.06 (1.29 - 7.26)	0.0113
SDANN 5	1.01 (0.99 - 1.02)	0.3822	0.99 (0.98 - 1.01)	0.4865	1.01 (1.00 - 1.02)	0.1276
Age	-	-	0.91 (0.87 - 0.95)	< 0.0001	-	-
Sex	-	-	-	-	3.67 (1.48 - 9.14)	0.0051
SDNN	1.00 (0.99 - 1.01)	0.6126	0.99 (0.98 - 1.00)	0.1239	1.00 (0.99 - 1.01)	0.8340
Age	-	-	0.91 (0.87 - 0.94)	< 0.0001	-	-
Sex	-	-	-	-	3.02 (1.26 - 7.19)	0.0128
RMSSD	0.99 (0.98 - 1.00)	0.0753	0.99 (0.98 - 1.00)	0.2187	0.99 (0.98 - 1.00)	0.0564
Age	-	-	0.91 (0.88 - 0.95)	< 0.0001	-	-
Sex	-	-	-	-	3.42 (1.39 - 8.42)	0.0074
QTc Min	1.00 (0.99 - 1.01)	0.8393	1.00 (0.98 - 1.01)	0.8016	1.00 (0.99 - 1.02)	0.6693
Age	-	-	0.91 (0.87 - 0.95)	< 0.0001	-	-
Sex	-	-	-	-	3.10 (1.30 - 7.41)	0.0107

Table 24: Statistically non-significant logistic regression results of HCM and FD with WMLs groups

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
ASDNN 5	1.00 (0.98 - 1.01)	0.6438	1.00 (0.98 - 1.01)	0.5417	1.00 (0.98 - 1.01)	0.6021
Age	-	-	0.99 (0.96 - 1.02)	0.4156	-	-
Sex	-	-	-	-	3.33 (1.43 - 7.75)	0.0053
SDANN 5	1.01 (0.99 - 1.02)	0.4018	1.00 (0.99 - 1.02)	0.4700	1.01 (1.00 - 1.03)	0.1075
Age	-	-	0.99 (0.97 - 1.02)	0.5693	-	-
Sex	-	-	-	-	4.09 (1.66 - 10.10)	0.0023
SDNN	1.00 (0.99 - 1.01)	0.4678	1.00 (0.99 - 1.01)	0.4138	1.00 (0.99 - 1.01)	0.6098
Age	-	-	0.99 (0.96 - 1.02)	0.4208	-	-
Sex	-	-	-	-	3.25 (1.40 - 7.58)	0.0063
RMSSD	1.00 (0.99 - 1.00)	0.3032	1.00 (0.99 - 1.00)	0.3330	1.00 (0.99 - 1.00)	0.1533
Age	-	-	0.99 (0.97 - 1.02)	0.5387	-	-
Sex	-	-	-	-	3.66 (1.53 - 8.73)	0.0035
QTc Min	1.00 (0.99 - 1.01)	0.7493	1.00 (0.99 - 1.01)	0.8024	1.00 (0.99 - 1.02)	0.5615
Age	-	-	0.99 (0.96 - 1.02)	0.4938	-	-
Sex	-	-	-	-	3.39 (1.45 - 7.94)	0.0049
QTc Mean	0.99 (0.98 - 1.00)	0.1212	0.99 (0.98 - 1.00)	0.1534	0.99 (0.97 - 1.00)	0.1121
Age	-	-	0.99 (0.97 - 1.02)	0.7125	-	-
Sex	-	-	-	-	3.38 (1.44 - 7.98)	0.0053
QTc > 450	0.99 (0.98 - 1.00)	0.1113	0.99 (0.98 - 1.00)	0.1468	0.99 (0.98 - 1.00)	0.1919
Age	-	-	1.00 (0.97 - 1.02)	0.7874	-	-
Sex	-	-	-	-	3.15 (1.34 - 7.38)	0.0083
Longest R-R	0.74 (0.47 - 1.16)	0.1882	0.75 (0.47 - 1.19)	0.2269	0.65 (0.40 - 1.06)	0.0843
Age	-	-	0.99 (0.97 - 1.02)	0.6381	-	-
Sex	-	-	-	-	3.79 (1.57 - 9.16)	0.0030

Table 25: Statistically non-significant logistic regression results of FD with WMLs and FD without WMLs groups

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	1.01 (0.95 - 1.07)	0.7366	0.97 (0.91 - 1.04)	0.4299	1.01 (0.95 - 1.08)	0.7794
Age	-	-	1.08 (1.04 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.04 (0.46 - 2.34)	0.9233
HR Mean	0.97 (0.93 - 1.02)	0.2273	0.99 (0.94 - 1.04)	0.6324	0.97 (0.92 - 1.02)	0.1894
Age	-	-	1.07 (1.04 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.26 (0.57 - 2.78)	0.5715
ASDNN 5	0.99 (0.98 - 1.01)	0.4841	1.02 (1.00 - 1.05)	0.0846	0.99 (0.98 - 1.01)	0.4937
Age	-	-	1.09 (1.05 - 1.13)	< 0.0001	-	-
Sex	-	-	-	-	1.06 (0.50 - 2.27)	0.8806
SDANN 5	1.00 (0.99 - 1.01)	0.8253	1.01 (0.99 - 1.02)	0.3228	1.00 (0.99 - 1.01)	0.8651
Age	-	-	1.08 (1.05 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.06 (0.49 - 2.33)	0.8766
SDNN	1.00 (0.99 - 1.01)	0.8750	1.01 (1.00 - 1.02)	0.2068	1.00 (0.99 - 1.01)	0.9068
Age	-	-	1.08 (1.05 - 1.12)	< 0.0001	-	-
Sex	-	-	-	-	1.07 (0.50 - 2.32)	0.8559
RMSSD	1.01 (1.00 - 1.02)	0.1127	1.01 (1.00 - 1.03)	0.1354	1.01 (1.00 - 1.02)	0.1144
Age	-	-	1.08 (1.04 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.03 (0.48 - 2.20)	0.9493
QT Max	1.00 (1.00 - 1.01)	0.1149	1.00 (0.99 - 1.00)	0.3091	1.00 (1.00 - 1.01)	0.1157
Age	-	-	1.08 (1.05 - 1.12)	< 0.0001	-	-
Sex	-	-	-	-	1.07 (0.50 - 2.30)	0.8566
QTc Min	1.00 (0.99 - 1.01)	0.8854	1.01 (0.99 - 1.02)	0.2954	1.00 (0.99 - 1.01)	0.8929
Age	-	-	1.08 (1.05 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.08 (0.51 - 2.30)	0.8388
QTc Max	1.00 (1.00 - 1.01)	0.3055	1.00 (0.99 - 1.00)	0.2795	1.00 (1.00 - 1.01)	0.3142
Age	-	-	1.08 (1.05 - 1.12)	< 0.0001	-	-
Sex	-	-	-	-	0.98 (0.45 - 2.14)	0.9546
Longest R-R	1.36 (0.75 - 2.46)	0.3145	1.13 (0.57 - 2.23)	0.7193	1.35 (0.75 - 2.45)	0.3158
Age	-	-	1.07 (1.04 - 1.11)	< 0.0001	-	-
Sex	-	-	-	-	1.08 (0.50 - 2.30)	0.8478

Table 26: Normality of HCM, FD with WMLs and FD without WMLs groups and homogeneity of variance results, for patients aged between 40 and 59 (including)

Variables	Normality - pvalues			Homogeneity of variances - pvalues
	HCM	FD with WMLs	FD without WMLs	
Age	p = 0.0439	p = 0.3298	p = 0.0925	-
HR Min	p = 0.2681	p = 0.7946	p = 0.5731	p = 0.2137
HR Mean	p = 0.7843	p = 0.5941	p = 0.2996	p = 0.0811
HR Max	p = 0.9374	p = 0.2568	p = 0.7093	p = 0.0454
ASDNN 5	p = 0.0866	p = 0.0008	p = 0.7723	-
SDANN 5	p = 0.9373	p = 0.4290	p = 0.0482	p = 0.0703
SDNN	p = 0.5008	p = 0.1656	p = 0.0183	-
RMSSD	p = 0.0002	p < 0.0001	p = 0.0693	-
QT Min	p = 0.0709	p = 0.1096	p = 0.6717	p = 0.5820
QT Mean	p = 0.0086	p = 0.3633	p = 0.0057	-
QT Max	p = 0.0247	p = 0.0987	p = 0.0002	-
QTc Min	p = 0.7429	p = 0.0418	p = 0.4230	-
QTc Mean	p = 0.6566	p = 0.4337	p = 0.0105	-
QTc Max	p = 0.2809	p = 0.1030	p = 0.2586	p = 0.0157
QTc > 450	p = 0.0101	p = 0.0001	p = 0.0001	-
Longest R-R	p = 0.0176	p = 0.0001	p = 0.0012	-

Table 27: Statistically non-significant logistic regression results of HCM and FD without WMLs groups with ages between 40 and 59 (including)

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Max	1.03 (1.00 - 1.07)	0.0838	1.01 (0.97 - 1.06)	0.5162	1.03 (0.99 - 1.07)	0.1201
Age	-	-	0.90 (0.78 - 1.04)	0.1478	-	-
Sex	-	-	-	-	3.23 (0.83 - 12.59)	0.0908
SDANN 5	0.98 (0.96 - 1.00)	0.0821	0.97 (0.95 - 1.00)	0.0279	0.98 (0.96 - 1.01)	0.1376
Age	-	-	0.83 (0.72 - 0.97)	0.0155	-	-
Sex	-	-	-	-	3.10 (0.79 - 12.07)	0.1035
RMSSD	0.97 (0.93 - 1.00)	0.0699	0.97 (0.93 - 1.00)	0.0876	0.96 (0.92 - 1.00)	0.0411
Age	-	-	0.90 (0.79 - 1.02)	0.1036	-	-
Sex	-	-	-	-	6.24 (1.24 - 31.48)	0.0265
QT Min	0.97 (0.95 - 1.00)	0.0541	0.98 (0.95 - 1.01)	0.1964	0.98 (0.95 - 1.01)	0.1180
Age	-	-	0.90 (0.79 - 1.03)	0.1192	-	-
Sex	-	-	-	-	2.76 (0.69 - 10.97)	0.1495
QTc Min	1.01 (0.99 - 1.03)	0.4962	1.00 (0.98 - 1.02)	0.8699	1.01 (0.99 - 1.03)	0.3714
Age	-	-	0.88 (0.77 - 1.00)	0.0436	-	-
Sex	-	-	-	-	3.88 (1.00 - 15.08)	0.0503
QTc Mean	0.98 (0.96 - 1.01)	0.1705	0.98 (0.95 - 1.01)	0.2002	0.98 (0.95 - 1.01)	0.1159
Age	-	-	0.88 (0.77 - 0.99)	0.0412	-	-
Sex	-	-	-	-	4.23 (1.04 - 17.11)	0.0432
QTc Max	0.99 (0.98 - 1.00)	0.0633	0.99 (0.98 - 1.00)	0.1299	0.99 (0.98 - 1.00)	0.0254
Age	-	-	0.89 (0.78 - 1.01)	0.0683	-	-
Sex	-	-	-	-	6.66 (1.34 - 33.08)	0.0204
QTc > 450	0.98 (0.96 - 1.00)	0.0873	0.98 (0.96 - 1.00)	0.1169	0.98 (0.96 - 1.00)	0.0783
Age	-	-	0.87 (0.77 - 1.00)	0.0453	-	-
Sex	-	-	-	-	3.96 (0.98 - 16.01)	0.0538
Longest R-R	0.29 (0.09 - 0.98)	0.0470	0.39 (0.11 - 1.42)	0.1542	0.18 (0.04 - 0.78)	0.0218
Age	-	-	0.91 (0.80 - 1.04)	0.1545	-	-
Sex	-	-	-	-	7.11 (1.41 - 36.01)	0.0177

Table 28: Statistically non-significant logistic regression results of HCM and FD with WMLs groups with ages between 40 and 59 (including)

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	1.08 (0.98 - 1.20)	0.1211	1.09 (0.98 - 1.21)	0.1112	1.09 (0.98 - 1.21)	0.1117
Age	-	-	0.90 (0.79 - 1.04)	0.1447	-	-
Sex	-	-	-	-	2.33 (0.60 - 9.09)	0.2230
HR Max	1.03 (0.99 - 1.08)	0.1443	1.03 (0.98 - 1.08)	0.2661	1.04 (0.99 - 1.08)	0.1266
Age	-	-	0.93 (0.81 - 1.07)	0.3016	-	-
Sex	-	-	-	-	2.34 (0.60 - 9.09)	0.2192
ASDNN 5	0.98 (0.95 - 1.01)	0.1700	0.98 (0.95 - 1.01)	0.1918	0.98 (0.95 - 1.01)	0.1626
Age	-	-	0.91 (0.80 - 1.04)	0.1770	-	-
Sex	-	-	-	-	2.26 (0.59 - 8.70)	0.2356
SDANN 5	1.01 (0.99 - 1.03)	0.4467	1.00 (0.98 - 1.02)	0.7280	1.01 (0.99 - 1.03)	0.3240
Age	-	-	0.92 (0.80 - 1.05)	0.2096	-	-
Sex	-	-	-	-	2.48 (0.64 - 9.64)	0.1888
SDNN	0.99 (0.98 - 1.01)	0.5143	0.99 (0.98 - 1.01)	0.4550	1.00 (0.98 - 1.01)	0.5863
Age	-	-	0.91 (0.79 - 1.04)	0.1478	-	-
Sex	-	-	-	-	2.10 (0.56 - 7.81)	0.2685
RMSSD	0.98 (0.96 - 1.00)	0.1247	0.98 (0.96 - 1.01)	0.1595	0.98 (0.96 - 1.00)	0.1073
Age	-	-	0.93 (0.81 - 1.06)	0.2857	-	-
Sex	-	-	-	-	2.76 (0.66 - 11.60)	0.1652
QT Min	0.98 (0.95 - 1.02)	0.3316	0.99 (0.96 - 1.02)	0.4028	0.99 (0.95 - 1.02)	0.3627
Age	-	-	0.91 (0.80 - 1.04)	0.1846	-	-
Sex	-	-	-	-	2.11 (0.56 - 7.90)	0.2681
QTc Min	1.02 (1.00 - 1.05)	0.0953	1.02 (0.99 - 1.05)	0.1258	1.03 (1.00 - 1.06)	0.0707
Age	-	-	0.92 (0.80 - 1.05)	0.2137	-	-
Sex	-	-	-	-	2.72 (0.67 - 11.10)	0.1626
Longest R-R	0.54 (0.22 - 1.33)	0.1813	0.64 (0.25 - 1.65)	0.3586	0.47 (0.18 - 1.19)	0.1120
Age	-	-	0.93 (0.81 - 1.07)	0.3072	-	-
Sex	-	-	-	-	2.85 (0.69 - 11.67)	0.1462

Table 29: Statistically non-significant logistic regression results of FD with WMLs and FD without WMLs groups with ages between 40 and 59 (including)

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Min	0.91 (0.82 - 1.01)	0.0868	0.91 (0.82 - 1.01)	0.0719	0.92 (0.82 - 1.02)	0.1061
Age	-	-	1.07 (0.96 - 1.19)	0.2447	-	-
Sex	-	-	-	-	0.71 (0.21 - 2.42)	0.5897
HR Mean	0.96 (0.89 - 1.03)	0.2188	0.96 (0.90 - 1.04)	0.3075	0.96 (0.89 - 1.03)	0.2901
Age	-	-	1.04 (0.93 - 1.17)	0.4516	-	-
Sex	-	-	-	-	0.73 (0.21 - 2.50)	0.6175
HR Max	0.98 (0.94 - 1.03)	0.4288	0.99 (0.95 - 1.04)	0.7129	0.99 (0.95 - 1.03)	0.5771
Age	-	-	1.05 (0.93 - 1.18)	0.4501	-	-
Sex	-	-	-	-	0.69 (0.20 - 2.39)	0.5542
ASDNN 5	1.03 (0.99 - 1.08)	0.1618	1.04 (0.99 - 1.09)	0.1060	1.03 (0.98 - 1.08)	0.1998
Age	-	-	1.08 (0.97 - 1.21)	0.1736	-	-
Sex	-	-	-	-	0.71 (0.21 - 2.42)	0.5876
RMSSD	1.03 (0.99 - 1.08)	0.1730	1.03 (0.99 - 1.08)	0.1479	1.03 (0.99 - 1.08)	0.1840
Age	-	-	1.07 (0.96 - 1.20)	0.2334	-	-
Sex	-	-	-	-	0.63 (0.19 - 2.11)	0.4571
QT Min	1.02 (0.99 - 1.04)	0.1564	1.02 (0.99 - 1.04)	0.2257	1.02 (0.99 - 1.04)	0.2173
Age	-	-	1.04 (0.93 - 1.16)	0.4904	-	-
Sex	-	-	-	-	0.79 (0.23 - 2.79)	0.7186
QT Mean	1.00 (0.98 - 1.03)	0.6414	1.00 (0.98 - 1.02)	0.8787	1.00 (0.98 - 1.02)	0.7580
Age	-	-	1.06 (0.94 - 1.18)	0.3520	-	-
Sex	-	-	-	-	0.64 (0.19 - 2.11)	0.4591
QT Max	0.99 (0.99 - 1.00)	0.3038	0.99 (0.98 - 1.00)	0.1745	0.99 (0.98 - 1.00)	0.2801
Age	-	-	1.08 (0.97 - 1.22)	0.1680	-	-
Sex	-	-	-	-	0.58 (0.18 - 1.94)	0.3785
QTc Min	1.01 (0.99 - 1.03)	0.4166	1.01 (0.99 - 1.03)	0.3701	1.01 (0.99 - 1.03)	0.3818
Age	-	-	1.06 (0.95 - 1.19)	0.2755	-	-
Sex	-	-	-	-	0.59 (0.18 - 1.93)	0.3797
QTc Mean	0.99 (0.97 - 1.02)	0.5175	0.99 (0.96 - 1.02)	0.4054	0.99 (0.96 - 1.02)	0.5227
Age	-	-	1.07 (0.96 - 1.19)	0.2472	-	-
Sex	-	-	-	-	0.61 (0.19 - 2.00)	0.4167
QTc Max	0.99 (0.98 - 1.00)	0.2440	0.99 (0.98 - 1.00)	0.1650	0.99 (0.98 - 1.01)	0.3228
Age	-	-	1.08 (0.96 - 1.21)	0.2025	-	-
Sex	-	-	-	-	0.72 (0.21 - 2.48)	0.6048
QTc > 450	0.99 (0.97 - 1.01)	0.4543	0.99 (0.96 - 1.01)	0.2947	0.99 (0.97 - 1.01)	0.3523
Age	-	-	1.08 (0.96 - 1.21)	0.2059	-	-
Sex	-	-	-	-	0.54 (0.16 - 1.83)	0.3213
Longest R-R	1.76 (0.56 - 5.46)	0.3311	1.64 (0.52 - 5.19)	0.3977	1.87 (0.59 - 5.97)	0.2881
Age	-	-	1.05 (0.94 - 1.17)	0.3669	-	-
Sex	-	-	-	-	0.56 (0.17 - 1.87)	0.3500

Table 30: Normality of HCM and FD with WMLs groups and homogeneity of variance results, for patients aged above 59

Variables	Normality - pvalues		Homogeneity of variances - pvalues
	HCM	FD with WMLs	
Age	p = 0.0355	p = 0.0185	-
HR Min	p = 0.2937	p = 0.3408	p = 0.3538
HR Mean	p = 0.4186	p = 0.6448	p = 0.1068
HR Max	p = 0.0247	p = 0.4152	-
ASDNN 5	p = 0.0001	p = 0.0107	-
SDANN 5	p = 0.5551	p = 0.3493	p = 0.5958
SDNN	p = 0.0003	p = 0.2821	-
RMSSD	p < 0.0001	p = 0.0001	-
QT Min	p = 0.3854	p = 0.0342	-
QT Mean	p = 0.3932	p = 0.8572	p = 0.1415
QT Max	p = 0.0002	p = 0.0007	-
QTc Min	p = 0.9707	p = 0.1330	p = 0.5846
QTc Mean	p = 0.7889	p = 0.8998	p = 0.8367
QTc Max	p = 0.0041	p = 0.0089	-
QTc > 450	p = 0.0219	p = 0.0060	-
Longest R-R	p < 0.0001	p < 0.0001	-

Table 31: Statistically non-significant logistic regression results of HCM and FD with WMLs groups with ages above 59

Variables	Univariate logistic regression		Multivariable logistic regression - Adjusted to age		Multivariable logistic regression - Adjusted to sex	
	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value	OR (95% C.I.)	p-value
HR Max	1.03 (0.99 - 1.06)	0.1342	1.03 (0.99 - 1.06)	0.1333	1.03 (0.99 - 1.06)	0.1281
Age	-	-	0.99 (0.92 - 1.07)	0.7602	-	-
Sex	-	-	-	-	3.94 (1.12 - 13.81)	0.0323
ASDNN 5	1.00 (0.98 - 1.02)	0.9291	1.00 (0.98 - 1.02)	0.9158	1.00 (0.98 - 1.02)	0.9507
Age	-	-	0.99 (0.92 - 1.06)	0.7775	-	-
Sex	-	-	-	-	3.74 (1.10 - 12.67)	0.0340
SDANN 5	1.01 (0.99 - 1.03)	0.4996	1.01 (0.99 - 1.03)	0.5088	1.02 (0.99 - 1.04)	0.1653
Age	-	-	0.99 (0.92 - 1.07)	0.8122	-	-
Sex	-	-	-	-	5.18 (1.34 - 20.04)	0.0172
SDNN	1.00 (0.98 - 1.01)	0.5939	1.00 (0.98 - 1.01)	0.5699	1.00 (0.98 - 1.01)	0.6474
Age	-	-	0.99 (0.92 - 1.06)	0.7352	-	-
Sex	-	-	-	-	3.71 (1.09 - 12.61)	0.0354
RMSSD	1.00 (0.99 - 1.01)	0.7086	1.00 (0.99 - 1.01)	0.7080	1.00 (0.99 - 1.00)	0.4582
Age	-	-	0.99 (0.92 - 1.07)	0.7820	-	-
Sex	-	-	-	-	4.06 (1.16 - 14.25)	0.0285
QT Max	1.00 (0.99 - 1.00)	0.1248	1.00 (0.99 - 1.00)	0.1227	1.00 (0.99 - 1.00)	0.1391
Age	-	-	0.99 (0.91 - 1.06)	0.7215	-	-
Sex	-	-	-	-	3.93 (1.11 - 13.98)	0.0344
QTc Min	1.00 (0.98 - 1.01)	0.4852	1.00 (0.98 - 1.01)	0.4951	1.00 (0.98 - 1.01)	0.5487
Age	-	-	0.99 (0.92 - 1.07)	0.8165	-	-
Sex	-	-	-	-	3.70 (1.09 - 12.57)	0.0363
QTc Mean	1.00 (0.98 - 1.02)	0.8502	1.00 (0.98 - 1.02)	0.8713	1.00 (0.98 - 1.02)	0.7139
Age	-	-	0.99 (0.92 - 1.07)	0.7956	-	-
Sex	-	-	-	-	3.82 (1.12 - 13.08)	0.0325
QTc Max	1.00 (0.99 - 1.00)	0.5775	1.00 (0.99 - 1.00)	0.5660	1.00 (0.99 - 1.00)	0.4064
Age	-	-	0.99 (0.92 - 1.06)	0.7564	-	-
Sex	-	-	-	-	4.07 (1.16 - 14.24)	0.0283
QTc > 450	1.00 (0.99 - 1.02)	0.5860	1.00 (0.99 - 1.02)	0.6058	1.01 (0.99 - 1.03)	0.3378
Age	-	-	0.99 (0.92 - 1.07)	0.8306	-	-
Sex	-	-	-	-	4.26 (1.19 - 15.23)	0.0256
Longest R-R	0.87 (0.51 - 1.49)	0.6124	0.88 (0.51 - 1.50)	0.6311	0.78 (0.44 - 1.40)	0.4109
Age	-	-	0.99 (0.92 - 1.07)	0.8225	-	-
Sex	-	-	-	-	4.08 (1.16 - 14.34)	0.0283

Appendix B

Machine Learning Appendix

Table 32: Hyperparameters values/status per machine learning model and classification problem for patients aged between 40 and 59 (inclusive)

ML Model	Model Parameters	Classification Problem		
		HCM vs FD w/o WMLs	HCM vs FD w/ WMLs	FD w/ WMLs vs FD w/o WMLs
LR	C	0.5	0.05	0.7
Linear SVM	C	0.1	0.05	3
RF	bootstrap	False	True (default)	False
	n_estimators	2	24	1
	max_depth	None (default)	6	5
	max_leaf_nodes	None (default)	None (default)	9
	min_samples_leaf	1 (default)	1 (default)	1 (default)
	min_samples_split	2 (default)	2 (default)	2 (default)

ML: machine learning; LR: logistic regression; SVM: support vector machine; RF: random forest

Table 33: Accuracy values per machine learning model and classification problem for patients aged between 40 and 59 (inclusive)

ML Model	Classification Problem		
	HCM vs FD w/o WMLs	HCM vs FD w/ WMLs	FD w/ WMLs vs FD w/o WMLs
LR	0.7561	0.6923	0.6250
Linear SVM	0.7317	0.6923	0.6875
RF	0.7561	0.7692	0.8125

ML: machine learning; LR: logistic regression; SVM: support vector machine; RF: random forest

Table 34: Hyperparameters values/status per machine learning model and number of features for HCM and FD without WMLs patients aged 40 to 59 (inclusive)

ML Model	Model Parameters	Number of features (Features)		
		4 Features (HR Mean, SDANN 5, RMSSD, and QTc > 450)	6 Features (HR Mean, SDANN 5, RMSSD, QTMin, QTc Max, and QTc > 450)	9 Features (All)
LR	C	2	0.8	0.5
Linear SVM	C	0.3	0.07	0.1
RBF SVM	C gamma	70 0.002	3 0.03	3 0.04
RF	bootstrap	True (default)	True(default)	False
	n_estimators	40	31	2
	max_depth	3	5	None (default)
	max_leaf_nodes	None (default)	None (default)	None (default)
	min_samples_leaf	1(default)	1 (default)	1 (default)
	min_samples_split	3	2 (default)	2 (default)
KNN	n_neighbors	8	3	8
	weights	uniform	uniform	uniform

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

Table 35: Hyperparameters values/status per machine learning model and number of features for HCM and FD with WMLs patients aged 40 to 59 (inclusive)

ML Model	Model Parameters	Number of features (Features)		
		5 Features (HR Mean, RMSSD, QTc Min, QTcMax and QTc > 450)	6 Features (HR Mean, HR Max, RMSSD, QTcMin, QTc Max and QTc > 450)	9 Features (All)
LR	C	0.06	0.4	0.05
Linear SVM	C	0.9	0.08	0.05
RBF SVM	C	500	30	100
	gamma	0.009	0.02	0.02
RF	bootstrap	False	False	True (default)
	n_estimators	900	40	24
	max_depth	5	8	6
	max_leaf_nodes	None (default)	None (default)	None (default)
	min_samples_leaf	1 (default)	1 (default)	1 (default)
	min_samples_split	2 (default)	0.3	2 (default)
KNN	n_neighbors	6	4	6
	weights	uniform	uniform	uniform

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

Table 36: Hyperparameters values/status per machine learning model and number of features for FD with WMLs and FD without WMLs patients aged 40 to 59 (inclusive)

ML Model	Model Parameters	Number of features (Features)		
		1 Feature (SDANN 5)	2 Features (SDANN 5 and QTc Min)	9 Features (All)
LR	C	6	20	0.7
Linear SVM	C	0.2	10	3
RBF SVM	C	9	900	30
	gamma	0.3	0.007	0.007
RF	bootstrap	True (default)	False	False
	n_estimators	50	46	1
	max_depth	2	1	5
	max_leaf_nodes	None (default)	None (default)	9
	min_samples_leaf	5	10	1 (default)
	min_samples_split	2 (default)	2 (default)	2 (default)
KNN	n_neighbors	16	8	2
	weights	uniform	uniform	uniform

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour

Table 37: Hyperparameters values/status per machine learning model and classification problem for patients aged above 59

ML Model	Model Parameters	Classification Problem
		HCM vs FD w/ WMLs
LR	C	1
Linear SVM	C	1.3
RF	bootstrap	False
	n_estimators	22
	max_depth	3
	max_leaf_nodes	None (default)
	min_samples_leaf	2
	min_samples_split	2 (default)

ML: machine learning; LR: logistic regression; SVM: support vector machine; RF: random forest

Table 38: Accuracy values per machine learning model and classification problem for patients aged above 59

ML Model	Classification Problem
	HCM vs FD w/ WMLs
LR	0.7143
Linear SVM	0.7755
RF	0.7959

ML: machine learning; LR: logistic regression; SVM: support vector machine; RF: random forest

Table 39: Hyperparameters values/status per machine learning model and number of features for HCM and FD with WMLs patients aged above 59

ML Model	Model Parameters	Number of features (Features)		
		5 Features (HR Min, SDANN 5, QT Min, QT Mean, and QTc > 450)	9 Features (HR Min, HR Max, ASDNN 5, SDANN5, QT Min, QT Mean, QTc Min, QTc >450, and Longest R-R)	10 Features (All)
LR	C	0.3	1.2	1 (default)
Linear SVM	C	20	0.9	1.3
RBF SVM	C	100	30	500
	gamma	0.02	0.008	0.001
RF	bootstrap	True (default)	False	False
	n_estimators	9	580	22
	max_depth	None (default)	None (default)	3
	max_leaf_nodes	None (default)	None (default)	None (default)
	min_samples_leaf	1 (default)	1 (default)	2
KNN	min_samples_split	2 (default)	2 (default)	2 (default)
	n_neighbors	7	18	2
	weights	distance	uniform	distance

ML: machine learning; LR: logistic regression; SVM: support vector machine; RBF: radial basis function; RF: random forest; KNN: K-Nearest Neighbour