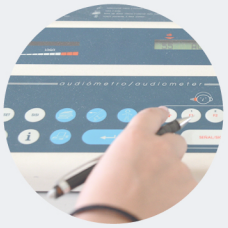
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA BIOMÉDICA






Caracterização Fenotípica do Sistema Visual em Modelos Animais de Doenças do Desenvolvimento Neurológico

INÊS JOSÉ BORGES PINTO VIGO
Novembro de 2022

P.PORTO

Politécnico do Porto
Instituto Superior de Engenharia do Porto

Phenotypic Characterization of the Visual System in Animal Models of Neurodevelopmental Diseases

Inês Borges Vigo

Master in Biomedical Engineering

Departamento de Física
Instituto Superior de Engenharia do Porto

November, 2022

This dissertation partially satisfies the requirements of the Thesis/Dissertation course of the program Master in Biomedical Engineering,

.

Candidate: Inês Borges Vigo, No. 1170969, 1170969@isep.ipp.pt

Scientific Guidance: João Martins, Post-DocatCIBIT,ICNAS

Scientific Co-Guidance: Luís Coelho, ProfessoratISEP

Departamento de Física
Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino de Almeida, 431, 4200–072 Porto

November, 2022

"Success is not the key to happiness. Happiness is the key to success. If you love what you are doing, you will be successful."

Albert Schweitzer

Acknowledgements

No project or goal can be achieved alone, we all need people who together help us to build our dreams. The realization of this master's thesis is no exception, it was only possible thanks to the essential support of many people. It is to all these people that I dedicate this work, as well as my most sincere thanks.

I thank, first of all, my mentors Engineer Luís Coelho and Dr. João Martins, for all the knowledge sharing and guidance during the elaboration of this study.

I thank, the Instituto Superior de Engenharia do Porto for having been the home that welcomed me throughout my academic career, as well as all the teachers and colleagues with whom I crossed paths in this magnificent institution.

I would also like to thank CIBIT (Coimbra Institute for Biomedical Imaging and Translational Research) for enabling this research, and its professionals for contributing to it.

I thank my friends, for all their emotional support and for always believing in me, encouraging me to do more and better every day.

Finally, I thank my family, especially my mother, sisters, godparents and boyfriend for being my first structure and source of support, being my example and safe harbour, without them none of this would be possible.

To all, my sincere gratitude..

Abstract

Neurofibromatosis Type 1 (NF1) is an inherited disease associated with neurodevelopmental disorders, including *Autism spectrum disorder* (ASD). One of the systems affected by this disease is the visual system, although it is little studied and little taken into consideration in the diagnosis of NF1. It is in this sense that the present investigation arises, where potential changes in retinal physiology are studied in a genetic model of NF1, using a non-invasive neurophysiological technique - the *Electroretinogram* (ERG) - to determine its diagnostic potential.

Thus, the main goal of this work was to find an accurate way to distinguish individuals with NF1 from healthy individuals. To this end, we evaluated possible features of the ERG signals, for both genotypes, that could demonstrate differences between them, and therefore be able to differentiate them. In general, the aim of this study was to find an algorithm capable of classifying an individual as sick or healthy with the greatest possible accuracy. To achieve this goal, a Python program was created that, using statistical analysis and *Machine Learning* (ML) techniques, was capable of returning a complete analysis of the data and the most relevant characteristics, as well as estimating the classification of the data in terms of accuracy value.

With this methodology, it was possible to conclude that the use of ERG in the diagnosis of NF1 in an individual is feasible, since the statistical analysis found two features with significant p-values and the ML classification obtained interesting accuracy values.

Keywords: Autism, Neurofibromatosis Type 1, Electroretinogram, Classification, Statistical Analysis, Machine Learning, Python.

Resumo

A neurofibromatose tipo 1 (NF1) é uma doença hereditária associada a perturbações do desenvolvimento neurológico, incluindo a doença do espectro do autismo (ASD). Um dos sistemas afetados por esta doença é o sistema visual, embora seja pouco estudado e pouco tido em consideração no diagnóstico da NF1. É neste sentido que surge a presente investigação, onde são estudadas as potenciais alterações na fisiologia da retina num modelo genético de murganho de NF1, utilizando uma técnica neurofisiológica não invasiva - o eletrorretinograma (ERG) - para determinar o seu potencial diagnóstico.

Assim, o principal objetivo deste trabalho passou por encontrar uma forma precisa de distinguir indivíduos com NF1 de indivíduos saudáveis. Para tal, foram avaliadas possíveis características dos sinais ERG, para os dois genótipos, que demonstrassem diferenças entre estes, e, por conseguinte, fossem capazes de os diferenciar. De forma geral, pretendeu-se com este estudo encontrar um algoritmo capaz de realizar a classificação de um indivíduo como doente ou saudável com a maior precisão possível. Para atingir este fim, foi então criado um programa em Python que, recorrendo a técnicas de Análise Estatística e de Aprendizagem de Máquina, fosse capaz de retornar uma análise completa dos dados e das características com maior relevância, assim como estimar a classificação dos dados em termos de valor de *accuracy*.

Com esta metodologia foi, então, possível concluir que a utilização do ERG no diagnóstico da NF1 num indivíduo é viável, uma vez que, pela análise estatística foram encontradas duas características com p-values significativos e pela classificação via Aprendizagem de Máquina se obtiveram valores de *accuracy* interessantes.

Palavras-Chave: Autismo, Neurofibromatose Tipo 1, Eletrorretinograma, Classificação, Análise Estatística, Aprendizagem de Máquina, Python.

Contents

List of Figures

List of Tables

List of Acronyms

ANOVA	*Analysis of Variance*
ASD	*Autism spectrum disorder*
CNS	*Central Nervous System*
CSV	*Comma-separated values*
CV	*Cross Validation*
CV	*Critical Value*
DF	*Degrees of Freedom*
DSP	*Digital Signal Processing*
EOG	*Electrooculography*
ERG	*Electroretinogram*
FA	*Fourier analysis*
fERG	*Focal Flash ERG*
ffERG	*Full-field Electroretinography*
FFT	*Fast Fourier Transform*
FS	*Frequency Sampling*
HSD	*Tukey Honestly Significant Difference*
ISCEV	*International Society of Clinical*
LR	*Logisitc Regression*
mfERG	*Multifocal Electroretinogram*
mfVEP	*Multifocal Visual Evoked Potential*
ML	*Machine Learning*
NB	*Naive Bayes*

NF1	*Neurofibromatosis Type 1*
OP	*Oscillatory Potential*
PCA	*Principal Component Analysis*
PERG	*Pattern Electroretinogram*
PNS	*Peripheral Nervous System*
RF	*Random Forest*
SVM	*Support Vector Machine*
T-Stat	*T-Statistic*
t-test	*Student's t-test*
VEP	*Visual Evoked Potential*
WA	*Wavelet Analysis*
WT	*Wild Type*

Chapter 1

Introduction

1.1 Definition of the Problem

Autism spectrum disorder (ASD) is a neurodevelopmental illness marked by communication and social interaction difficulties, as well as restricted and repetitive behavioural patterns and interests [1]. It is still a complex condition with little understanding of its cause and therapy. The prevalence of ASD is estimated to be between 0.3 and 0.9 percent. The cause has yet to be determined, but genetic factors are expected to play a significant influence [2].

Although the aforementioned characteristics remain the most important diagnostic markers of ASD, the existence of sensory symptoms has recently been given a specific importance, in line with studies of autism-related individual differences in visual perception. Furthermore, more than 90% of people with ASD have hyper- and/or hyposensitive responses to particular stimuli, which can lead to sensory overload [3].

In this sense, the scientific community has sought to study and find ways to evaluate the impact of these stimuli on individuals with ASD, and it is also with this purpose that this research arises, where, with the application of software and computational tools, we intend to evaluate and analyze automatically the response to various intensities of light to which laboratory mice with the disease are exposed. In this study, mice were used since animal models are similar enough to humans in their anatomy, physiology or response to a pathogen that researchers can extrapolate the results of animal model studies to better understand human physiology and disease.

1.2 Background and Objectives

This dissertation is part of the Master's Degree in Biomedical Engineering, of the academic year 2021/2022, and its main objective is the phenotypic characterization of the visual system in animal models associated with neurodevelopmental disorders. The aim of this study is to evaluate potential changes in retinal physiology in a genetic model of NF1 mouse, using a non-invasive neurophysiological technique, the electroretinogram (ERG), to determine its diagnostic potential. In this sense, it is intended to develop computer tools for signal processing, waveform analysis, feature extraction, and classification.The choice of this methodology is due to the fact that the animal model allows dispensing with the human testing, ensuring similar anatomy and physiology, besides the fact that the ERG is a method that allows acquiring a large number of data efficiently and staging different environments and stimuli. As for classification, it is believed that with the advances in Machine Learning techniques interesting results can be achieved.

Thus, and completing the aforementioned objectives, it is also intended to:

- To exploit a wide range of stimuli, that is, to enable the analysis of any intensity used in the acquisition of the ERGs.
- Improve the feature extraction methods, as well as increase and vary the set of features to be extracted.
- To explore machine learning algorithms for the classification of NF1 and WT genotypes.

1.3 Dissertation Structure

This dissertation is divided into seven chapters that follow the logical way in which it was prepared and thought out. This first chapter presents a brief summary of the content of this work. Chapter 2 presents the theoretical foundations deemed important for the understanding and interpretation of the theme explored. This is then followed by chapter 3, where the state of the art is presented, where the key ideas of the various studies carried out within the same scope as the present research are presented. Chapter 4 presents the materials and methods used for the execution of this research. In turn, chapter 5 describes the whole process of algorithm development, whose results and discussion are presented in chapter 6. Finally, there is chapter 7 where the final considerations are presented, as well as some proposals for future improvements.

Chapter 2

Theoretical Bases

To carry out any research or study, it is necessary to acquire a priori theoretical knowledge so that the necessary guidelines can be outlined for the execution of the same in order to achieve the stipulated objectives. Thus, this section presents the main scientific foundations for this dissertation.

2.1 Autism Spectrum Disorder

Autism spectrum disorder (ASD) is an overarching concept for neurodevelopmental disabilities that affect social and communication skills, and its prevalence has been increasing in recent years [4]. According to the World Health Organization, the international prevalence of ASD is 0.76 percent, which equates to approximately 16 percent of the global child population [5].

ASD has progressed from a narrowly defined and rare childhood onset disorder to a well-publicized, advocated, and researched life condition that is now recognized as quite common and very heterogeneous [6]. The main characteristics of ASD, which are social communication deficits and repetitive and unusual sensory-motor behaviors, have not changed significantly since its initial definition. However, autism is now recognized as a spectrum disorder with symptoms ranging from mild to severe [7].

Indeed, autism can be viewed as a common and early behavioral constellation of social and communicative atypicality associated with apparently limited interests and repetitive behaviors, as well as an uneven ability profile. Over the last 75 years,

its definition has steadily broadened, introducing phenotypes that are increasingly distant from its initial description, heterogeneous in intelligence and speech, and associated conditions. When it is not associated with other conditions, its cause is primarily genetic, transmissible, and favored by frequent polymorphisms with minor effects found in the general population [8].

Autism is a neurobiological disorder influenced by genetic and environmental factors that affect the developing brain. Although research continues to uncover factors associated with ASD risk, and these findings may guide further aetiological investigation, no final causal pathway has been identified [5].

In fact, despite advances in our understanding of the neurobiology and genetics of ASD, diagnosis is still based on the identification and reporting of behaviorally defined clinical symptoms [9]. Several researchers, however, have focused on new methods of diagnosis, specifically visual markers.

2.1.1 Clinical manifestations of the visual system

Given the extent of the neurological underpinnings of the visual system, it is not surprising that a condition with widespread differences in brain function would also significantly affect visual function. Indeed, individuals with autism frequently struggle to process and integrate sensory information, including visual information [10].

Several studies have been conducted to assess the effects of ASD on the visual system, and it has been discovered that children with ASD have a higher incidence of refractive error, but the distribution of refractive error does not show a clear trend. Individuals with ASD have differences in eye motility function when compared to their typical peers, including atypical optokinetic reflexes at higher target speeds [11], indicating that eye movement impairment is one of the significant clinical features [12].

Individuals with ASD have less control of their saccades, which results in more saccades occurring in response to stimuli and hypometric saccades [13]. Also, regardless of the diagnostic category, eye contact, eye abnormalities, and face recognition are types of behavior that characterize individuals with ASD and have been linked to eye movement impairments [12].

In contrast sensitivity research, children with ASD performed worse in the target pattern but not in flicker detection. There have been few studies that directly address color performance in ASD, however, based on the existing results, it can be stated that, in general, color perception is poor in autism [14].

Other studies have found that children with autism have "locally oriented" perception and low-level enhanced functioning. In enhanced perceptual function, their abilities in processing three-dimensional images tend to prioritise detailed information from short-exposure stimuli or long-exposure stimuli [15]. Individuals with ASD outperform in visual tasks requiring attention to local details, but perform poorly

in holistic-oriented tasks, according to these studies. Because detail-oriented and holistic tasks necessitate the use of high and low spatial frequency information, respectively, the'spatial frequency hypothesis' in ASD proposes that the detail-oriented tendency in ASD reflects increased sensitivity to high spatial frequencies and/or decreased sensitivity to low spatial frequencies [16].

Individuals with ASD have a sensitivity to light, particularly bright light, that spans all age groups and genders. When stimulated by a flash of light, the pupil goes through a distinct contraction and recovery process known as the pupillary light reflex. The atypical pupillary light reflex has been observed in people with autism [17].

Thus, the visual clinical manifestations found in individuals with ASD are ubiquitous, multimodal and often severe in their intensity, because the symptoms of the disease are biologically based and are linked to physiological differences in the processing of visual and other sensory information.

2.1.2 Neurofibromatosis type 1

Neurofibromatosis type 1 (NF1), also known as von Recklinghausen disease [18], is an inherited disease that affects one in every 3000 people worldwide, regardless of gender or ethnicity [19]. People with this condition are more likely to develop benign and malignant *Central Nervous System* (CNS) and *Peripheral Nervous System* (PNS) tumors, as well as malignant diseases affecting other parts of the body [20].

NF1 is caused by a mutation in the NF1 tumor suppressor gene, which is located on the long arm of chromosome 17 and codes for a cytoplasmic protein known as neurofibromin[18]. This protein functions as a negative regulator of the Ras proto-oncogene and is found primarily in neurons, Schwann cells, oligodendrocytes, and astrocytes. This protein does, in fact, influence multiple signaling pathways that affect numerous cellular processes throughout the body[21].

Because of its highly variable characteristics, early diagnosis of NF1 is difficult. Some people may be mildly affected with few symptoms, while others may be severely affected. NF1 is typically diagnosed through clinical evaluation, which includes a thorough history and physical examination [22]. The National Institutes of Health considers two or more of the clinical features shown in Figure 2.1 to be sufficient for the diagnosis of NF1 [21, 20]:

Café-au-lait spots are symmetrical flat areas of skin over-pigmentation with a brownish color and rounded edges that appear from birth. These patches are the first clinical manifestations of NF1, and they tend to grow in size and number throughout childhood and puberty. In adulthood, these spots are present in approximately 95% of NF1 patients. Furthermore, 70% of NF1 patients have freckles in the intertriginous areas of the axilla and in the inguinal region [24].

	NIH criteria for the diagnosis of neurofibromatosis type 1
1	Six or more cafe-au-lait skin macules >5 mm in prepubertal individuals and >15 mm in postpubertal individuals
2	Two or more neurofibromas of any type or one plexiform neurofibroma
3	Axillary or inguinal freckling
4	Two or more Lisch nodules
5	Optic glioma
6	Bone lesion with sphenoid dysplasia or thinning of the long bone cortex with or without pseudarthrosis
7	A first-degree relative (parent, sibling, or offspring) that meets NIH criteria

*The diagnosis of NF1 requires at least two of the seven NIH criteria.

Figure 2.1: National Institutes of Health Diagnostic Criteria for NF1 [23].

The most common ophthalmological feature of NF1 is Lisch nodules, which are haematomas of the iris pigment epithelium. However, they are not pathognomonic. They appear as a cluster of pale, yellow-brown, oval to round, dome-shaped papules protruding from the iris surface [25].

Optic pathway gliomas, which affect about 15-20% of NF1 patients, are typically low-grade astrocytomas that can develop in the optic nerve, optic chiasm, optic tract, and hypothalamus[26].

Approximately 50% of patients meet the diagnostic criteria for NF1 by the age of one year, 97% meet the criteria by the age of eight years, and nearly all patients meet the criteria by the age of twenty years [27].

Genetic tests are available in addition to clinical evaluation, but they are not recommended in most cases. It is not a required component of diagnosis due to the cost of the test and the accuracy of diagnosis through clinical assessment [22]. Some skin findings, however, are not always easily visible and may necessitate the use of ultraviolet light to identify. To analyze the clinical features of the eye, additional ophthalmological examinations may be performed [28].

2.2 Anatomophysiology of the eye

The eye is frequently compared to a camera because it acts like a self-sharpening machine, with the ability to automatically adjust to light intensity, a self-cleaning lens, and powers the brain, which is a computer with extremely advanced parallel processing capabilities [29].

Both the human and mouse eyes are vertebrate eyes, which are a common feature of mammals in general. This organ contains several structures that are essential for us to see and perceive the world around us, including the cornea, iris, pupil, retina, and optic nerve.

The cornea is a clear tissue with significant refractive and barrier functions, with the epithelium acting as the primary barrier to fluids and pathogens. This is further supported by the basal membrane and Bowman's layer, which help to keep the stroma dehydrated. The stroma accounts for the majority of the volume of the cornea, providing support and clarity as well as aiding in ocular immunity. The posterior cornea, which is made up of Descemet's membrane and endothelium, is necessary for stromal dehydration [30].

The iris is a coloured circular muscle whose pigments determine the color of the eyes. This circular muscle regulates the size of the pupil, allowing more or less light to enter the eye depending on the circumstances. Eye color, or more precisely, iris color, is caused by varying amounts of eumelanin, brown/black melanins, and phaeomelanin, red/yellow melanins, which are produced by melanocytes and are thus primarily responsible for the eye color we have [31].

The retina is the third and innermost layer of the eye. This layer, which is located at the focal plane of the eye's optical system, is in charge of converting relevant image information from the external environment into neural impulses that are transmitted to the brain for decoding and analysis. It has two primary layers: an inner neurosensory retina and an outer simple epithelium known as the retinal pigment epithelium. Bruch's membrane surrounds the retina on the outside and the vitreous on the inside. It connects to the optic nerve posteriorly, where axons from ganglion cells of the eye exit [32].

Our most important cranial nerve is the optic nerve. It is divided into four segments as it travels from the eyeball to the brain: intraocular, intraorbital, intracanalicular, and intracranial. The axons of the optic nerve are formed by nearly 1.2 million ganglion cells that line the inner retina. These flow to the lamina cribrosa, where they merge to form the optic papilla. They form the optic nerve and transport action potentials to the lateral geniculate body via the optic chiasm and optic tract, where they are processed by the brain [33].

In Figure 2.2, these elements are shown together with the other components of the eye of a mammal. In the case of this image, it is the human eye in an almost axial section.

When light enters the eye, it travels along a pathway that converts it into electrochemical impulses that the brain can process and interpret. As this light enters the eye, the first structure it encounters is the cornea, which refracts the light, causing it to converge towards the inside of the eye, where it meets the iris and pupil. The optical structures do not always react in the same way; in fact, depending on the intensity and availability of light, the iris will contract or expand to adjust the size of the pupil, thereby regulating the amount of light that can enter the eye. Light enters the pupil and is received by the lens, which, with the assistance of other surrounding muscles, can change shape and bring objects of varying distances into focus through

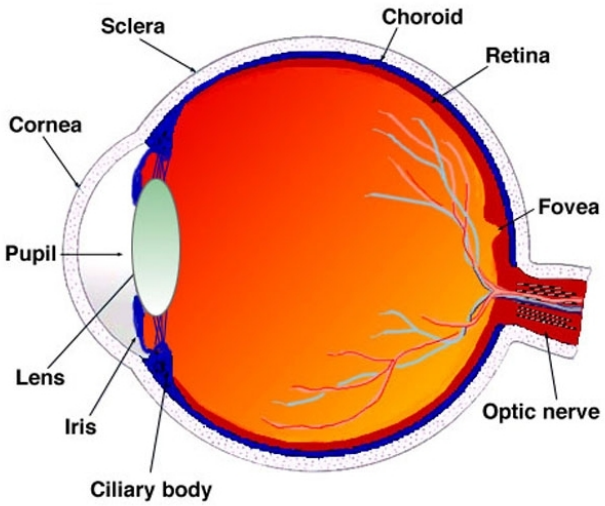


Figure 2.2: Quasi-axial (quasi-horizontal) section of the adult human eye [31].

a process known as accommodation. This lens is also in charge of enhancing the image created and refined by the cornea at the start of the process before projecting it onto the retina. The retina, in turn, collects light via photoreceptors and pigmented epithelial cells. The photopigmenting molecules in these photoreceptors absorb light, causing an electrical signal to change. This conversion of energy into electrical impulses starts a chain reaction of signals that travels through the neurons of the retina to the optic nerve and then to the brain. The brain then receives and processes these signals as perceived images [34].

2.3 Electroretinogram

Electroretinography, which generates an ERG, can be used to study normal retinal function in both humans and animals. The rod and cone systems of the retina can be examined separately with appropriate stimulus selection, and specific levels of processing can be assessed. ERGs are currently useful for non-invasive, in vivo assessment of retinal function in physiologically intact animals, including humans, and can thus supplement functional information obtained from other tests. Indeed, ERGs are useful for non-invasive tracking of the development and loss of retinal function as retinas degenerate and modify their circuits as a result of acquired diseases and genetic mutations in humans and animal models [35].

Electroretinography is a test that measures the electrical responses of different types of cells in the retina, such as photoreceptors (rods and cones), inner retinal cells like bipolar and amacrine cells, and ganglion cells [36]. ERG measurements can be taken from the corneal surface using a contact lens embedded with a corneal ring

electrode and a conjunctival reference, or from the periorbital skin using small blade electrodes attached to the lateral canthi. The standard ERG stimulus is a high-intensity light stimulus that can be modulated to optimize rod or cone response evaluation. Because cones have a shorter refractory period, a higher frequency stimulus, such as 30 Hz, can be used to assess their function. Rods, on the other hand, have a very low response at frequencies above 20 Hz [37].

The ERG is typically recorded with the patient's pupils dilated following the administration of mydriatics. This aids in ensuring uniform retinal illumination and preventing pupillary responses from influencing the nature of ERG tracking. The ERG can also be recorded after light adaptation (photopic) to reduce rod effects or dark adaptation (scotopic) to increase rod contribution [38].

In most literature, ERG waveforms are described as negative when the trace falls below the baseline and positive when they cross above it. As shown in Figure 2.3, the ERG is typically analysed by evaluating the first negative deflection, referred to as the a-wave, and the subsequent positive deflection, referred to as the b-wave.

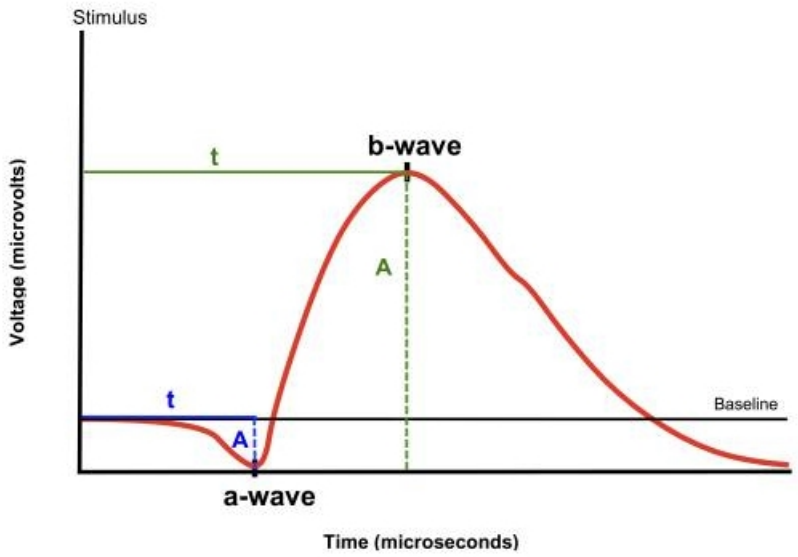


Figure 2.3: Illustration of the characteristic waveform of an ERG. A, amplitude; t, implicit time [39].

A large amplitude positive wave may precede the a-wave in photopic recordings, particularly when a bright flash is used as a stimulus. The a-wave represents photoreceptor mass depolarization, while the b-wave represents the activation of retinal interneurons from the inner nuclear layer [40]. In photopic ERGs, the a-wave is observed at about 15 msec and the b-wave at about 30 to 35 msec in normal subjects. The implicit times, commonly referred to as latency, in photopic recordings are generally shorter, i.e. faster, but the amplitude of the response is generally smaller than in dark-adapted ERGs [41]. The c and d waves can also be described, which are

attributed to the retinal pigment epithelium and bipolar cells, respectively, however, it is not common for these to be evaluated in the literature [42].

Aside from the presence of a- and b-waves, a normal ERG also contains other peaks caused by *Oscillatory Potential* (OP)s. The OP is a high frequency, low amplitude transretinal light field potential that overlaps with the rising phase of the ERG's b-wave. OPs are typically triggered by a single, brief flash and are thought to be primarily caused by light activity in the inner retina, specifically neural interactions between bipolar cells, amacrine cells, and ganglion cells [43].

A summary of these components as well as their behaviour and origin in the ERG is presented in Table 2.1.

Table 2.1: Resume of Photopic ERG Components [44].

ERG Components	**Origin**	**Component Feature**	**Existence Time**	**Frequency Band**
a-wave	Cones and Rods	Negative First Trough	Early Response	Wide Band
b-wave	ON Bipolar Cells and Muller Cells	Positive Highest Peak	Early Response	Wide Band
OPs	Amacrine Cells	Low amplitude Oscillations Across the ERG	Exists Across Entire ERG	Narrow Band

2.3.1 Full-Field ERG

Full-field Electroretinography (ffERG) or conventional electroretinography measures the electrical response of the entire retina to intermittent light from an arc basin (Ganzfeld's stimulation arc) scattered into the eye through a dilated pupil [45].

The ffERG is recorded in the cornea and triggered by a full-field stimulus, resulting in a mass response generated by cells throughout the retina. To obtain reproducible amplitudes and implicit timing of responses, the stimulus and background light should be homogeneous and cover the entire retina, so that all receptors are stimulated or adapted in a relatively homogeneous manner [46]. This same stimulus is reproduced by the full-field stimulator, or Ganzfeld. This device consists of a large diameter hemispherical dome (approximately 40 centimetres in diameter) with a xenon strobe lamp mounted on top of the dome. This configuration enables the presentation of diffuse and homogeneous stimuli while also ensuring background illumination for the entire retina. The recorded electrical activity is a fusion of positive and negative components from the retina's various processing phases [47].

The *International Society of Clinical* (ISCEV) recommended this stimulus system for use in obtaining ERG clinical records, and this protocol is now used globally [48].

2.3.2 Pattern ERG

A *Pattern Electroretinogram* (PERG) is a specialized electrophysiological test of central retinal function in response to a pattern reversal stimulus. The PERG is a valuable diagnostic tool that can help providers determine the correct diagnosis in patients with retinal diseases that are beyond the scope of standard clinical examination [49].

The PERG is typically measured non-invasively from the cornea under adapted lighting conditions using a control frame or grid pattern with an equal number of bright and dark elements. The stimulus pattern encompasses the central retina and typically reverses in contrast at full-cycle temporal frequencies ranging from 1 to 8 Hz [50].

As a result, the PERG provides both an objective index of macular function and, due to its internal retinal origins, a direct measurement of ganglion cell function. It thus aids in the electrophysiological differentiation of macular nerve and optic nerve dysfunction and provides invaluable additional data for a better interpretation of the cortically generated visual evoked potential for a similar stimulus [51].

2.3.3 Multifocal ERG

The *Multifocal Electroretinogram* (mfERG) is a recent advancement in electroretinographic testing that allows for the rapid assessment of retinal function from multiple areas at the same time. Since its discovery by Sutter and Tran in 1992, mfERG has transformed the objective assessment of retinal diseases. This method stimulates multiple areas of the retina at the same time and records each response independently, providing a topographic measure of the retinal electrophysiological activity. Many local retinal responses are recorded under light-adapted conditions and through dilated pupils, typically 61 or 103, representing the central and mid-peripheral visual field, allowing detection of small areas of retinal dysfunction in this region [52, 53].

The mfERG is a valuable tool for detecting macular, perimacular, and middle peripheral retinal abnormalities that are not visible on fundus examination, such as in chloroquine or hydroxychloroquine toxicity, siderosis, anorexia nervosa, tilted disc syndrome, and keratoconus. The mfERG can detect objectively central loss of activity in all types of maculopathies, including age-related macular degeneration, central serous chorioretinopathy, vitelliform maculopathy, macular hole, juvenile retinoschitis, and others. The extent of a central lesion can be estimated using mfERG in the early stages of Stargardt's maculopathy and toxic maculopathy. Furthermore, when combined with visual evoked potentials, mfERG is useful in the differential diagnosis of retinal and optic nerve diseases [53].

Although the technique is relatively new, there are already hundreds of centres around the world with the necessary equipment to record the responses of this ERG technique [54].

2.3.4 Focal Flash ERG

The *Focal Flash ERG* (fERG) measures the electrical response from a small area of the retina, approximately 3 to 5 degrees. In general, the fERG is used to elicit a response from a specific area of the retina in response to a focal stimulus (small spot of light). The fERG is commonly used to assess macular function, but it can be used to assess any part of the retina. Because the stimulus activates only a small portion of the retina, there are fewer dipole sources contributing to the field potential recorded in the cornea; thus, the fERG signal is quite small, on the order of 10 microvolts. Furthermore, the spot is surrounded by a ring of constant brightness that is typically brighter than the spot to reduce the influence of areas of the retina outside the desired spot that may be stimulated by stray light. This is perhaps the most straightforward approach to adapting to electrical stimulation, where electrodes in contact with the retina are intuitively focal stimuli [55, 56].

2.3.5 Applications of Electroretinography

As previously stated, the ERG is produced by retinal cellular responses to light stimulation. As a result, electrical currents can be recorded in the cornea when current flows in cells oriented radially from the electrical source to the sinks. Indeed, the recording and evaluation of these radial currents as an electroretinogram provides insight into the function of different retinal cell populations, allowing the characterisation and measurement of retinal dysfunction, the progression of retinal changes in disease, and the efficacy of therapy for retinal conditions [57].

The ERG, in fact, provides an objective and quantitative measure of retinal function, allowing the clinician to monitor the function of rod cells, cone cells, and ganglion cells in each eye. The specialist can analyze different retinal cell types and layers by carefully manipulating the stimulus and test conditions [58].

Several different types of ERG tests have been described throughout this chapter, these provide specific information about the patient's visual function. The ffERG assesses overall retinal function and can differentiate between cell types, revealing the function of photoreceptors, bipolar cells, ganglion cells, and amacrine cells, but it cannot provide specific information on individual retinal sectors. Such information can, however, be obtained using mfERG, which measures the response in each of a large number of small sectors of the retina, typically 61 or 103, providing a map that allows the clinician to pinpoint specific areas. PERG, on the other hand, measures

the response to a rapidly changing contrast pattern at a constant luminance level, providing information on ganglion cells and generalised macular function [59].

Moreover, ERG can aid in the diagnosis of a patient suffering from a variety of visual symptoms such as nyctalopia, photophobia, field defects, and decreased visual acuity. This follow-up can also help determine the significance of background abnormalities and provide an early diagnosis for patients with a family history of inherited retinal diseases [60]. Thus, ERG testing is especially important because it provides an objective and quantifiable assessment of visual function.

2.3.6 ERG-signal Processing

The ERG, as a biological signal, becomes difficult to interpret because it is the result of physiological process outputs that reflect the activity of a complex biological system. Biological signals are frequently non-stationary and difficult to model due to the inherent randomness of signals that vary from person to person; however, there are currently several signal processing mechanisms as well as feature extraction that have shown good results [61].

As previously stated, the ERG has two significant amplitudes: the negative a-wave and the positive b-wave. Because ERG signals have small amplitudes, around microV, it is critical to separate the signal from the noise wave and interference that can occur due to motion. An artifact in the ERG may appear, interfering with the recording and interpretation of the ERG b-wave. The noise line is the electrical interference caused by the cable that connects the electrode to the amplifier. The most common type of interference is caused by 50 or 60 Hz generated by power lines and electrical outlets. This falls within the ERG's bandwidth ranges (1-300 Hz), severely reducing the quality of the recorded data. A low pass filter can be used to reduce the component of this noise [44].

Different techniques can be used in ERG signal processing, including *Fourier analysis* (FA), *Principal Component Analysis* (PCA) and *Wavelet Analysis* (WA) in recognising pathological traces from the healthy ones [62]. Indeed, the use of frequency domain analysis can allow the extraction of useful information in very low amplitude ERG responses barely detectable above noise, as it separates different frequency contributions to the overall waveform and removes potential noise contamination in the time domain [57].

In turn, the lower and upper frequency limits for recording OP components are 75 to 100 Hz and 300 to 1000 Hz, respectively. Analyzing the OPs with digital filtering and frequency spectrum analysis using a *Fast Fourier Transform* (FFT) is more convenient. The frequency spectrum is calculated from the raw ERG waveform during this process, and the amplitudes of the components at different frequencies are derived. The frequency spectrum can be compared directly to that of the control group. Another method is to use band-pass filtering to extract the OPs waveform

by selecting components of the frequency spectrum in the range of 100-150 Hz. Following that, the amplitudes of the various OP waveforms can be analyzed [63].

Chapter 3

State of the Art

The state of the art is a fundamental step in any research and development process that serves to demonstrate and present the latest developments in the area of the project or research in question, as well as current trends. Thus, for a brief literature review five different Google Schoolar searches were conducted. Each one was differentiated by the key words used, being them: "ERG analysis and processing"; "diagnosis and ERG"; "classification and ERG"; "ERG and *Machine Learning* (ML)"; "ERG autism classification". The first article selection was based on reading the title of the articles and a brief reading of the abstract, this resulted in a total of 33 articles. After a more detailed reading of each of the articles collected, 8 articles were excluded, leaving a total of 18 articles in the state of the art presented below.

3.1 ERG Signal Analysis

In 1972, Algvere and Westbe, presented a study that aimed to analyse the oscillatory potentials of a human ERG in response to double flashes of light during dark adaptation. The sample of this study consisted of ERGs recorded from the right eye of a healthy woman. A bandwidth of 10-2500 Hz was used to adequately record oscillatory potentials that have a frequency of 120-160 Hz, and to ensure that the ERG was not contaminated by other slower components. To estimate the amplitude of the oscillatory potentials, the authors summed the heights of the wave peaks, and the height of each oscillation was measured from the baseline drawn between successive wavelet channels. The authors then proceeded to Fourier analysis of these

potentials, in order to obtain accurate estimates of their energy and frequency, in this analysis were used Fourier transforms described by specific equations for the acquisition of the above parameters. With this, the authors inferred that in light adaptation, the retina is organised so that oscillatory potentials can develop. While in dark adaptation, the organisation of the retina is different and oscillations cannot be distinctly seen. They added that retinal reorganisation occurs at a fixed level of retinal sensitivity, which corresponds to the absolute visual threshold for cones [64].

Burns, Elsner and Kreitz, in 1992, used their own ERGs to perform an analysis uncommon in the reviewed literature, this analysis involved studying the nonlinearities of the flicker ERG. To do this, the researchers divided their study into three experiments. In the first experiment they measured both the fundamental components and the second harmonic response of the ERG as a function of frequency and intensity. Next, in the second experiment they used beat frequencies in the human ERG as a probe to determine the temporal properties of the early visual system. Finally, in the third experiment they compared the effect of varying the modulation of a flicker light in the ERG with explicit predictions of two classes of nonlinearities: saturation-type nonlinearities and rectification-type nonlinearities. With this, this study concluded that the fundamental retinal response can be recorded at frequencies above 100 Hz and the second and higher harmonic response has measurable response frequencies up to 200 Hz, as well as, both the fundamental and second harmonic response components have multiple local maxima as a function of frequency. Adding further that the measurement of the retinal response to the sum of two sine waves indicates that there is an early low-pass temporal filter in the retina. This early filter has a cutoff frequency between 40 and 50 Hz. Finally, they concluded that the high frequency nonlinearity is not a compressive nonlinearity [65].

Rufiange et al. presented, in 2003, a method for the analysis of the ERG photopic curve based on easily identifiable and reproducible characteristics of the ascending and descending elements of the function described by the ERG. To carry out this method of analysis, the researchers started from a sample of 48 subjects, 31 women and 17 men, this was divided into two groups, to be submitted to different procedures. In both, the photopic ERGs were recorded, and in the first group the signals were recorded against a 17 cdm2 photopic background, and in the second, against a 30 cdm2 background. After collecting the signals, the analysis of each ERG involved measuring the peak times and amplitude of the a- and b-waves, where the amplitude of the a-wave was measured from the baseline to the depression and the amplitude of the b-wave from the depression of the a-wave to the peak of the b-wave. Peak times were measured from the onset of the flash to the peak of each wave. From the collection of these data, the authors established a list of seven characteristics to take into account: the maximum amplitude of the b-wave; the amplitude of the a-wave when the amplitude of the b-wave is maximum; the flash intensity at the

maximum amplitude of the b-wave; the ratio of the amplitude of the b-wave over that of the -a-wave measured in intensity; intensity of the stimulation that will generate a b-wave half of the maximum amplitude in the rising and falling part; and the intensity of stimulation required to generate an ERG where the amplitude of the b-wave is equal to the a-wave. This study concluded that the maximum amplitude reached by the b-wave together with that of the corresponding a-wave is not significantly modified following an increase in background luminance, however, there is an increase in the intensity of the stimulus required to reach these maximum values. Similarly, the intensity of the stimulus required to generate a b-wave with half the maximum amplitude in the rising limb of the photopic curve, as well as the intensity of the flash required to generate an ERG where the amplitude of the a-wave equals that of the b-wave demonstrate an increase in responses collected against a brighter background [66].

In order to analyse the contribution of computational analysis of high-intensity a-waves, Marmor et al., in 2002, developed a study, where during one year, ffffERGs of 38 individuals with retinal diseases and 8 normal subjects were recorded as a control group. As parameters the authors opted for the amplitudes of the a and b waves and their implicit times, or latencies, as well as measuring the intensity of the responses of each eye to calculate the values of Rmp3 and S, through the Hood-Birch method. These parameters were measured for both rods and cones, and the cone responses were subtracted from dark-adapted responses to isolate the rod parameters. The group of researchers concluded that a-wave analysis is an important tool for clinical research and the study of visual patients, however coupling it with the standard ERG protocol does not add significantly to clinical assessments in a routine clinical setting [67].

Aiming to analyse the long-term effect of light on mice, Richard et al. in 2006 carried out a study where 8 mice were exposed to specific light cycles for 20 days in order to cause retinal damage. ERGs were recorded at the end of each phase of the light cycles and then loaded into Matlab. In Matlab the authors proceeded to extract the amplitudes and implicit times of a- and b-waves and oscillatory potentials, where a-wave amplitudes were measured from baseline to depression and b-wave amplitudes from a-wave depression to b-wave peak, and implicit time was measured from stimulus onset to either a-wave peak or b-wave peak. Each individual waveform was subsequently smoothed using an average of 11 sliding window points. The smoothed waveform was then subtracted from the original raw waveform to produce the oscillatory potentials. The four OPs were subsequently measured with an algorithm written to identify the times and determine the amplitudes of each. For each OP, the previous time and its peaks were identified using the maxima and minima of the second derivative of the OP waveform, respectively. That said, the authors were able to infer that the inner retinal function recovered significantly, with b-wave

amplitudes remaining suppressed at 60% of their baseline values, OP amplitudes recovered completely, and the implied times of all inner retinal b-wave components and oscillatory potentials recovered to a level close to baseline values [68].

Barraco et al., proposed, in 2006, several mathematical methods capable of analyzing the a-wave component of the human ERG. For the preparation of these methods, a sample of ERGs from 10 subjects, from both eyes, was used. The authors organized this investigation into two parts: the first part concerns the onset and the initial slope of the -a wave, up to the first minimum (about 10-15 ms), the second part deals with the main part of the wave, up to about 30 ms. In both cases, the a-waves, recorded at various luminance levels, were provided with a set of appropriate functions representing possible models of physiological behaviour that would take place in the early stages of phototransduction. These simulations were achieved through the computer program provided with several functions describing the behaviour of the ERG wave. This program, in fact, consisted of two subprograms, one for the analysis of the onset of the a-wave and the other for the analysis of the main wave. In the first case, the working time interval was fixed by the program according to the occurrence of the first minimum, while, in the second case, the time interval was fixed at 30.27 ms. Based on the results obtained, the authors were able to infer that correlations between the responses of photoreceptor units under a light stimulus occur in the early stages, while random processes are created later [69]. Then in 2011, a group of researchers that also included Barraco, conducted a similar study, where the Fourier transforms were replaced by the wavelet transform. This transform was applied in order to analyse the ERG a-wave in the time-frequency domain. In this way, the authors detected and identified the time-frequency stable components of the a-wave, using six representative luminance values. The results indicated the occurrence of three frequencies that are in the 20-200 Hz range [70].

3.2 Classification Using the ERG Signal

Using ffERG, Devos et al. presented a study in 2005 that aimed to demonstrate photopic and scotopic retinal dysfunction in patients with dementia with Lewy bodies and visual hallucinations compared to patients with Parkinson's disease, patients without hallucinations, and controls. The ffERG were recorded from 16 patients with dementia with Lewy bodies, 17 subjects with Parkinson's disease, and 16 controls. As features of analysis and classification of the different groups, the authors selected the implicit times and amplitudes of a- and b-waves from both cones and rods. Once these data were collected, the researchers carried out a statistical analysis, which showed that implicit a and b wave times were higher in the dementia with Lewy bodies group when compared to the Parkinson's disease group and to the

controls, with no difference between the Parkinson's disease group and the controls [71].

Barraco et al., in 2013, designed a new method for the automatic recognition of ERG records in order to accelerate and objectify the identification of visual pathologies. To this end, they established a comparison between different ways of processing and classifying human ERGs. As sample they gathered 24 ERGs from 12 different people, where 8 subjects had congenital achromatopsia, 6 had X-linked congenital stationary night blindness, and 10 controls who had no visual deficit. For processing and classification of the acquired signals, the authors used three methods: (1) Principal Component Analysis, using the Statistical Package for Excel and the XLSTAT Addinsoft xlstat 2009 software; (2) Fourier Analysis, for the wavelet frequency decomposition, where they used a custom software based on the use of the fast Fourier transform algorithm; (3) Wavelet Analysis, for the extraction of features in the time-frequency domain through MATLAB's wavelet toolbox. In addition, the researchers performed a simple analysis of the implicit times and amplitudes of the a- and b-waves, which turned out to be the method with the highest accuracy and precision. With these methods, it was then possible to distinguish pathological subjects from healthy subjects, as well as differentiating subjects with different pathologies [72].

Benchorin et al. in 2017 developed a method to measure the functional response of photoreceptors and downstream retinal cells in mice. By way of a protocol, the authors present a set of steps to assess retinal function in mice. In the same they point out the use of data analysis software like Microsoft Excel or GraphPad Prism to generate traces. As features, they indicate how to measure the amplitudes and latency of the a and b waves. Stating that: the amplitude of a-wave can be measured by calculating the difference in the amplitude of the trace between time (t) = 0 and the lowest point of the trace, while the amplitude of b-wave can be measured by calculating the difference in the amplitude of the trace between the bottom of a-wave and the top of the highest curve; the latency of a-wave can be measured by calculating the time elapsed between t = 0 and the time point at the lowest point of the trace, and that of b-wave can be measured by calculating the time elapsed between the bottom of a-wave and the top of b-wave. To assess the relevance of the collected features, the authors recommend statistical analysis through two-way *Analysis of Variance* (ANOVA) [73].

In 2019, Wright et al, described an approach using ML to detect retinal toxicity associated with hydroxychloroquine using mfERG. The authors divided this approach into two parts. First, they trained a neural network to encode the temporal and amplitude information present in the mfERG waveforms. And in the second part, they used another neural network to identify the spatial information in the mfERG output. The first phase of the neural network was trained using 43432 waveforms from 356 subjects. The second phase of the network, on the other hand,

was trained to identify patients with indicators of hydroxychloroquine toxicity. The second phase of the neural network was trained using 10-fold cross validation. With this methodology, the authors were able to identify the mfERG of patients with toxicity with 66% specificity and 70% sensitivity [74].

In 2021, Dorado et al, conducted a study in order to find a method to diagnose multiple sclerosis from mfERGs. For the same, they were obtained with records from 15 eyes of patients diagnosed with the disease, and from 6 eyes of healthy subjects. They then pooled the mfERGs to perform correlation measurement with a normative database of filtered signals, based on the empirical model decomposition and three features from the continuous wavelet transform (CWT) domain. From the initial 40 features, the authors selected the 4 most relevant features using a filtering method and using a wrapper feature selection method. To proceed to classify the data, they resorted to *Support Vector Machine* (SVM) from which they obtained a Matthews correlation coefficient value of 0.89 (precision = 0.95, specificity = 1.0 and sensitivity = 0.93). With this, this study was able to identify an outer retinal dysfunction in patients with recent multiple sclerosis by analysing retinal responses in mfERG and employing an SVM algorithm as a classifier [75].

In order to evaluate and compare the performance of different classifiers, Diao et al., in 2021, used ffERGs in its complete form of 217 eyes, where 109 suffered from visual pathologies, and the other 108 were healthy, representing the control group. The authors processed the records using Matlab, where they collected the values of the a and b wave amplitudes, PhNR72 and PhNRmin (PhNR - Photopic Negative Response). For the classification, the following methods were used: temporal analysis; Nearest Neighbor Dynamic Time Warping; Linear Support Vector Machine; Linear Support Vector Machine With a Radial Basis Function Kernel; Random Forest; Gradient Boosting; Time Series Forest; Long-Short Term Memory Networks. For the training of the algorithms the waveforms were divided into training, validation and test sets. For each of the classifiers studied, the respective data were used for training, and the validation set was then used to fine-tune the parameters and select the best models for the test set. From this comparison, the classifiers with the best results were the Linear Support Vector Machine With a Radial Basis Function Kernel with a specificity of 0.86 and the Time Series Forest with an F1 score of 0.76 [76].

Schwitzer et al. presented a study in 2022 that aimed to analyse retinal changes in patients with major depressive disorder. In this study, PERGs of 24 people with the pathology and 29 control subjects were used. These signals were recorded every 4 weeks for 3 months. As features, the researchers chose to evaluate the amplitude and implicit time of P50 (positive peak) and N95 (negative peak), which in the literature have been shown to be the most used components in the PERG. Next, the authors proceeded to extract the time/frequency characteristics of the PERG

time series based on wavelet analysis. Finally, a statistical model was applied to this feature space and from which a metric was derived to quantify the pathological patient status, which provided discrimination between patients and controls with a p-value of 0.0001 [77].

3.3 Classification of Neurodevelopmental disorders Using the ERG Signal

In the previous sections we have seen studies that aimed at the analysis of the ERG and its potential use in classification mechanisms, in this section we present studies that address the classification of disorders with neurodevelopmental disorders, specifically ASD, using the ERG, which is more in line with the focus of this dissertation.

In February 1988, Ritvo et a., exposed a pilot study with the purpose of evaluating b-wave amplitudes in individuals with autism. For the same, they used a sample of 27 subjects with autism and 20 healthy volunteers. From this study, the authors concluded that around 48% of autistic patients presented abnormal b-wave amplitudes when compared to the normative data from the laboratory where the study was carried out. In fact, through the graphical representation of the ERGs collected the authors were able to observe a trend towards a decrease in b-wave amplitude in people with autism, showing that this fact may be a parameter of interest for the investigation of visual disorders associated with other diseases such as autism [78].

In order to explore the b-wave amplitudes of individuals with ASD, Constable et al. in 2016 carried out a study in this scope. In this study, the authors used light-adapted and dark-adapted ffERGs from two groups, one with 11 people with ASD and another for control with 29 people. From the ERGs, artefacts originating from blinking or motion were manually removed, and the a and b amplitudes of the ERG waves and peaks of the eye with the highest amplitude for each individual were used in the final statistical analysis. The authors also used the Naka-Rushton parameters (Km, associated with retinal sensitivity, n, with retinal homogeneity and Vmax with retinal responsiveness), which were derived from regression line analysis of the transformed fitted raw data, to calculate the b-wave amplitude. For statistical analysis, the non-parametric Krustal Wallis test, Mann-Whitney test and ANOVA were used. As conclusions, the researchers reported that in non-polarized light-adapted conditions, the b-wave is lower in the whole ASD group, and also, some individuals with ASD show subnormal b-wave amplitudes in the dark-adapted ERG [79].

Subsequently, in 2020, a group of researchers also represented by Constable, investigated the light-adapted ERGs of 90 individuals with ASD and compared them to 87 control subjects, average age. These ERGs were produced by a random series

of nine different Troland [1] types based on the strength of full-field flash and the standard ISCEV flash at 2/s on a white background of 30 cd m-2. As in 2016, Constable et al. chose the a- and b-wave amplitudes as well as their peak times as parameters and performed a statistical analysis to compare the obtained characteristics. This analysis showed that the ASD group had lower b-wave and a-wave amplitudes when flash intensities were higher and slower b-wave peaks. They also used Photopic Hill Models which showed that the Gaussian component peaks and logistic functions differed significantly between the groups under study [81].

Constable et al. in 2021 sought to investigate whether the negative photopic response of ffERGs could help explain visual perceptual differences in the autism spectrum. In this study, 55 subjects with autism, and 87 healthy subjects participated. As analysis parameters, the amplitudes of the negative photopic response at Tmin and at t = 72 ms as well as the a and b wave values were studied, then a comparison between the sample groups was established. Through statistical analysis of the data, the authors observed that there were no significant differences between groups for the measurements of the amplitude or time of the negative photopic response; however, they noted a difference between groups for the amplitude of the b-wave [82].

With this state of the art, it can be observed that the features used in the various studies are mostly based on the main components of the ERG signal, which are the a-wave and b-wave, more specifically on the values of their amplitudes and implicit times. In addition to the reduced range of features, statistical analysis is the methodology of choice for assessing the value of the features in differentiating healthy from unhealthy individuals, with very few studies using Machine Learning techniques for this purpose.

[1] A troland is a widely used measure of retinal illuminance in vision science and optical science [80]

Chapter 4

Materials and Methods

NF1 is one of the most common inherited diseases, this disease is characterised by a wide spectrum of ophthalmological manifestations, where Lisch nodules (iris hamartomas) stand out as one of the main criteria for the diagnosis of NF1. Choroidal nodules, retinal vascular abnormalities, café-au-lait spots and neurofibromas also stand out in this range of clinical features. Patients with NF1 also show the development of optic nerve gliomas and leukaemia, as well as frequent cognitive deficits and behavioural abnormalities. Although the clinical manifestations are well defined, the variety and multiplicity of their origin make it so important to invest in early and accurate detection of the disease so that appropriate action can be taken, especially as it is a disease with a higher prevalence in children.

4.1 Data Acquisition and Pre-Processing

In this study, the RETI-PORT/SCAN 21 equipment from Roland Consult Stasche & Finger was used to acquire the ERGs of the mice. This equipment is usable for: the *Visual Evoked Potential* (VEP) Standard, the ERG Standard, the VEP Flash, the scotopic and photopic ERG, the fast and slow *Electrooculography* (EOG), the mfERG Flash stimulation and the *Multifocal Visual Evoked Potential* (mfVEP) Stimulation Standard, and all ISCEV standards and guidelines are included. The unit includes a biosignal amplifier, which in turn includes a preamplifier near the patient. All patient data and results are stored in a database. The curves of the biosignal and

the average of all channels can be displayed on the monitor. In the analysis mode, the system automatically sets all markers and calculates all set parameters [83].

The RETI-port/scan 21 device offers the ISCEV standard programs, giving the user the option to create custom programs; the average result and biological signal are displayed on screen, artefacts are automatically rejected, amplifier data is controlled by software, and an automated impedance test can be performed at any moment; the multi-focal examination of each signal with N1, P1 marker, 2D, and 3D, as well as user groups; a straightforward graphical user interface with numerous separate windows; A database that is stored on DVDRW, the export of all data to an EXCEL file for statistical analysis, special mathematical functions to calculate various parameters, the FFT to analyze steady state results, the *Digital Signal Processing* (DSP) filter, and small interface technology for common computers are all included [83].

Although this whole system seems to present several functionalities, the ERGs analysis process presents different problems and difficulties for the user, it stands out: the general difficulty in identifying regions and points of interest, since the user needs prior knowledge of the typical ranges of each component to manually select the point of interest; the inefficiency in the automation of the analysis of the collected signals, since after collection it is necessary to manually apply filters for the identification of elements, such as the b-wave and the OPs of the signals, for each luminous intensity; finally, the method of data analysis, being numerical, makes it difficult to extract features and subsequently does not facilitate the visualization and interpretation of the results.

Problems with signal processing are exacerbated by the lack of correlation in this data, which makes it impossible to distinguish between pathogenic and non-pathological data. It is evident from this that these data require a statistical analysis using a number of ERGs to identify groupings with differentiating traits, as well as perhaps the use of classification algorithms.

4.2 Animals

In this work, both male and female transgenic NF1 animal models were used. Since each animal was genotyped, we were able to establish whether it belonged to the heterozygous NF1 (NF1+/-) or *Wild Type* (WT) genotype groups. In the housing facility, the animals were kept at a constant temperature of 22°C and were given unlimited access to food and water. 0.5% tropicamine eye drops were administered to dilate the pupils.

60-day-old mice male and female mice were sedated by receiving intraperitoneal injections of 80 mg/kg ketamine and 5 mg/kg xylazine. Throughout the treatment, a heating pad set to 37°C was used to maintain the animals body temperatures.

The mouse was chosen as a model due to its similarity to humans in terms of mechanisms and traits when displaying the NF1 gene mutation. The best alternative for studying the cognitive and behavioral aspects, which are also seen in people with this illness, is the NF1 gene product because of its 98% resemblance to humans [84].

4.3 Light stimulation and ERG Recordings

For light stimulation the Ganzfeld stimulator was used. The use of a Ganzfeld stimulator affects both clinical results and ease of operation. The main advantages of using this stimulator are the proximity of the flash electrode to the surface of the sphere to increase the intensity and uniformity of the stimulus, the provision for interposing colour filters in the background light, the use of a diffusing filter in the background light; and placing the stimulus flash earlier to avoid uneven illumination of the retina [85]. Thus, based on the recommendations laid out in the ERG acquisition procedure [86], four different responses were recorded:

- Scotopic Luminance Responses: The dark-adapted animals were subjected to a series of seven distinct white light flashes, ranging in intensity from 0.0095 to 9.49 cd-s/m2, three times at 0.1 Hz.
- Photopic Adaptometry: When measuring the light adaptation process using a white background light (25 cd/m), bright white flashes (9.49 cd-s/m2) were delivered three times at 1.3 Hz at the beginning of light adaptation and at 2, 4, 8, and 16 minutes.
- Photopic Luminance Responses: White flashes were provided three times at 1.3 Hz with a white background light (25 cd/m) and seven varied light intensities (0.0095 to 9.49 cd-s/m2).
- Photopic Flicker: Ten white brilliant flashes (3.00 and 9.49 cd-s/m2) were delivered at 6.3 Hz against a white background light (25 cd/m2).

Since they constitute a richer signal in rodent ERG, only the scotopic luminance responses were examined in the current study. After an overnight period of darkness adaption, ERGs were acquired; no animals were given any medication or other forms of treatment.

The pupil was completely dilated with topical tropicamide (0.5%) before to the ERG recording. For the duration of the recording experiment, Celluvisc® eye drops were adequate to keep the corneal potential stable (impedance 6 kohm). With a ground electrode in the tail, a reference electrode at the head, and a gold wire electrode at the cornea, the electrical responses were recorded. Throughout the entire process, electrode impedance stayed below 6 kohm. With the exception of

the flicker test, which employed a 0.65 kHz sampling rate, the ERG waveforms were recorded with a band width of 1 to 300 Hz and sampled at 3.4 kHz by a digital acquisition system (Roland Consult GmbH, Brandenburg, Germany), and they were analyzed using Python.

4.4 Database

The research team, through the process described above, saved several .csv files with the data of the mice, as well as their ERGs recorded. Each file represents one tested individual including: identification, sex, test date and the ERGs collected, corresponding to 16 readings taken with eight different light intensities for each eye. Each ERG is composed of 512 points representing 150 milliseconds of recording. The intensities tested were infrared light as the lowest light intensity and 7 readings for each eye with intensities ranging from -25dB to +5dB (0.0095 to 9.49 cd-s/m2), always adding +5dB in each new reading. Fifty files were used, 10 female and 40 male individuals for a final number of 23 wild type and 27 NF1 pathological individuals. Thus, the database used in this study can be represented by 4.1.

Table 4.1: Summary table of the database used in this study. The first column indicates the number of individuals, male and female, for the Neurofimatosis type 1 (NF1) genotype. The second column represents the same but for the Wild Type (WT) genotype.

	NF1	WT	Total
Female	7	3	10
Male	20	20	40
Total	27	23	50

Chapter 5

Technical Development

The main objective of the study in question is the creation of an algorithm that, taking into account a series of specific features, is capable of efficiently and accurately distinguishing pathological individuals, NF1, from non-pathological individuals, WT. With this in mind, and taking into account the other objectives of the study presented, a procedure was outlined that would allow these to be achieved in an organised and coherent manner, as can be seen in Figure 5.1.

The process began with the organisation of the files, where an analysis was made of the CSVs received and where they were separated according to their genotype - Wild and NF1 - with the aim of facilitating the use of the data.

The Python program developed is waiting to receive the raw CSV files, once it reads and processes them, converting them into a set of ERGs, representing the various light intensities with which they were recorded. Once the ERG signals are gathered, the algorithm collects the features of interest, which in turn undergo statistical analysis and classification by ML algorithms.

Finally, a summary file is generated with a resume of the processing and analysis of the data initially entered.

5.1 Processing of CSV Files

When the program starts, the user is asked to select the files to be analysed, both the wild files and the NF1 files, as well as for which channel (which corresponds to the light intensity) they want the analysis to be performed. These files are raw, that

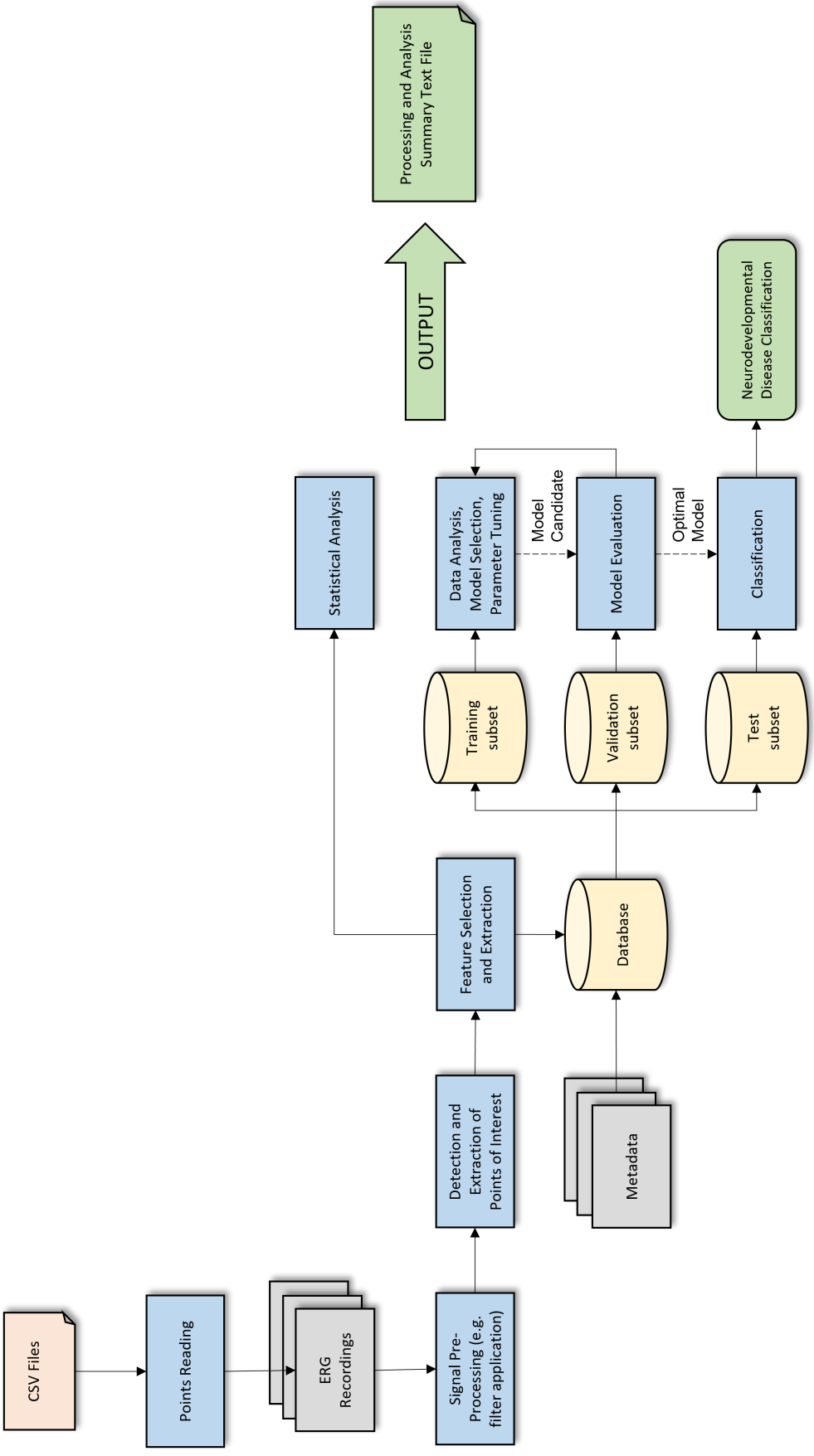


Figure 5.1: Pipeline of the procedure adopted for the project.

is, with too much information and not ready for direct translation to be interpreted as an ERG signal, and for this reason they go through a processing phase that allows collecting the points contained in the CSV files and converting them into a signal as shown in Figure 5.2.

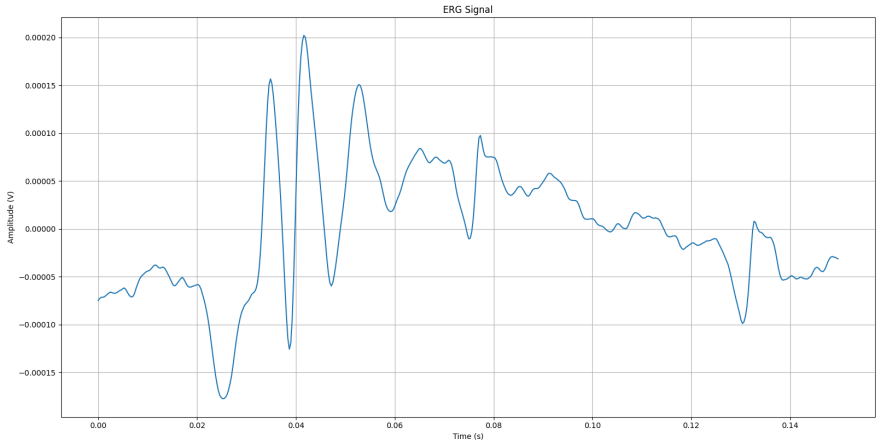


Figure 5.2: Example of a single ERG from channel 15.

For each CSV file a list of 16 ERGs is returned, similar to those shown in Figure 5.3, corresponding to each light intensity.

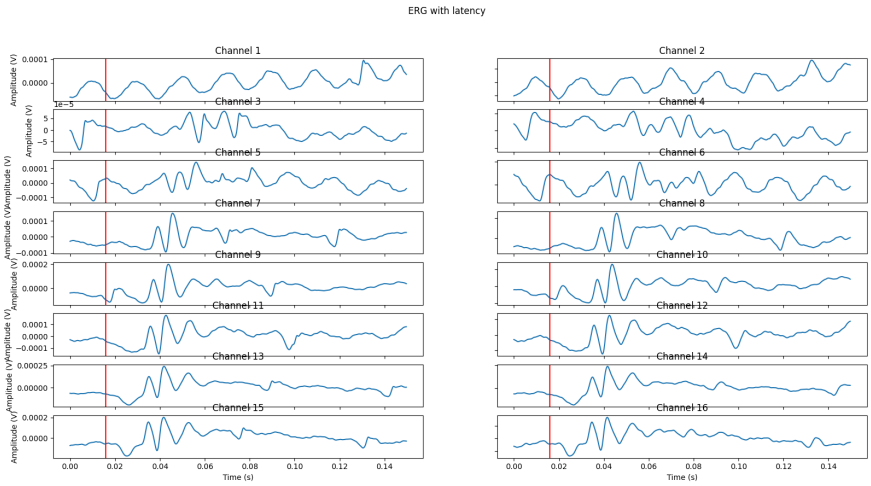


Figure 5.3: Example of 16 ERGs obtained from a CSV file.

Each time the program is run, only one of the channels is analysed, more specifically, the user chooses, at the beginning of the program, which channel to analyse. Although all the channels of the initial file are initially collected, signal preprocessing and the collection of points and characteristics of interest, as well as

statistical analysis and classification via ML are only associated with the selected channel.

5.2 Signal Pre-Processing and Selection/Extraction of Features

After the literature review, it is clear that the amplitudes and implicit times of the a and b waves are used as features of interest in the analysis of ERGs, since they are the most prominent components in the signal and the easiest to detect. However, in order to evaluate new features that may be potential indicators in the distinction between pathological and non-pathological signals, NF1 and WT respectively, four more features were used in this study: the maximum signal frequency, the minimum signal frequency, the maximum of the first signal derivative and the minimum of the first signal derivative.

A *Frequency Sampling* (FS) equal to the split between the number of signal points, 512 points, and the total length of the signal window, 15 milliseconds, was used to pre-process the ERG in order to gather the values of the amplitudes and implicit times:

$$FS = \frac{Numberofsignal Points}{WindowTime} = \frac{512}{0.15} \approx 3413.33Hz \tag{5.1}$$

Since the a and b waves are in the first half of the signal window, the reading window of the signal was reduced to half its size, that is, as each signal is composed of 512 points, for the treatment of the signal it was decided to use the first 256 points, ensuring the presence of the a and b waves in this window. In addition to reducing the window to half, the baseline corresponding to the signal was also subtracted from it, and this was calculated as the mean values of the first 16 millisenconds of the signal, where we calculate the time as the multiplication between the FS and the time elapsed in the recording until the luminous flash occurred, which occurs at 16 milliseconds.

As mentioned earlier, the amplitude of the a-wave corresponds to the distance between the trough of the a-wave and the baseline of the signal, in this way, the amplitude and the implicit time of the a-wave are directly taken from the signal resulting from the treatment stated above, with the amplitude being obtained as the minimum of the signal and the implicit time the moment at which this minimum occurs.

For the b-wave, it was necessary to apply a low-pass filter with a cutoff frequency of 47.5 Hz before extracting the values of interest, in order to prevent other maximum points, such as the OPs, from being confused with the peak of the b-wave. This signal pre-processing was done using the *scipy* library, more specifically the *signal* module

that Python provides for free and which allows the use of a variety of filters and specific functions for signal processing. Having said that, to calculate the amplitude of the b-wave we must take into account that it corresponds to the distance between the valley of the a-wave and the peak of the b-wave, so to determine this metric we found the maximum of the signal and added to it the value of the amplitude of the a-wave, calculated previously. Subsequently, through the time at which the maximum of the signal occurs, the implicit time of wave b was obtained.

The frequency of a signal is traditionally obtained by the FFT which is translated into Equation 5.2. This feature comprises the time domain and can be calculated using the function *fft()*, which belongs to the same library used for the low-pass filter. Since the frequency values are given in negative values, for a normalization of the data the modulus function was applied, *abs()*.

$$X\left(jw\right) = \int_{-\infty}^{\infty} x\left(t\right) \ . \ e^{-jwt}. \ dt \tag{5.2}$$

In turn, the first derivative of a signal represents the rate of change of y with x, that is, $\frac{dy}{dx}$, which can be interpreted as the inclination of the tangent to the signal at each point of the signal, that is to say, it translates the inclinations present in the original signal. Thus, to find the value of the derivative of the ERG signal, we resorted to the numpy library, more specifically to the *diff()* function, which returns a vector of points referring to this derivative, meaning less a value than the original signal.

Finally, since the output of both functions is a vector of values, the functions *max()* and *min()*, which also belong to the *numpy* library, were applied to find the maximum and minimum of each component. The summary of the features collected for this study can be seen in Table 5.1.

Table 5.1: Set of features under study together with the method of extraction.

Features	**Extraction**
a-wave Amplitude	Minimum of the ERG signal.
a-wave Implicit Time	Time where the minimum of the ERG signal occurs.
b-wave Amplitude	Minimum of the ERG signal, after applying the low-pass filter.
b-wave Implicit Time	Time where the minimum of the ERG signal, after applying the low-pass filter.
Maximum of Frequency	Module value of the maximum of *fft()* applied to the ERG signal.
Minimum of Frequency	Module value of the minimum of *fft()* applied to the ERG signal.
Maximum of Derivative	Maximum value of the first derivative by applying the *diff()* function to the ERG signal.
Minimum of Derivative	Minimum value of the first derivative by applying the *diff()* function to the ERG signal.

5.3 Statistical Analysis

In any research, the mechanism for distinguishing two groups is to find patterns, trends and/or relationships within them that allow them to be characterised differently from one another. However, this analysis is not always possible with the naked eye or at first sight, and it becomes necessary to resort to more advanced methods of analysis, such as statistical analysis. Therefore, inherent to a statistical study are the hypotheses of the case study, which are in fact the summary of what one wants to assess, thus, in this study, the hypotheses are at issue:

Null Hypothesis: There are no characteristics with relevant weight that allow the distinction between pathological and non-pathological individuals (No significant differences between genotypes).

Alternative Hypothesis: There are characteristics with relevant weight that allow the distinction between pathological and non-pathological individuals (There are significant differences between genotypes).

There are several statistical techniques to infer this analysis, for this research Independent *Student's t-test* (t-test) and One-Way ANOVA were applied.t-test and

ANOVA are statistical procedures for comparing means between groups in hypothesis testing. The test variable (dependent variable) must be on a continuous scale and approximately regularly distributed for these approaches.

Before performing the statistical analysis with these two methods, there is a set of assumptions that need to be met, such as: the dependent variable (the variable of interest) must be a continuous scale; the data must have independence of observations, that is, there should be no data common to the two groups; the dependent variable must be normally or quasi-normally distributed for each group; finally, the data must have homogeneity of variances [87].

A t-test is s a type of statistical test that is used to compare the means of two groups. As a ramification arises the independent samples t-test, which is one of the most widely used inferential tests and is a common first choice for when two groups are examined against a continuous variable. In general, the t-test is a very common test to determine differences when a singular measured variable (e.g. amplitude of a-wave, implicit time of b-wave, maximum of the first derivative of the signal, etc) is compared with differences between a cluster variable with two disaggregation groups (e.g. NF1 vs WT) [88], so the t-test was applied to each feature in order to infer as to its statistical significance.

In this way, to perform the statistical analysis of the data, we began by preparing a dataframe containing the variables/features that we wanted to infer associated with the genotype of the individuals, NF1 and WT. From this, each characteristic was isolated according to the genotypes in order to perform an independent t-test for each feature, that is, a t-test was performed for each variable in which the differences between the genotypes were compared.

Thus, seven t-tests were performed, with a 95% confidence interval, using two different mechanisms: one resorting to the standard Python function called *ttest_ind()* belonging to the *scipy* library, more specifically to the *stats* module; and another to a custom function that receives the values of the characteristic for the pathological genotype and the values for the non-pathological one, as well as the alpha indicating the confidence interval, returning the values of the t-statistic (t-stat), the *Degrees of Freedom* (DF), the *Critical Value* (CV) and the p-value. These variables were calculated according to their definition, where [89]:

- The t-statistic value is calculated as the ratio between the differences of the means of the input values and the value of the standard error of the difference between the means.
- The number of DF for the test is calculated as the sum of the observations in both samples minus two.

- The CV can be calculated using the *ppf()* percentage point function from the *scipy* library, provided by Python, for a given significance level, such as 0.05 (95% confidence).
- The p-value can be calculated using the cumulative distribution function *cdf()*, again provided by the *scipy* library.

While the t-test is used to compare the means of two groups without the need for multiple comparisons because a single p-value is observed, ANOVA is used to compare the means of three or more groups without the need for multiple comparisons because a single p-value is observed. There are several types of ANOVA tests, each with its own set of objectives. One-way ANOVA and one-way repeated measures ANOVA are the two primary forms of ANOVA. The first method is used for independent observations, while the second method is used for dependent observations. When only one categorical independent variable is used, it is referred to as one-way ANOVA, however, when two categorical independent variables are used, it is referred to as two-way ANOVA [90].

In our case study, only one-way ANOVA was applied since only the independent variable "Genotype" is analysed. In fact, one-way ANOVA is an extension of the independent samples t-test. The significant p-value of this test is related to the multiple comparisons test used to find the significant pair/s. Since it is an omnibus test, this tests for an overall difference, that is at least one of the groups is statistically significantly different from the others. However, if the ANOVA is significant, you cannot tell which group is different. To find out which group is different, it is necessary to perform planned or post-hoc comparisons. In this test, a continuous dependent variable and a categorical independent variable (e.g. genotype), each with at least three categories (e.g. the amplitude of the b-wave, the implied time of the a-wave, the maximum frequency of the signal, etc.), are used [91]. The reason it is called one-way or one-factor ANOVA, even though there are 3 or more groups to be tested, is because these groups are under a categorical variable, and the name refers to the number of variables in the analysis, not the number of groups.

To apply the one-way ANOVA test, the *pandas*, *researchpy* and *statsmodels* libraries were used. Through the pandas library, more specifically the *stats* module, it is possible to directly apply the test using the *f_oneway()* function; however, this only returns the t-statistic and p-value values. To perform a more detailed test, the same method was performed, but using another procedure that actually encompasses two phases:

1. The model was fitted using an estimation method using the *ols()* function from the *statsmodels.formula.api* library. OLS stands for Ordinary Least Squares and is a general function to perform a linear regression.

2. The fitted model was ran with the ANOVA method to produce the ANOVA table using the *anova_lm()* function from the *statsmodels.api.stats* library.

Thus, as a result we have an ANOVA table with the values of the variables: sum of squares (sum_sq), the number of degrees of freedom (df), the F value, the p-value (PR), the mean of squares (mean_sq), the effective size (eta_sq) and the partial effective size (omega_sq). As mentioned earlier, in order to know which groups differ significantly, it is necessary to perform post-hoc tests. There are several methods for applying post-hoc tests, such as the HSD test, the Bonferroni test, and the Dunn-Šidák test. In the present study, HSD was used as it is the most common when using the Python language in applying the One-way ANOVA.

In fact, HSD is used to test differences between sample means to determine their statistical significance. HSD tests all pairwise differences while controlling for the probability of making one or more Type I errors [92]. That said, the HSD was applied using the *statsmodels.stats.multicomp* library that by applying the functions *MultiComparison()* and *tukeyhsd()*, allows to obtain a table with: the groups being compared (group 1 and group 2), the difference between the means of the groups (meandiff), the corrected p-value (p-adj), the value of the lower range of the confidence interval (lower), the value of the upper range of the confidence interval (upper) and the value of the reject (true or false) based on the p-value.

This procedure was carried out for each of the features under study.

5.4 Classification via Machine Learning

Inference and prediction are the two fundamental objectives of biological systems research. Inference develops a mathematical representation of the data creation process to formalize knowledge or test a theory about how the system acts. The goal of prediction is to foresee future actions or unobserved results [93].

In theory, many statistical and ML techniques can be applied to both prediction and inference. But inference is the primary goal of statistical methods, and it is accomplished by developing and modifying a probability model that is unique to the job at hand. The goal of ML, in contrast, is prediction. It employs general-purpose learning algorithms to identify patterns in frequently dense and voluminous data. When working with "big data" where the quantity of input variables outweighs the quantity of subjects, ML techniques are especially helpful. In the midst of complex non-linear interactions and data collection without a carefully planned experimental design, ML algorithms can still be useful [94].

Thus, although statistical analysis allows inference as to the patterns and relationships of the variables under study, the best way to predict and classify data with better accuracy is the use of ML, more specifically by resorting to the classification algorithms it makes available.

Having said this, and one of the main objectives of this research being the distinction and accurate classification of healthy and unhealthy individuals, four classification algorithms were used: *Logisitc Regression* (LR), SVM, *Naive Bayes* (NB) and *Random Forest* (RF); using the eigth features, already mentioned above, as a function of the "Genotype".

The process employed for classifying the data was identical for the various algorithms, as there is a library in Python called *sklearn* that contains specific functions for all the algorithms listed above. Therefore, there is an initial procedure common to all the algorithms:

1. A dataset was created from the dataframe previously used in the statistical analysis;
2. The values "NF1" and "WT", present in the "Genotype" column of the dataset, were mapped to 0 and 1, respectively, to make the variables binary;
3. To the dataset were added the eigth features for classification;
4. For each classification algorithm, the function *train_test_split()* was used to split the dataset into the training and test subdatasets of the classifier, and this function has as parameters:
    - the input data;
    - the size of the test subdataset, where we used a value of 0.2 that means using 20% of the initial dataset for testing and 80% for training;
    - the random state that allows controlling the shuffle applied to the data before applying the division. This value was being fine-tuned in order to find the best result.

Once the division of the initial dataset into the two subgroups of data was completed, the classification was carried out using procedures customised for each of the classification algorithms described in the following paragraphs.

Logistic Regression: after splitting the dataset, we scaled our feature set with the help of the *StandardScaler()* library to then apply logistic regression on the training set using the *LogisticRegression()* function and check the accuracy score with the help of the *accuracy_score* library.

Support Vector Machine: for the application of SVM, a classifier was created through the *svm.SVC()* function, with a linear kernel, in order to apply it to the training set. The function *predict()* was applied on the test set in order to perform the prediction. Afterwards, as in LR, the *accuracy_score* library was used to infer the accuracy of the classifier.

Naive Bayes: the classification through NB can have several variants, in our study the Gaussian Naive Bayes was applied, that is possible through the function

GaussianNB(). After using the classifier in the training set, the confusion matrix was created using the *metrics.confusion_matrix()* library. The prediction was possible by applying the *predict()* function. As metrics for evaluating this algorithm we have the score and the roc_auc. The first is obtained by the *score()* function associated to the classifier itself, and the second is obtained in three steps using the *predict_proba()*, *roc_curve()* and *auc()* functions sequentially.

Random Forest: After splitting, the model was trained on the training set using the *RandomForestClassifier()* function and then predictions were made on the test set using the *predict()* function. Again, to evaluate the accuracy of the classifier the library *accuracy_score* was applied.

These algorithms were chosen because they are among the most popular classifiers in the ML world, particularly in their application to signal classification [95, 96, 97].

At the end of all this process, a text file is generated with all the relevant data from the analysis performed by the program described here, from the input data to the classification of the data, through feature extraction and statistical analysis.

The full code of the algorithm presented in this section can be accessed via the GitHub link: https://github.com/inesbvigo/ergClassification.git.

Chapter 6

Overview and Analysis of the Results

Having described the procedure of the program developed, the results are presented and discussed in this section. Given that there are 15 ERG reading channels corresponding to the light intensities with which they were recorded, in order to enrich the discussion of the results we chose to compare them for three different channels, channel 3 (intensity 1), channel 9 (intensity 4), and channel 15 (intensity 7), so as to represent a lower, an intermediate and a higher intensity, respectively.

6.1 Data Collection

As described above, after the selection of the files for analysis by the user, they are processed for the posterior collection of the features of interest, which are the base data for all subsequent analysis. These data were organised in a table, more specifically in a dataframe, with a column dedicated to the Genotype (independent variable) and a column dedicated to each of the features under study (dependent variables) resulting in a table with 9 columns (Genotype + maximum signal frequency, minimum signal frequency, the maximum of the first derivative of the signal, the minimum of the first derivative of the signal, the amplitude of wave a, the implicit time of wave a, the amplitude of wave b and the implicit time of b-wave) and 50 lines (23 of WT genotype and 27 of NF1 genotype, equal to the number of files introduced), as illustrated in Figure 6.1, in this case for channel 15 as representative.

Genotype	Max_Frequency	Min_Frequency	MAX_Diff	MIN_Diff	A_Wave_Amplitude	A_Wave_Time	B_Wave_Time	B_Wave_Amplitude
WT	0,0165	8,58E-08	5,84E-05	-3,22E-05	-0,0001	0,0252	0,0416	0,0002
WT	0,0225	4,05E-08	5,05E-05	-2,44E-05	-0,0002	0,0270	0,0583	0,0004
WT	0,0115	2,71E-07	2,52E-05	-1,17E-05	-0,0001	0,0267	0,0589	0,0002
WT	0,0153	1,97E-07	3,87E-05	-1,66E-05	-0,0001	0,0270	0,0545	0,0003
WT	0,0327	6,55E-07	2,95E-05	-1,37E-05	-0,0001	0,0267	0,0744	0,0003
WT	0,0294	2,07E-07	4,62E-05	-2,42E-05	-0,0002	0,0270	0,0653	0,0004
WT	0,0207	1,75E-08	5,09E-05	-2,23E-05	-0,0002	0,0264	0,0472	0,0003
WT	0,0193	4,30E-09	1,87E-05	-8,40E-06	-0,0001	0,0281	0,0703	0,0003
WT	0,0290	1,18E-07	4,77E-05	-2,36E-05	-0,0001	0,0264	0,0469	0,0003
WT	0,0179	2,05E-07	2,49E-05	-1,08E-05	-0,0001	0,0270	0,0592	0,0003
WT	0,0255	2,66E-07	2,59E-05	-1,03E-05	-0,0001	0,0278	0,0744	0,0003
WT	0,0154	2,74E-07	2,25E-05	-1,17E-05	-0,0001	0,0267	0,0448	0,0002
WT	0,0207	3,40E-07	4,25E-05	-1,94E-05	-0,0002	0,0270	0,0548	0,0003
WT	0,0368	9,54E-07	3,91E-05	-1,61E-05	-0,0002	0,0272	0,0642	0,0005
WT	0,0402	2,35E-07	3,03E-05	-1,60E-05	-0,0002	0,0284	0,0744	0,0005
WT	0,0188	2,04E-09	2,85E-05	-1,45E-05	-0,0001	0,0272	0,0677	0,0002
WT	0,0148	1,91E-07	2,07E-05	-1,04E-05	-0,0001	0,0272	0,0744	0,0002
WT	0,0127	1,10E-08	1,51E-05	-7,16E-06	0,0000	0,0272	0,0744	0,0002
WT	0,0212	1,38E-07	2,59E-05	-1,30E-05	-0,0001	0,0281	0,0744	0,0003
WT	0,0244	7,00E-09	1,37E-05	-6,84E-06	-0,0001	0,0278	0,0744	0,0003
WT	0,0242	3,23E-07	2,05E-05	-9,77E-06	-0,0001	0,0284	0,0744	0,0003
WT	0,0166	4,24E-08	2,67E-05	-1,51E-05	-0,0002	0,0267	0,0744	0,0003
WT	0,0181	2,13E-08	2,25E-05	-9,61E-06	-0,0001	0,0275	0,0744	0,0003
NF1	0,0492	6,24E-07	7,45E-05	-3,77E-05	-0,0004	0,0281	0,0598	0,0007
NF1	0,0194	4,37E-07	3,13E-05	-1,77E-05	0,0000	0,0281	0,0445	0,0002
NF1	0,0146	2,52E-07	5,32E-05	-2,17E-05	-0,0002	0,0287	0,0425	0,0003
NF1	0,0298	1,53E-07	3,83E-05	-1,92E-05	-0,0002	0,0270	0,0668	0,0004
NF1	0,0393	3,82E-07	4,04E-05	-2,27E-05	-0,0002	0,0272	0,0609	0,0005
NF1	0,0091	2,94E-08	2,33E-05	-1,16E-05	-0,0001	0,0267	0,0650	0,0002
NF1	0,0287	4,17E-07	3,76E-05	-1,32E-05	-0,0001	0,0284	0,0694	0,0003
NF1	0,0163	1,54E-08	2,33E-05	-1,35E-05	-0,0001	0,0267	0,0674	0,0003
NF1	0,0431	2,73E-07	5,55E-05	-1,88E-05	-0,0001	0,0281	0,0744	0,0005
NF1	0,0305	2,12E-07	3,78E-05	-1,48E-05	-0,0001	0,0272	0,0665	0,0004
NF1	0,0249	2,77E-07	3,63E-05	-1,58E-05	-0,0001	0,0275	0,0486	0,0003
NF1	0,0123	2,56E-07	3,00E-05	-1,40E-05	-0,0001	0,0270	0,0618	0,0002
NF1	0,0148	4,47E-08	4,00E-05	-1,87E-05	-0,0002	0,0264	0,0469	0,0003
NF1	0,0152	8,08E-08	1,42E-05	-6,83E-06	0,0000	0,0267	0,0551	0,0002
NF1	0,0152	3,15E-07	3,87E-05	-1,55E-05	-0,0001	0,0270	0,0592	0,0002
NF1	0,0266	3,36E-07	3,27E-05	-1,58E-05	-0,0001	0,0272	0,0527	0,0004
NF1	0,0248	3,95E-07	1,90E-05	-7,50E-06	-0,0001	0,0284	0,0744	0,0003
NF1	0,0390	2,50E-09	4,09E-05	-2,07E-05	-0,0002	0,0284	0,0744	0,0005
NF1	0,0299	9,85E-08	3,74E-05	-1,45E-05	-0,0002	0,0284	0,0744	0,0004
NF1	0,0121	2,51E-08	1,12E-05	-6,35E-06	-0,0001	0,0287	0,0744	0,0002
NF1	0,0167	1,74E-07	1,59E-05	-1,14E-05	-0,0001	0,0281	0,0744	0,0002
NF1	0,0272	1,25E-07	2,23E-05	-9,80E-06	-0,0001	0,0275	0,0744	0,0003
NF1	0,0295	2,47E-07	1,95E-05	-8,47E-06	0,0000	0,0275	0,0744	0,0003
NF1	0,0248	3,13E-07	2,59E-05	-9,70E-06	-0,0002	0,0278	0,0744	0,0003
NF1	0,0305	2,10E-07	1,71E-05	-1,05E-05	-0,0002	0,0287	0,0721	0,0004
NF1	0,0274	1,78E-07	2,28E-05	-1,30E-05	-0,0001	0,0284	0,0744	0,0004
NF1	0,0224	5,39E-07	2,95E-05	-1,68E-05	-0,0001	0,0267	0,0647	0,0003

Figure 6.1: Example of a dataframe made from channel 15.[1]

In a first phase of data analysis, the graphic representation becomes an essential tool for allowing a direct visualisation of possible differences between the groups under study. For this very reason, using the dataframe presented above, box plots were generated grouping the values of each characteristic as a function of genotype, for the three channels under study, represented in figures 6.2, 6.3 and 6.4. These same graphs were generated using the *boxplot()* function of the *seaborn* library, which receives as input the values of the x and y axes, as well as the dataset, and in this case even the variable showfliers with the value "false" in order not to contemplate in the representation the outliers, that is, the values that move abnormally from the other values.

Observing these graphs, we can notice that there are significant differences in the data distribution for some features under study. Namely, for channel 3, we can see that the graph corresponding to the data of the implicit time characteristic of the b-wave is not the desired, not even getting a box for the values of WT, this is because, removing the outliers, the values for this variable do not show great variance having most of them the same value, as we can see in figure 6.1. However, this graph contains important information, as it shows that there is a significant difference between WT and NF1 values. For this channel, no other feature showed, at first sight, significant visual differences to distinguish the two groups.

For channel 9 a small difference can be noted in the box plots for the implicit time feature of the a-wave, noted by the deviation that exists between the two medians of the groups (represented by the value that divides the box into two).

Finally, for channel 15, the graphs with possible significant differences can be highlighted as those of the characteristics: implicit time of a-wave, amplitude of b-wave and minimum of signal frequency. Of these three it can also be noted that the implicit time of the a-wave is the one which presents the most deviations between the genotypes.

[1]Max_Frequency: Maximum frequency of the signal (measured in terms of energy); Min_Frequency: Minimum frequency of the signal (measured in terms of energy); Max_Diff: Maximum of the first derivative; Min_Diff: Minimum of the first derivative; B_Wave_Amplitude: B-wave amplitude (measured in Volts); B_Wave_time: B-wave implicit time (measured in seconds); A_Wave_Amplitude: A-wave amplitude (measured in Volts); A_Wave_Time: A-wave implicit time (measured in seconds)

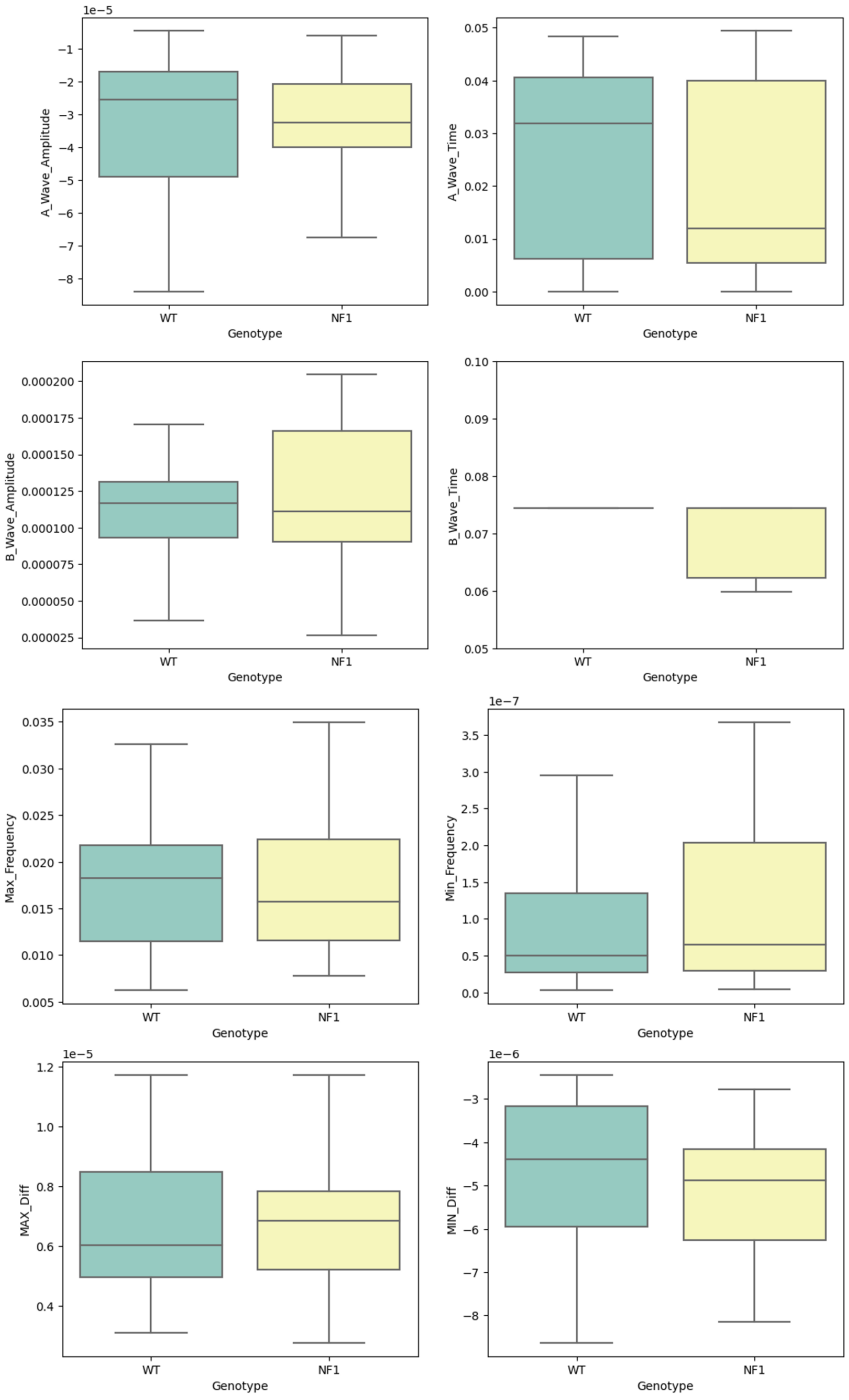


Figure 6.2: Graphic representation of the data grouped by feature as a function of genotype for channel 3.

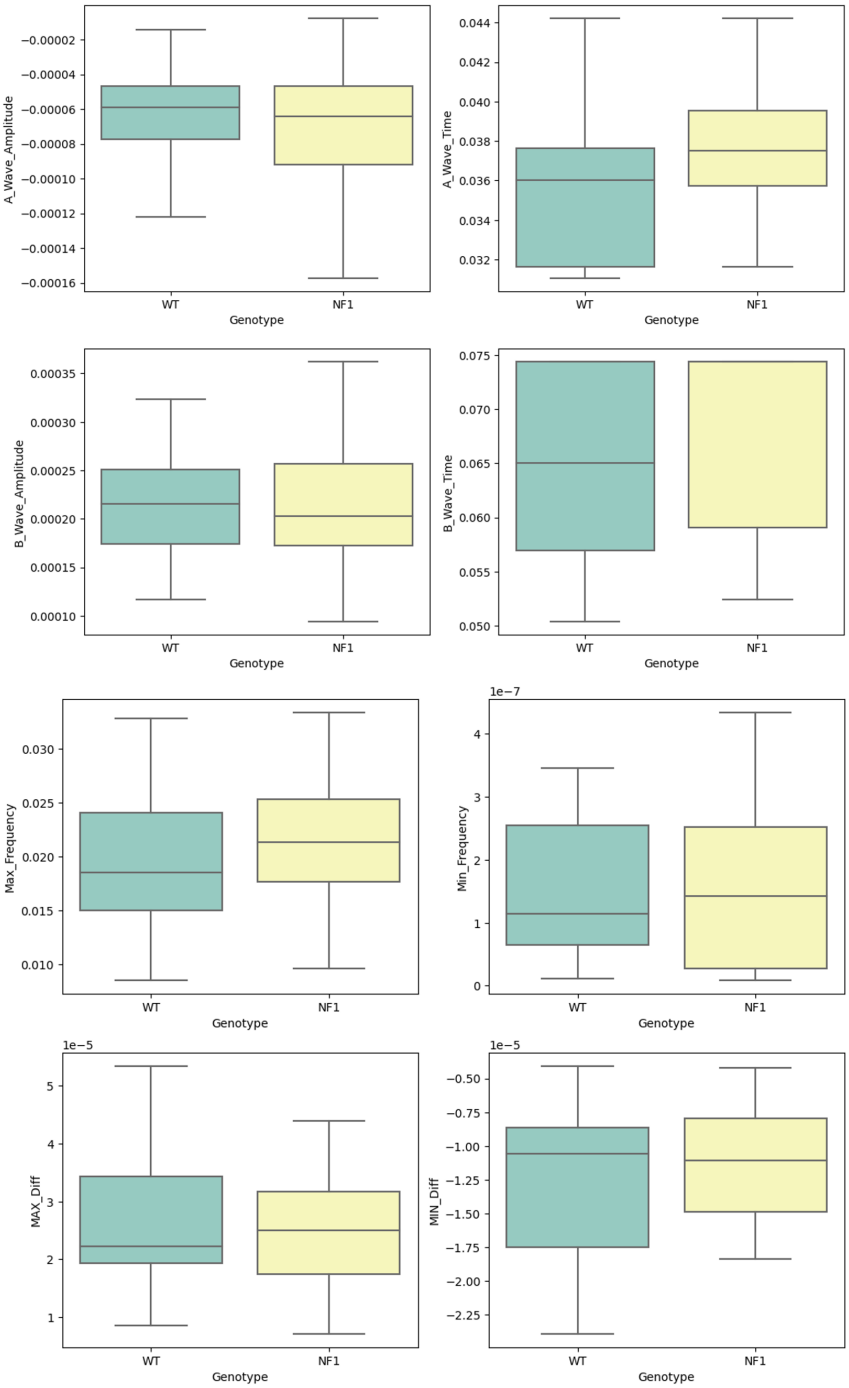


Figure 6.3: Graphic representation of the data grouped by feature as a function of genotype for channel 9.

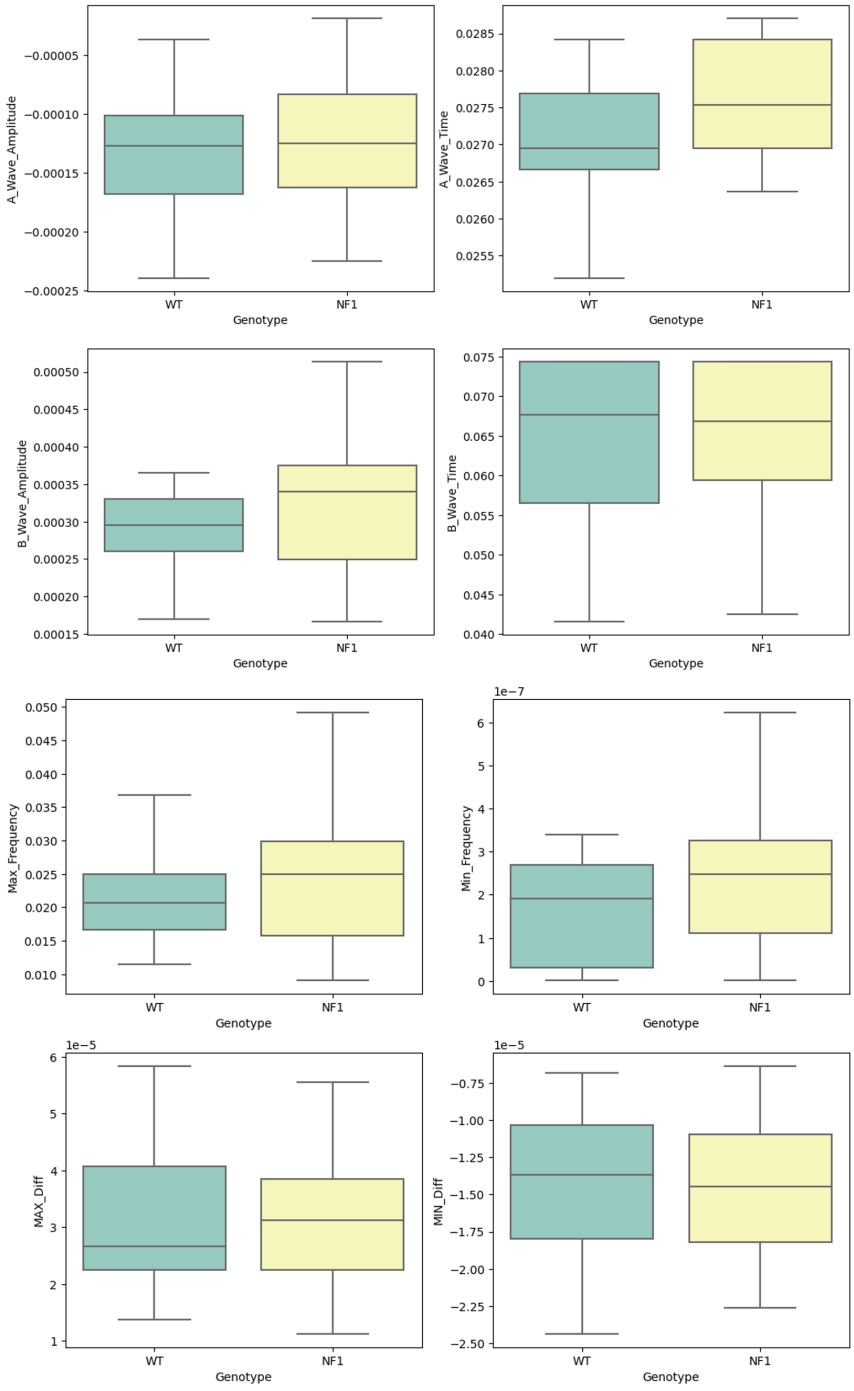


Figure 6.4: Graphic representation of the data grouped by feature as a function of genotype for channel 15.

In summary, the study material for this research can be summarised by table 6.1. It shows the number of files analysed, the features taken into account and the channels used to acquire and study them.

Table 6.1: Summary of the data used in the present investigation.

NF1	WT	Features	Channels Analized
27 files	23 files	8 features	channels 3, 9 and 15

6.2 Statistical Analysis

As established earlier, before statistical analysis can be carried out, a number of assumptions need to be met, of which the following are important to note:

1. **Independence:** Independence stipulates that all groups are mutually exclusive, that is, an individual can only belong to one group. Furthermore, this means that the data contains no repeated measures. This assumption was soon ensured when selecting and collecting the data for the study.

2. **Normality:** Assess whether populations assume a normal distribution. To test this normality assumption we used the Shapiro-Wil test using the shapiro() method from the scipy.stats library.

3. **Homogeneity of variance:** homogeneity of variance assumes that all groups have equal variances. To verify this assumption we used Levene's homogeneity of variance test, which was run using the levene() method from the scipy.stats library [98].

The results for the p-value of the Shapiro-Wil test and Levene's test for homogeneity of variances are shown in tables 6.2, and 6.3, respectively.

Table 6.2: p-value result for the Shapiro Wil Normality Test for the three channels.

Feature	Channel 15	Channel 9	Channel 3
A-Wave Amplitude	0.00768	**0.2011**	0.0163
A-Wave Implicit Time	**0.20806**	3.29E-09	0.0001
B-Wave Amplitude	0.01050	**0.1235**	0.0049
B-Wave Implicit Time	0.00001	3.30E-06	3.11E-10
Maximum Frequency	**0.12426**	**0.6945**	**0.0878**
Minimum Frequency	0.00008	1.78E-05	3.27E-08
Maximum Derivative	0.02546	**0.1553**	3.40E-08
Minimum Derivative	0.00116	0.0042	3.79E-06

In this follow-up, to say that a population is normally distributed the p-value returned from Shapiro's test must be greater than or equal to the alpha value, 0.05.

Table 6.3: p-value result for the Levene's Variance Equality Test for the three channels.

Feature	Channel 15	Channel 9	Channel 3
A-Wave Amplitude	0.2499	0.6775	0.2911
A-Wave Implicit Time	0.4263	0.6239	0.6948
B-Wave Amplitude	0.1070	0.2780	0.5968
B-Wave Implicit Time	0.5279	0.5616	**0.0391**
Maximum Frequency	0.1625	0.8781	0.7791
Minimum Frequency	0.5788	0.9378	0.3886
Maximum Derivative	0.7380	0.8293	0.4700
Minimum Derivative	0.6681	0.9048	0.5777

To infer that the groups meet the variance homogeneity criterion, the p-value returned from Levene's test must be less than or equal to the alpha value, that is, p-value <=0.05.

Thus, and taking into account the values in tables 6.2 and 6.3, it is possible to verify that none of characteristic meets the two assumptions mentioned above. However, it is important to note that the implicit time of the a-wave for the channel 15 meets the normality assumption, as well as, the amplitude of the a-wave (channel 9), the maximum frequency of the signal (channel 15 and 9) and maximum of the first derivative (channel 9). Also, the implicit time of the b-wave for channel 3 meets the variance equality test.

Although the characteristics did not meet the assumptions described above, statistical tests were still applied to the data, taking this into account when interpreting them.

6.2.1 t-test

By definition, the t-test works by checking the means of two samples to infer whether they are significantly different from each other. This is possible by calculating the standard error in the difference between the means. In this way, we can interpret the statistical value in a two-tailed test by comparing it with the positive critical value, that is:

- If |t-statistic| <= critical value: Accept the null hypothesis.
- If |t-statistic| > critical value: Reject the null hypothesis.

In addition, the p-value can also be compared with the significance level, the alpha value, 0.05, to determine whether the null hypothesis can be rejected, translating as:

- If p > alpha: Accept the null hypothesis.
- If p <= alpha: Reject the null hypothesis [99].

In short, in order to have a variable with statistical significance for this test, it should have a p-value less than or equal to the alpha value, and the critical value should be less than the *T-Statistic* (T-Stat) value module.

In the statistical analysis, we have that the null hypothesis claims that both means are statistically equal, on the other hand the alternative hypothesis states that both means are statistically different from each other. Remember the hypotheses previously stipulated:

Null Hypothesis: There are no characteristics with relevant weight that allow the distinction between pathological and non-pathological individuals.

Alternative Hypothesis: There are characteristics with relevant weight that allow the distinction between pathological and non-pathological individuals.

Having said that, the summary of the results obtained for channels 15, 9 and 3, are represented in Tables 6.4, 6.5 and 6.6, respectively, these were obtained, as described in Section 5, using the *scipy.stats* library. Each of these tables consists of 5 columns, one with the features under study, the next with the T-Stat value, the third with the degrees of freedom (DF) value, the fourth with the critical value (CV) and the last with the p-value.

Table 6.4: Results of the t-test for channel 15.

Feature	T-Stat	DF	CV	p-value
Maximum Frequency	-1.211706	48	1.677224	0.231556
Minimum Frequency	-0.659788	48	1.677224	0.512544
Maximum Derivative	-0.176228	48	1.677224	0.860856
Minimum Derivative	-0.044187	48	1.677224	0.964938
A-Wave Amplitude	-0.017795	48	1.677224	0.985876
A-Wave Implicit Time	-2.346728	48	1.677224	**0.023114**
B-Wave Amplitude	-1.190022	48	1.677224	0.239890
B-Wave Implicit Time	-0.160006	48	1.677224	0.873548

Table 6.5: Results of the t-test for channel 9.

Feature	T-Stat	DF	CV	p-value
Maximum Frequency	-0.720817	48	1.677224	0.474518
Minimum Frequency	0.249615	48	1.677224	0.803950
Maximum Derivative	0.407827	48	1.677224	0.685214
Minimum Derivative	-0.214391	48	1.677224	0.831150
A-Wave Amplitude	0.484343	48	1.677224	0.630346
A-Wave Time	-1.050041	48	1.677224	0.298957
B-Wave Amplitude	-0.412019	48	1.677224	0.682160
B-Wave Time	-1.231685	48	1.677224	0.224067

Table 6.6: Results of the t-test for channel 3.

Feature	T-Stat	DF	CV	p-value
Maximum Frequency	0.129201	48	1.677224	0.897738
Minimum Frequency	-0.917081	48	1.677224	0.363685
Maximum Derivative	0.284380	48	1.677224	0.777342
Minimum Derivative	0.629885	48	1.677224	0.531757
A-Wave Amplitude	-0.145529	48	1.677224	0.884902
A-Wave Time	0.555265	48	1.677224	0.581292
B-Wave Amplitude	-0.337966	48	1.677224	0.736862
B-Wave Time	2.290597	48	1.677224	**0.026421**

Therefore, in view of the premise presented above, and analysing the results of the tables, we can verify that only channels 15, higher intensity, and 3 lower intensity, have a variable with statistical importance, although this is different for both:

- **Channel 15 - A-wave Implicit Time:**
  - p-value $\approx 0.023 < 0.05$;
  - T-Stat $\approx$ -2.35; critical value ≈ 1.68; $|-2.35| > 1.68$.
- **Channel 3 - B-wave Implicit Time:**
  - p-value $\approx 0.026 < 0.05$;
  - T-Stat ≈ 2.29; critical value ≈ 1.68; $|2.29| > 1.68$.

So,interpreting the hypotheses presented before and adapting them to the results of the t-test, it is correct to conclude through the same that:

- For channel 15, the implicit time of the a-wave is a feature with relevant weight that allows the distinction between pathological and non-pathological individuals, the remaining features being not statistically significant, since the p-value is approximately 0.023.It is noteworthy that this was also one of the features that in the box plots there was a significant visual difference between the study groups.
- For channel 3, the implicit time of b-wave is a relevant characteristic that allows for the distinction between pathological and non-pathological individuals; the remaining characteristics are not statistically significant, since the p-value is approximately 0.026. This result is in line with the conclusions drawn by the graphical representation of the values in section 6.1.
- For channel 9, there is no characteristic with relevant weight that allows the distinction between pathological and non-pathological individuals, none of the p-values is less or equal to the alpha value.

As such, although a statistically significant feature was found for two of the channels under analysis, the fact that it was not the same for both does not allow us to conclude with any certainty that these features may be distinctive factors in differentiating pathological from non-pathological ERGs.

6.2.2 One-Way ANOVA

The one-way ANOVA examines the means of the groups of interest and checks whether any of the means differ statistically from the others. In particular, it evaluates the null hypothesis:

$$NullHypothesis: \ \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

where μ represents the group mean and k the number of groups. If, however, the one-way ANOVA returns a statistically significant result, we accept the alternative hypothesis, which stipulates that there are at least two group means that are statistically significantly different from each other.

At this point, it is important to remember that the one-way ANOVA is an omnibus test which means that it cannot tell which specific groups are statistically different from each other, but that at least two groups were. To determine which specific groups differ from each other, a *post-hoc* test must be used [94].

In our case study, only two groups are evaluated, so the one-way ANOVA test determines whether for the feature in question there is a significant difference between the two groups, WT and NF1. However, we applied the *post-hoc* test to confirm this difference.

That said, the study hypotheses for the one-way ANOVA test are:

Null Hypothesis:

$$mean\_feature\_x_{NF1} = mean\_feature\_x_{WT}$$

Alternative Hypothesis:

$$mean\_feature\_x_{NF1} \neq mean\_feature\_x_{WT}$$

Where x represents the feature under study.
Whereby the null hypothesis is rejected if:

- alfa <= p-value
- F value <= p-value

Although only one of the characteristics fulfilled the assumptions mentioned before, the one-way ANOVA test and the *post-hoc* test, more specifically the HSD test, were still applied to all the characteristics under study, for the three channels, through the *anova_lm()* and *tukeyhsd()* methods. With this analysis, it was confirmed what had been verified in the previous validation, that the most statistically relevant characteristics would be the implicit time of a-wave (channel 15) and the implicit time of b-wave (channel 3), as shown in Tables 6.7 and 6.8 for the ANOVA test, and 6.9 and 6.10 for the HSD. Since the complete results of these tests for all features are very extensive, only the statistically relevant tables are presented here, the rest can be found in Appendix A.1.

Table 6.7: Summary table of the one-way ANOVA test for 'A-Wave Implicit Time' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000003	1.0	5.50371	0.023154	3.123543e-06
Residual	0.000027	48.0	NaN	NaN	5.675341e-07

Table 6.8: Summary table of the one-way ANOVA test for 'B-Wave Implicit Time' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.001203	1.0	4.501702	0.03905	0.001203
Residual	0.012825	48.0	NaN	NaN	0.000267

The ANOVA table consists of 5 columns, one with the values of the sum of squares, another with the degrees of freedom (DF), a third with the statistical value

of F, the fourth with the p-value and the last with the mean of squares. This table also consists of two rows, one representing the data analyzed in the test, and the other with the residual values. The residuals for the ANOVA model represent the amount of the sum of squares of error that represents the total error in the experiments [100].

Table 6.9: Results of the HSD Test for 'A-Wave Implicit Time' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0005	0.0232	-0.0009	-0.0001	True

Table 6.10: Results of the HSD Test for 'B-Wave Implicit Time' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0098	0.0391	0.0005	0.0192	True

The table of the HSD test, in turn, consists of 7 columns, the first two with the two groups being compared, the next with the value of the difference in means between the two groups (mean_diff), the fourth column with the adjusted p-value (p-adj), the fifth and sixth with the values of the lower and upper bands of the confidence interval, respectively, and finally the value of the rejection or not of the null hypothesis ('True' or 'False' value).

As can be seen, once again, and in agreement with the t-test, through the ANOVA tables and the *post-hoc* test, and taking into account the hipothesis clauses made before, the features that may have statistical significance, and for this reason represent a differentiation factor between individuals with NF1 and WT, are the implicit time of the a-wave for channel 15, and the implicit time of the b-wave for channel 3. This conclusion is possible to be drawn, since in Table 6.7 it is possible to verify that for the implicit time of the a-wave for channel 15 there is a p-value equal to 0.023 and the HSD test has a 'reject' result equal to 'TRUE' which means rejecting the null hypothesis, in the same way, for the amplitude of the b-wave in channel 3 there is a p-value equal to 0.039 and a 'reject' equal to 'TRUE'.

6.3 Classification via Machine Learning

Machine learning is one of the essential and effective tools in data analysis in any area, especially in the health area where it can be applied to improve traditional diagnostic methods. Thus, in our research four ML classifiers were applied in order to complement and confront the statistical analysis previously performed. These

classifiers allow us to infer whether there are substantial differences between groups for the eight features analyzed. This difference is translated into several parameters of score evaluation, but usually the classification is summarized with the accuracy value. This value translates how accurately the algorithm was able to identify a signal correctly as WT or NF1.

Then, as mentioned above, four classifiers were used to classify the data under study through the methods that the *sklearn* library provides. Thus, the data were classified using: the LR using the *LogisticRegression()* method; the SVM with a linear kernel using the *svm.SVC()* method; the NB using the *GaussianNB()* method; and the RF using the *RandomForestClassifier()* method.

For each classifier two subsets were generated from the initial data set, one for classifier training and one for testing. This division was made 80% for training and 20% for testing. One of the parameters that the method of creating the training and test sets has is the *random_state*, which is responsible for fixing and ensuring that the division of the data is always the same. If this parameter was not determined the data would always be randomly split getting different subsets of data every time. With this, to avoid that the results were always alternating, and to ensure that the best possible result was achieved, a study was made of the best *random_state* for each of the classifiers for each channel under analysis. In this study, the *random_state* value was varied between 0 and 1000, keeping the value of the one that achieved the highest accuracy value, in other words, 1000 *random_state* values were tested for each classifier, for each channel, in order to find the one that best fit the data model. To the values found, the values 0 and 42 were added because they are the most commonly used in the literature [101]. That said, the data was classified using all the *random_state* found, as well as the two literature values, as can be seen in tables 6.11, 6.12 and 6.13, where the accuracy results are shown for each of the classification algorithms as a function of the *random_state* used.

The basis of the accuracy calculation by the *accuracy_score()* function is the expression:

$$Accuracy = \frac{NumberofCorrectPredictions}{TotalNumberofPredictions} \tag{6.1}$$

Where the accuracy value is determined by the ratio between the number of cases well predicted by the algorithm and the total number of cases used for the test. These values can also be checked by the confusion matrix.

Table 6.11: Accuracy score of each classifier as a function of *random_state* for channel 15.

random_state =	0	4	42	45	83	300	**Mean**
Logistic Regression	0.50	0.60	0.60	0.50	**0.90**	0.70	**0.63**
Support Vector Machine	0.40	0.60	0.50	0.50	0.70	**0.90**	0.60
Naive Bayes	0.60	**1.00**	0.40	0.60	0.70	0.50	**0.63**
Random Forest	0.50	0.60	0.20	**0.90**	0.50	0.70	0.57
Mean	0.50	**0.70**	0.43	0.63	**0.70**	0.70	

For channel 15, the results of the values for *random_state* were 4, 45, 83 and 300 for the NB, RF, LR and SVM classifiers respectively. In Table 6.11, it is possible to observe that for the optimized *random_state* of each classifier there is none with accuracy lower than 90%, and an accuracy of 100% was even reached for NB. This table also shows that, taking into account an overall average, the classifiers with better results were the NB and LR with an average of 63% and the best *random_state* were the 4 and 83.

Table 6.12: Accuracy score of each classifier as a function of *random_state* for channel 9.

random_state =	0	10	42	57	264	784	**Mean**
Logistic Regression	0.60	**0.80**	0.60	0.30	0.50	0.60	0.57
Support Vector Machine	0.60	0.60	0.50	0.60	0.50	**0.80**	**0.60**
Naive Bayes	0.50	0.50	0.30	0.70	**0.80**	0.70	0.58
Random Forest	0.30	0.40	0.30	**0.80**	0.30	0.60	0.45
Mean	0.50	0.58	0.43	0.60	0.53	**0.68**	

In turn, for channel 9, the accuracy values were a little lower than the previous ones, since all algorithms managed to achieve an accuracy of 80% for their best *random_state*. For this channel, the values for this parameter were 10, 57, 264 and 784, for the LR, RF, NB and SVM classifiers, in that order. For this scenario, the mean values were also lower, with the best classifier being SVM with a mean accuracy of 60%, and the best *random_state* equal to 784 with an accuracy of 68%. As we can see from table 6.12.

Table 6.13: Accuracy score of each classifier as a function of *random_state* for channel 3.

random_state =	0	21	42	152	267	294	**Mean**
Logistic Regression	0.40	0.40	0.40	0.30	**0.90**	**0.90**	0.55
Support Vector Machine	0.50	0.40	0.40	0.40	0.60	**0.90**	0.53
Naive Bayes	0.60	0.50	0.40	**1.00**	0.60	0.90	**0.67**
Random Forest	0.20	**0.90**	0.40	0.40	0.60	0.80	0.55
Mean	0.43	0.55	0.40	0.53	0.80	**0.88**	

Finally, for channel 3, the four *random_state* chosen were 21 for RF, 152 for NB, 267 for LR and 294 for SVM, the latter having the same accuracy value for LR. For channel 3 we have the best average accuracy of the three channels, being this of 67% for the NB classifier, where the best *random_state* reached the value of 80% accuracy.

In the previous tables only the accuracy value for each of the classifiers was presented, however, in Appendix A.2 you can find a more complete summary table for each of the classifiers for each channel. These tables were produced by employing the *classification_report()* method also belonging to the sklearn library. Each of the summary tables contains values such as precision, recall, f1-score and support for the WT group values, for the NF1 group values, for the accuracy value, for the macro average value and for the weighted average value.

In order to complement the data obtained by using the classifiers, a confusion matrix was generated for each of these, for the three channels. The confusion matrix is a very popular measure used in solving classification problems by representing the counts from the predicted values and the real values [102]. In our study, these matrices were obtained using the *confusion_matrix()* function, which receives as parameters the subset generated for testing, as well as the prediction made by the classifier. The confusion matrices created in this research can be seen in section A.2, where it can be observed that for the *random_state* fitted to our models the number of NF1 false or WT false did not exceed the value of 2, and this value occurred for channel 9 for the SVM, where the classifier classified two WT signals as NF1.

Something that is prone to happen when applying once only the *train_test_split()* on a small dataset is model overfitting. Overfitting means that the model that was trained is now overfitted to the training dataset, with this, this model will be very accurate on the training data, but probably not very accurate on the untrained or new data. Basically, when this happens, the model learns or describes the noise in the training data, rather than the actual relationships between the variables in the data [103], this is because the main set is only split once at the beginning of the method, and the training set is always the same. As we see from the high accuracy

values obtained in this study by the *train_test_split()* application, the robustness of its results can be called into question. In order to avoid this, something called *Cross Validation* (CV) can be performed. CV is very similar to the method initially used, however this is applied to more subsets, this is, we divide the data into k subsets, and train on k-1 one of those subsets. The logic is to keep the last subset for testing by doing it for each of the subsets. Within CV there are several other division stratergies like K-fold, Stratified K-Fold, Leave-One-Out, Leave-P-Out or Shuffle Split. Given that we have unbalanced data, that is, there are more NF1 files than WT files, the Stratified K-Fold stratergy was chosen as it is the most adequate methodology for this type of data organisation, since it guarantees that, through the stratification of the target classes (NF1 and WT), both sets will have an equal proportion of all classes.

To apply this method, we used the *StratifiedKFold()* function to create the stratified k value and the *cross_val_score()* function to perform the CV, both from the *sklearn* library. The function *StratifiedKFold()* receives the value of divisions that is intended to be performed, the k value. The *cross_val_score()* function, on its turn, receives as inputs the classifier to be used, the dataset under study, and the "cv" value that represents the CV division strategy to be used, in this case the value of the result from applying the first function. The CV was applied to all the classifiers under study, and three values of k were tested, known as the most common in the application of ML to data science, which are k=3, 5 and 10. The results for this validation are represented in tables 6.14, 6.15 and 6.16, which correspond to k-fold 3, 5 and 10, respectively. Depending on the values of k, the function returns 3, 5 or 10 values in an array of scores, the value that is presented in each cell, is the average of these scores for each channel and each classifier.

Table 6.14: Results of Cross Validation with Stratified K-Fold [k=3].

Classifier	Channel 3	Channel 9	Channel 15	**Mean**
Logistic Regression	0.46	0.50	0.58	**0.51**
Support Vector Machine	0.48	0.42	0.54	0.48
Naive Bayes	**0.59**	0.40	0.50	0.50
Random Forest	0.46	0.38	0.50	0.45
Mean	**0.50**	0.43	0.53	

From the table 6.14 we can see that the classifier with more accuracy, for a k-fold value of 3, is the NB, with a value of 59% for channel 3. It is important to note that, in the previous method, for this same channel, NB was also the best classifier.

Table 6.15: Results of Cross Validation with Stratified K-Fold [k=5].

Classifier	Channel 3	Channel 9	Channel 15	**Mean**
Logistic Regression	0.54	0.50	0.56	0.53
Support Vector Machine	**0.60**	0.52	0.56	**0.56**
Naive Bayes	**0.60**	0.42	0.50	0.51
Random Forest	0.46	0.40	0.40	0.42
Mean	**0.55**	0.46	0.51	

For 5-fold, we have, in general, better accuracy values than for 3-fold, having the same maximum accuracy value of 60% for both SVM and NB classifiers for channel 3.

Table 6.16: Results of Cross Validation with Stratified K-Fold [k=10].

Classifier	Channel 3	Channel 9	Channel 15	**Mean**
Logistic Regression	0.52	0.46	0.52	0.50
Support Vector Machine	0.50	0.38	0.52	**0.57**
Naive Bayes	0.58	0.46	**0.60**	0.55
Random Forest	0.48	0.42	0.46	0.45
Mean	0.52	0.43	**0.53**	

For a value of k equal to 10, the scenario is repeated, with the NB classifier obtaining the best accuracy value, with 60%, this time for channel 15. Having an overview of the three tables, it can be noted that an higher value of k - an higher value of subsets - leads to an increase in the accuracy values, and, in the case under study, the results obtained for k equal to 5 do not differ much from those obtained for k equal to 10.

As can be seen, with the application of CV the accuracy values decrease significantly and come closer to the mean values for the various random_state used in train_test_split(). This is because having more subsets prevents the model from adapting to noise and unrelated elements of the data. Having too little data leads to the training having less accurate results and for that reason, to strengthen the classifiers it would be necessary to use a larger number of data. However, it can be concluded that, for a reduced number of data, CV should be used to infer as to the performance of the algorithm, or else perform several measurements with *train_test_split()* with different *random_state* that guarantee the randomness of data partitioning in the training and test sets.

Finally, in order to evaluate the importance of each feature in the data classification process, we took advantage of the functionality associated with the RF classifier and generated a bar graph illustrating the weight of each feature in the

classification process using that classifier. These graphs are illustrated in figures 6.5, 6.6 and 6.7.The legend of the features represented in the figures can be found in the footnote [2].

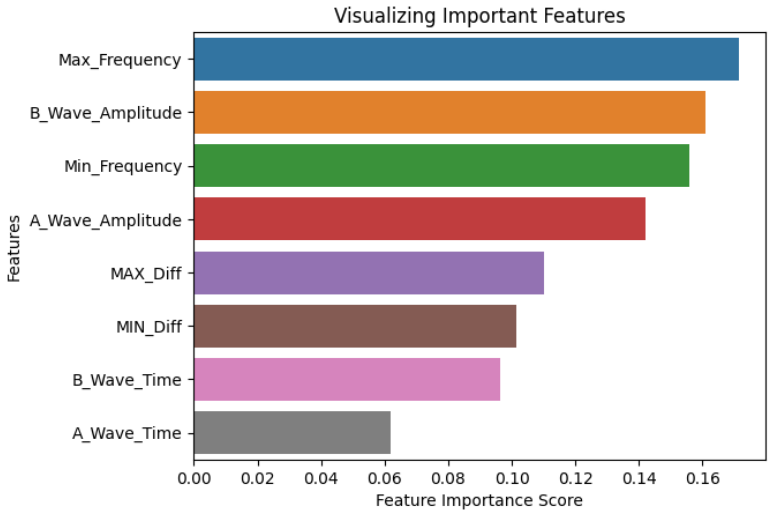


Figure 6.5: Chart with the features used in the RF classifier in order of importance for channel 15.

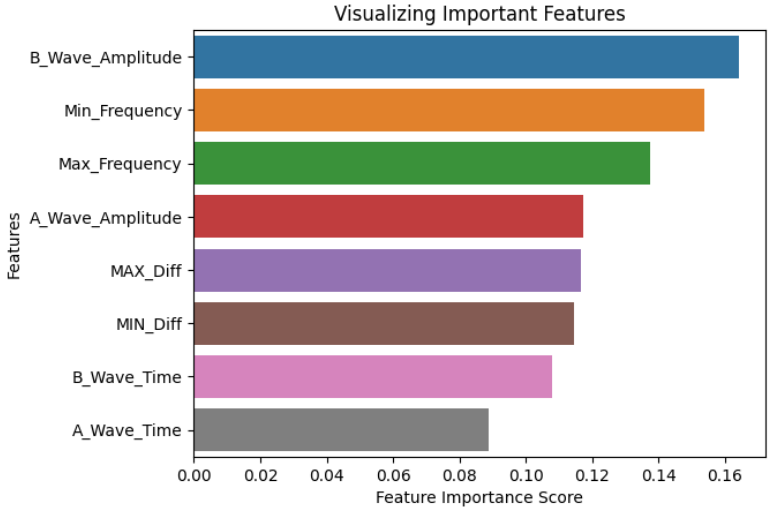


Figure 6.6: Chart with the features used in the RF classifier in order of importance for channel 9.

[2]Max_Frequency: Maximum frequency of the signal (measured in terms of energy); Min_Frequency: Minimum frequency of the signal (measured in terms of energy); Max_Diff: Maximum of the first derivative; Min_Diff: Minimum of the first derivative; B_Wave_Amplitude: B-wave amplitude (measured in Volts); B_Wave_time: B-wave implicit time (measured in seconds); A_Wave_Amplitude: A-wave amplitude (measured in Volts); A_Wave_Time: A-wave implicit time (measured in seconds)

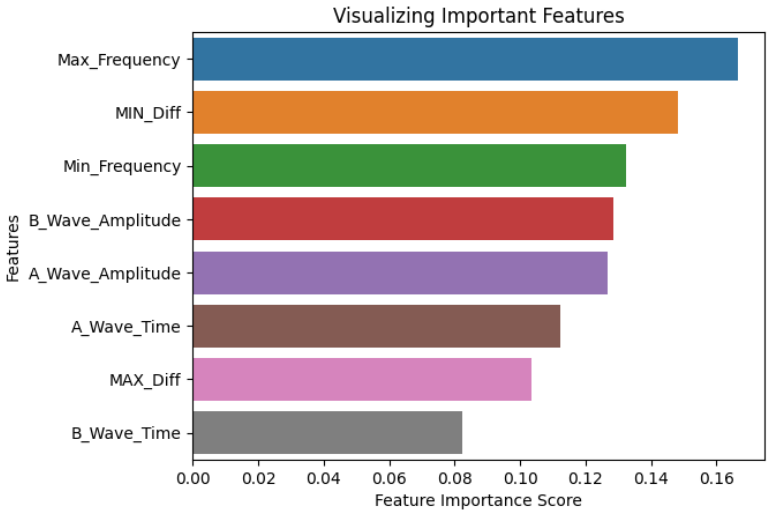


Figure 6.7: Chart with the features used in the RF classifier in order of importance for channel 3.

From the graphs presented above we can see that the features with the most weight for the RF algorithm are the maximum frequency of the signal, appearing in the top features for all three channels, the minimum frequency of the signal and the amplitude of the b-wave appearing twice in the top three.

Having said this, it is possible to conclude, in view of the results obtained for the ML classification, that any of the classifiers can be used for the distinction of pathological signals, and it is important to note that the NB classifier stands out with better results and high accuracy for the both methods used, and can then be pointed out as a potential classifier for the data model used. It is also possible to conclude that the maximum frequency of the signal, as well as the minimum frequency of the signal and the amplitude of the b-wave are characteristics with high distinguishing power between pathological and non-pathological ERGs.

6.4 Creation of the summary file of the analysis made by the algorithm

Throughout the analysis that the program does, a text file is generated and filled with a summary of the same. This file begins by containing the date and time when the program was started, which are used to create the name of the file so that it is easier to identify. Next, it saves the name of the files selected by the user, both the WT and NF1 files. As main data it saves the table resulting from the execution of the t-test, as well as all the tables originated by the application of the one-way ANOVA test. Likewise, for the ML classification, the accuracy values for each of the classification algorithms are registered in the file.

Figure 6.8 and 6.9 show an example of a file generated for channel 15, where the above described components can be observed.

```
Analysis Resume

Analysis Date:2022/10/09 09:54

Wild Type Files: 929.6_M_WT.csv, 930.3_F_WT.csv, 942.0_M_WT.csv, 942.2_M_WT.csv, 942.6_M_WT.csv, 958.4_M_WT.csv, 958.6_M_WT.cs
NF1 Files: 926.5_F_NF.csv, 929.1_M_NF.csv, 929.5_M_NF.csv, 930.0_F_NF.csv, 930.4_F_NF.csv, 942.3_M_NF.csv, 943.1_F_NF.csv, 943


---------------      Statistical Analysis      ---------------

------ T-Test ------

A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups ar

                    T-stat  DF        CV   P-value
Max_Frequency    -1.211706  48  1.677224  0.231556
Min_Frequency    -0.659788  48  1.677224  0.512544
MAX_Diff         -0.176228  48  1.677224  0.860856
MIN_Diff         -0.044187  48  1.677224  0.964938
A_Wave_Amplitude -0.017795  48  1.677224  0.985876
A_Wave_Time      -2.346728  48  1.677224  0.023114
B_Wave_Amplitude -0.160006  48  1.677224  0.873548
B_Wave_Time      -1.190022  48  1.677224  0.239890

T-stat: ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error.
DF: shape of the t-distribution.
CV (Critical Value): defines regions in the sampling distribution of a test statistic. In hypothesis tests, critical values de
P-value: the probability that the results from the sample data occurred by chance

------ One-way ANOVA ------
One-way ANOVA is a hypothesis test that evaluates two mutually exclusive statements about two or more population means.

sum_sq (Sum of the squares): helps express the total variation that can be attributed to various factors.
df: Degrees of Freedom
F: It is calculated by dividing two mean squares. This calculation determines the ratio of explained variance to unexplained v
PR: p-value
mean_sq (Mean of the squares): used to determine whether factors (treatments) are significant
eta_sq: effect size
omega_sq: partial effect size


One-way ANOVA Test - A_Wave_Amplitude ~ C(Genotype)
                 sum_sq    df         F    PR(>F)       mean_sq    eta_sq  omega_sq
C(Genotype)  1.158851e-12   1.0  0.000298  0.986292  1.158851e-12  0.000006 -0.020402
Residual     1.864909e-07  48.0       NaN       NaN  3.885228e-09       NaN       NaN
```

Figure 6.8: Example of the text file generated with the analysis performed by the algorithm.

```
One-way ANOVA Test - B_Wave_Time ~ C(Genotype)
                 sum_sq    df         F    PR(>F)       mean_sq    eta_sq  omega_sq
C(Genotype)  1.282562e-08   1.0  1.326356  0.255159  1.282562e-08  0.026889  0.006485
Residual     4.641514e-07  48.0       NaN       NaN  9.669822e-09       NaN       NaN

One-way ANOVA Test - Max_Frequency ~ C(Genotype)
                 sum_sq    df         F    PR(>F)    mean_sq    eta_sq  omega_sq
C(Genotype)  0.000113   1.0  1.401492  0.242304  0.000113  0.028369  0.007966
Residual     0.003872  48.0       NaN       NaN  0.000081       NaN       NaN

One-way ANOVA Test - Min_Frequency ~ C(Genotype)
                 sum_sq    df         F    PR(>F)       mean_sq    eta_sq  omega_sq
C(Genotype)  1.719975e-14   1.0  0.458378  0.501633  1.719975e-14  0.009459 -0.010951
Residual     1.801108e-12  48.0       NaN       NaN  3.752308e-14       NaN       NaN

One-way ANOVA Test - MAX_Diff ~ C(Genotype)
                 sum_sq    df         F    PR(>F)       mean_sq    eta_sq  omega_sq
C(Genotype)  5.504834e-12   1.0  0.030509  0.862074  5.504834e-12  0.000635 -0.019773
Residual     8.660674e-09  48.0       NaN       NaN  1.804307e-10       NaN       NaN

One-way ANOVA Test - MIN_Diff ~ C(Genotype)
                 sum_sq    df         F    PR(>F)       mean_sq    eta_sq  omega_sq
C(Genotype)  8.157188e-14   1.0  0.00196  0.964875  8.157188e-14  0.000041 -0.020367
Residual     1.998131e-09  48.0       NaN       NaN  4.162773e-11       NaN       NaN


---------------      Classification via Machine Learning      ---------------

Accuracy score of Logistic Regression:
0.9

Accuracy score of Support Vector Machine:
0.9

Score of Naive Bayes:
1.0

Accuracy of Random Forest:
0.9
```

Figure 6.9: Example of the text file generated with the analysis performed by the algorithm (Continued).

Having presented and discussed the results obtained in the present investigation, it is important to retain the following points:

1. Through the representation of the box plots, differences between the two genotypes are noted for the features implicit time b-wave (channel 3 and 15), the implicit time a-wave (channel 9 and 15), and the minimum frequency (channel 15).

2. Both the t-test and the one-way ANOVA test allowed to statistically evaluate the data model.

3. The statistical analysis allowed the conclusion that the implicit time of the a-wave and the implicit time of the b-wave may be characteristics of the ERG signals capable of distinguishing healthy from unhealthy individuals.

4. With the correct parameterization, namely by using a suitable random state, any of the four classifiers can be used for the classification and distinction of pathological and non-pathological ERGs.

5. All algorithms were able to achieve an accuracy of at least 90%, when used the optimal random_state for a single application of *train_test_split()*.

6. Using cross-validation, maximum accuracy values of 60% were obtained, and for the average of the various *random_state* values, values of 67% were reached.

7. Although the accuracy values are lower, it is good practice to apply cross-validation or several repetitions of *train_test_split()* to ensure that the model is not overffited, this when the number of data under analysis is small.

8. The algorithm with the best mean accuracy was Naive Bayes, and it can be pointed as the best classifier for the case under study.

9. The weighting charts of features used by the classifiers indicate that the most important features are the amplitude of the b-wave, the maximum frequency of the signal and the minimum frequency of the signal.

10. Of the channels evaluated, the ones that showed the best results were 15 and 3, and can thus be considered as preference channels when analyzing the ERG.

11. The use of the ERG is a possible tool for diagnosing patients with NF1.

Chapter 7

Final Considerations

Neurofibromatosis type 1 is an inherited autosomal dominant condition and is known to affect all organ systems, including the visual system. Currently, traditional diagnostic methods lead to the late detection of this disease, delaying the beginning of the treatment process and patient follow-up. It was in order to improve the current diagnostic systems that the research presented throughout this document emerged. In fact, the main objective of this thesis was to evaluate how a precise distinction between WT and NF1 genotypes would be possible through an ERG, that is, how to correctly classify and diagnose an individual with NF1 disease.

Thus, we started by selecting a set of features that could represent distinctive factors between the two genotypes, which were: the amplitude of waves a and b, the implicit time of waves a and b, the maximum and minimum frequencies of the signal and the maximum and minimum of the first derivative of the signal, the latter representing the inclinations in the original signal. These features were used to investigate their importance for the study in question, using two methodologies, statistical analysis and classification using ML.

Two tests were used in the statistical analysis: the independent t-test and the one-way ANOVA. Both were applied to each feature and the results were consistent with each other, since they revealed the same characteristics with a statistically significant p-value, which were the implicit time of a-wave and the implicit time of b-wave , with p-values of 0.023 and 0.026, respectively. Thus, these are relevant characteristics when comparing pathological ERGs with non-pathological ERGs.

Subsequently, and in order to complement the statistical analysis, machine learning techniques were used, and the classifiers Logistic Regression, Support-Vector Machine, Naive Bayes and Random Forest were applied to the data under study. From this classification process, it was possible to obtain for the appropriate *random_state* and applying one single time the **train_test_split()**, accuracy values varying between 80% and 100%, and the classifier that presented the best accuracy mean, and therefore the best performance, was the Naive Bayes with a maximum mean of 67%. As for the CV, 60% accuracy values were obtained, also for the NB classifier. Still within the classification scope, the weight of each feature for the RF classifier was measured, which allowed us to conclude that the features with the greatest importance for this classification were the maximum signal frequency, the amplitude of the b-wave and the amplitude of the a-wave. As a final conclusion, it is possible to state that the use of ERG is a viable tool for the diagnosis of patients with NF1, and can be an added value in its association with traditional methods.

Therefore, in general, it can be concluded that the set of features chosen for the research were adequate, as well as the methods of data analysis and the same, however, this research can be enriched with the addition of more and varied features that allow a more accurate and precise distinction. Increasing the number of data being studied is also a great asset to enrich the study, particularly when using machine learning techniques, where the more data, the better the output of the classifiers. Finally, something that would be of relevance to add to this research would be the possibility, through the introduction of a single ERG, to have its output classified as NF1 or WT.

To conclude, it is also important to note that, although this entire study has been applied to the NF1 pathology, the combination of ML with ERG is still very little studied and scarce in the current state of the art, however, just as it shows to have potentiality to be used as a diagnostic tool for NF1, interesting results may also arise for other diseases, thus requiring further exploration of these tools.

References

[1] V. Nejati, L. Moradkhani, S. Suggate, and P. Jansen, "The impact of visual-spatial abilities on theory of mind in children and adolescents with autism spectrum disorder," *Research in Developmental Disabilities*, vol. 114, p. 103960, July 2021. [Cited on page 1]

[2] S. Dakin and U. Frith, "Vagaries of visual perception in autism," *Neuron*, vol. 48, pp. 497–507, Nov. 2005. [Cited on page 1]

[3] R. A. Seymour, G. Rippon, G. Gooding-Williams, J. M. Schoffelen, and K. Kessler, "Dysregulated oscillatory connectivity in the visual system in autism spectrum disorder," *Brain*, vol. 142, pp. 3294–3305, Aug. 2019. [Cited on page 1]

[4] M. B. Usta, "Prediction of outcome in children with autism spectrum disorders," in *Neural Engineering Techniques for Autism Spectrum Disorder*, pp. 1–8, Elsevier, 2021. [Cited on page 3]

[5] H. Hodges, C. Fealko, and N. Soares, "Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation," *Translational Pediatrics*, vol. 9, pp. S55–S65, Feb. 2020. [Cited on pages 3 and 4]

[6] R. A. Muhle, H. E. Reed, K. A. Stratigos, and J. Veenstra-VanderWeele, "The emerging clinical neuroscience of autism spectrum disorder," *JAMA Psychiatry*, vol. 75, p. 514, May 2018. [Cited on page 3]

[7] C. Lord, M. Elsabbagh, G. Baird, and J. Veenstra-Vanderweele, "Autism spectrum disorder," *The Lancet*, vol. 392, pp. 508–520, Aug. 2018. [Cited on page 3]

[8] L. Mottron, "Autism spectrum disorder," in *Neurocognitive Development: Disorders and Disabilities*, pp. 127–136, Elsevier, 2020. [Cited on page 4]

[9] S. L. Hyman, S. E. Levy, S. M. Myers, D. Z. Kuo, S. Apkon, L. F. Davidson, K. A. Ellerbeck, J. E. Foster, G. H. Noritz, M. O. Leppert, B. S. Saunders, C. Stille, L. Yin, C. C. Weitzman, D. O. Childers, J. M. Levine, A. M. Peralta-Carcelen, J. K. Poon, P. J. Smith, N. J. Blum, J. I. Takayama, R. Baum, R. G. Voigt, and C. B. and, "Identification, evaluation, and management of children

with autism spectrum disorder," *Pediatrics*, vol. 145, Jan. 2020. [Cited on page 4]

[10] R. Coulter, "Understanding the Visual Symptoms of Individuals with Autism Spectrum Disorder (ASD).," *Optometry and vision development*, vol. 40, no. 3, p. 164, 2009. [Cited on page 4]

[11] J. O. Phillips, D. E. Parker, C. M. Jacobs, R. J. Groen, A. H. Weiss, and S. J. Webb, "Oculomotor and vestibular performance in children with autism spectrum disorder.," *Investigative Ophthalmology & Visual Science*, vol. 50, pp. 2883–2883, 2009. [Cited on page 4]

[12] A. Bakroon and V. Lakshminarayanan, "Visual function in autism spectrum disorders: a critical review," *Clinical and Experimental Optometry*, vol. 99, pp. 297–308, July 2016. [Cited on page 4]

[13] D. S. Manoach, "Deficient saccadic inhibition in asperger's disorder and the social-emotional processing disorder," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, pp. 1719–1726, Dec. 2004. [Cited on page 4]

[14] R. L. Schulman, "Optometry's role in the treatment of autism.," *Journal of Optometric Vision Development*, vol. 25, no. 4, pp. 259–268, 1994. [Cited on page 4]

[15] L. Mottron, M. Dawson, I. Soulières, B. Hubert, and J. Burack, "Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception," *Journal of Autism and Developmental Disorders*, vol. 36, pp. 27–43, Jan. 2006. [Cited on page 4]

[16] H. C. Koh, E. Milne, and K. Dobkins, "Spatial contrast sensitivity in adolescents with autism spectrum disorders," *Journal of Autism and Developmental Disorders*, vol. 40, pp. 978–987, Mar. 2010. [Cited on page 5]

[17] C. Daluwatte, J. H. Miles, S. E. Christ, D. Q. Beversdorf, T. N. Takahashi, and G. Yao, "Atypical pupillary light reflex and heart rate variability in children with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 43, pp. 1910–1925, Dec. 2012. [Cited on page 5]

[18] S. F. A. Júnior, G. Zanetti, A. S. A. de Melo, A. S. Souza, L. S. Souza, G. de Souza Portes Meirelles, K. L. Irion, B. Hochhegger, and E. Marchiori, "Neurofibromatosis type 1: State-of-the-art review with emphasis on pulmonary involvement," *Respiratory Medicine*, vol. 149, pp. 9–15, Mar. 2019. [Cited on page 5]

[19] B. N. Wilson, A. M. John, M. Z. Handler, and R. A. Schwartz, "Neurofibromatosis type 1: New developments in genetics and treatment," *Journal of the American Academy of Dermatology*, vol. 84, pp. 1667–1676, June 2021. [Cited on page 5]

[20] A. C. Hirbe and D. H. Gutmann, "Neurofibromatosis type 1: a multidisciplinary approach to care," *The Lancet Neurology*, vol. 13, pp. 834–843, Aug. 2014. [Cited on page 5]

[21] T. Karaconji, E. Whist, R. V. Jamieson, M. P. Flaherty, and J. R. B. Grigg, "Neurofibromatosis type 1: Review and update on emerging therapies," *Asia-Pacific journal of ophthalmology (Philadelphia, Pa.)*, vol. 8, no. 1, p. 62—72, 2019. [Cited on page 5]

[22] N. Julian, N. E. Edwards, S. DeCrane, and C. M. Hingtgen, "Neurofibromatosis 1: Diagnosis and management," *The Journal for Nurse Practitioners*, vol. 10, pp. 30–35, Jan. 2014. [Cited on pages 5 and 6]

[23] M. Brahmi, P. Thiesse, D. Ranchere, T. Mognetti, S. Pinson, C. Renard, A.-V. Decouvelaere, J.-Y. Blay, and P. Combemale, "Diagnostic accuracy of pet/ct-guided percutaneous biopsies for malignant peripheral nerve sheath tumors in neurofibromatosis type 1 patients," *PloS one*, vol. 10, p. e0138386, 10 2015. [Cited on pages vii and 6]

[24] N. Hirsch, A. Murphy, and J. Radcliffe, "Neurofibromatosis: clinical presentations and anaesthetic implications," *British Journal of Anaesthesia*, vol. 86, pp. 555–564, Apr. 2001. [Cited on page 5]

[25] R. E. Ferner, S. M. Huson, N. Thomas, C. Moss, H. Willshaw, D. G. Evans, M. Upadhyaya, R. Towers, M. Gleeson, C. Steiger, and A. Kirby, "Guidelines for the diagnosis and management of individuals with neurofibromatosis 1," *Journal of Medical Genetics*, vol. 44, pp. 81–88, Aug. 2006. [Cited on page 6]

[26] J. W. Chan, "Neuro-ophthalmic features of the neurocutaneous syndromes," *International Ophthalmology Clinics*, vol. 52, no. 3, pp. 73–85, 2012. [Cited on page 6]

[27] K. P. Boyd, B. R. Korf, and A. Theos, "Neurofibromatosis type 1," *Journal of the American Academy of Dermatology*, vol. 61, pp. 1–14, July 2009. [Cited on page 6]

[28] E.-M. Jouhilahti, S. Peltonen, A. M. Heape, and J. Peltonen, "The pathoetiology of neurofibromatosis 1," *The American Journal of Pathology*, vol. 178, pp. 1932–1939, May 2011. [Cited on page 6]

[29] H. D.H., *Eye, Brain, and Vision*. Scientific American Library/Scientific American Books, 1995. [Cited on page 6]

[30] A. O. Eghrari, S. A. Riazuddin, and J. D. Gottsch, "Overview of the cornea," in *Progress in Molecular Biology and Translational Science*, pp. 7–23, Elsevier, 2015. [Cited on page 7]

[31] H. Kolb, *Gross Anatomy of the Eye*. University of Utah Health Sciences Center, Salt Lake City (UT), 1995. [Cited on pages vii, 7, and 8]

[32] J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman, "Anatomy of the eye and orbit," in *The Eye*, pp. 1–102.e2, Elsevier, 2016. [Cited on page 7]

[33] J. Selhorst and Y. Chen, "The optic nerve," *Seminars in Neurology*, vol. 29, pp. 029–035, Feb. 2009. [Cited on page 7]

[34] J. Zhu, E. Zhang, and K. Rio-Tsonis, "Eye anatomy," Nov. 2012. [Cited on page 8]

[35] L. Frishman, "Electroretinography," in *Reference Module in Neuroscience and Biobehavioral Psychology*, Elsevier, 2017. [Cited on page 8]

[36] T. J. Lewis and C. L. Trempe, "A new diagnostic paradigm," in *The End of Alzheimer's*, pp. 130–191, Elsevier, 2017. [Cited on page 8]

[37] A. J. Green, "Visual evoked potentials, electroretinography, and other diagnostic approaches to the visual system," in *Aminoff's Electrodiagnosis in Clinical Neurology*, pp. 477–503, Elsevier, 2012. [Cited on page 9]

[38] D. L. McCulloch, M. F. Marmor, M. G. Brigell, R. Hamilton, G. E. Holder, R. Tzekov, and M. Bach, "ISCEV standard for full-field clinical electroretinography (2015 update)," *Documenta Ophthalmologica*, vol. 130, pp. 1–12, Dec. 2014. [Cited on page 9]

[39] S. Asanad and R. Karanjia, *Full-Field Electroretinogram*. StatPearls Publishing, Dec. 2021. [Cited on pages vii and 9]

[40] I. S. Samuels, G. M. Sturgill, G. H. Grossman, M. E. Rayborn, J. G. Hollyfield, and N. S. Peachey, "Light-evoked responses of the retinal pigment epithelium: Changes accompanying photoreceptor loss in the mouse," *Journal of Neurophysiology*, vol. 104, pp. 391–402, July 2010. [Cited on page 9]

[41] J. W. Chan, "Electrophysiology of vision: Clinical testing and applications," *Journal of Neuro-Ophthalmology*, vol. 26, p. 301, Dec. 2006. [Cited on page 9]

[42] M. Moschos and D. Brouzas, "C wave of electroretinogram and visual evoked response in optic neuritis due to demyelinating diseases," *Ophthalmologica*, vol. 204, no. 3, pp. 149–155, 1992. [Cited on page 10]

[43] C.-J. DONG, P. AGEY, and W. A. HARE, "Origins of the electroretinogram oscillatory potentials in the rabbit retina," *Visual Neuroscience*, vol. 21, pp. 533–543, July 2004. [Cited on page 10]

[44] A. Mohammad Alaql, *Analysis and Processing of Human Electroretinogram.* PhD thesis, University of South Florida Scholar Commons, 2016. [Cited on pages ix, 10, and 13]

[45] M. Azarmina, "Full-field versus multifocal electroretinography," *J. Ophthalmic Vis. Res.*, vol. 8, pp. 191–192, July 2013. [Cited on page 10]

[46] E. L. Berson, P. Gouras, and M. Hoff, "Temporal aspects of the electroretinogram," *Archives of Ophthalmology*, vol. 81, pp. 207–214, Feb. 1969. [Cited on page 10]

[47] A. G. Fernandes, S. R. Salomão, J. M. Pereira, and A. Berezovsky, "Full-field electroretinogram recorded with skin electrodes in normal adults," *Arquivos Brasileiros de Oftalmologia*, vol. 79, no. 6, 2016. [Cited on page 10]

[48] A. G. Robson, J. Nilsson, S. Li, S. Jalali, A. B. Fulton, A. P. Tormene, G. E. Holder, and S. E. Brodie, "ISCEV guide to visual electrodiagnostic procedures," *Documenta Ophthalmologica*, vol. 136, pp. 1–26, Feb. 2018. [Cited on page 10]

[49] G. Holder, A. Robson, C. Hogg, M. Kurz-Levin, N. Lois, and A. Bird, "Pattern erg: clinical overview, and some observations on associated fundus autofluorescence imaging in inherited maculopathy," *Documenta Ophthalmologica*, vol. 106, no. 1, pp. 17–23, 2003. [Cited on page 11]

[50] X. Luo and L. J. Frishman, "Retinal pathway origins of the pattern electroretinogram (PERG)," *Investigative Opthalmology & Visual Science*, vol. 52, p. 8571, Nov. 2011. [Cited on page 11]

[51] G. E. Holder, "Pattern electroretinography (PERG) and an integrated approach to visual pathway diagnosis," *Progress in Retinal and Eye Research*, vol. 20, pp. 531–561, July 2001. [Cited on page 11]

[52] G. E. Holder, "Electrophysiological assessment of optic nerve disease," *Eye*, vol. 18, pp. 1133–1143, Nov. 2004. [Cited on page 11]

[53] M. Dettoraki and M. M. Moschos, "The role of multifocal electroretinography in the assessment of drug-induced retinopathy: A review of the literature," *Ophthalmic Research*, vol. 56, no. 4, pp. 169–177, 2016. [Cited on page 11]

[54] D. C. Hood, J. G. Odel, C. S. Md; Chen, and J. Bryan, "BA the multifocal electroretinogram," *Journal of Neuro-Ophthalmology*, vol. 23, pp. 225–235, 2003. [Cited on page 12]

[55] M. Yanoff, J. Duker, and J. Augsburger, *Ophthalmology*. An expert consult title online + print, Mosby Elsevier, 2009. [Cited on page 12]

[56] Mark S. Humayun, J. D. Weiland, G. Chader, and E. Greenbaum, *Artificial Sight*. Biological and Medical Physics, Biomedical Engineering, New York, NY: Springer New York, 2007. [Cited on page 12]

[57] N. Pasmanter and S. M. Petersen-Jones, "A review of electroretinography waveforms and models and their application in the dog," *Veterinary Ophthalmology*, vol. 23, pp. 418–435, Mar. 2020. [Cited on pages 12 and 13]

[58] F. Carpi and F. Tomei, "Non-invasive electroretinography," *Biomedicine & Pharmacotherapy*, vol. 60, pp. 375–379, Sept. 2006. [Cited on page 12]

[59] K. Wolpert and S. Tsang, "Electroretinography," in *Electroretinograms*, InTech, Aug. 2011. [Cited on page 13]

[60] O. R. Marmoy and S. Viswanathan, "Clinical electrophysiology of the optic nerve and retinal ganglion cells," *Eye*, vol. 35, pp. 2386–2405, June 2021. [Cited on page 13]

[61] S. Behbahani, H. Ahmadieh, and S. Rajan, "Feature extraction methods for electroretinogram signal analysis: A review," *IEEE Access*, vol. 9, pp. 116879–116897, 2021. [Cited on page 13]

[62] R. Barraco, D. P. Adorno, M. Brai, and L. Tranchina, "A comparison among different techniques for human ERG signals processing and classification," *Physica Medica*, vol. 30, pp. 86–95, Feb. 2014. [Cited on page 13]

[63] M. Yu, D. Creel, and A. Iannaccone, *Handbook of clinical electrophysiology of vision*. Springer Nature Switzerland, 2019. [Cited on page 14]

[64] P. Algvere and S. Westbeck, "Human ERG in response to double flashes of light during the course of dark adaptation: A fourier analysis of the oscillatory potentials," *Vision Research*, vol. 12, no. 2, pp. 195–214, 1972. [Cited on page 16]

[65] S. A. BURNS, A. E. ELSNER, and M. R. KREITZ, "Analysis of nonlinearities in the flicker ERG," *Optometry and Vision Science*, vol. 69, pp. 95–105, Feb. 1992. [Cited on page 16]

[66] M. Rufiange, J. Dassa, O. Dembinska, R. K. Koenekoop, J. M. Little, R. C. Polomeno, M. Dumont, S. Chemtob, and P. Lachapelle, "The photopic ERG luminance-response function (photopic hill): method of analysis and clinical application," *Vision Research*, vol. 43, pp. 1405–1412, June 2003. [Cited on page 17]

[67] M. F. Marmor, A. Serrato, and R. Tzekov, "Comparison of conventional ERG parameters and high-intensity A-wave analysis in a clinical setting," *Documenta Ophthalmologica*, vol. 106, no. 3, pp. 281–287, 2003. [Cited on page 17]

[68] A. Richards, A. A. Emondi, and B. Rohrer, "Long-term ERG analysis in the partially light-damaged mouse retina reveals regressive and compensatory changes," *Visual Neuroscience*, vol. 23, no. 1, pp. 91–97, 2006. [Cited on page 18]

[69] R. Barraco, L. Bellomonte, M. Brai, and M. Anastasi, "Analysis of the human a-wave ERG component," *Physiological Measurement*, vol. 27, pp. 881–899, July 2006. [Cited on page 18]

[70] R. Barraco, D. P. Adorno, and M. Brai, "ERG signal analysis using wavelet transform," *Theory in Biosciences*, vol. 130, pp. 155–163, Apr. 2011. [Cited on page 18]

[71] D. Devos, M. Tir, C. A. Maurage, N. Waucquier, L. Defebvre, S. Defoort-Dhellemmes, and A. Destee, "ERG and anatomical abnormalities suggesting retinopathy in dementia with lewy bodies," *Neurology*, vol. 65, pp. 1107–1110, Oct. 2005. [Cited on page 19]

[72] R. Barraco, D. P. Adorno, M. Brai, and L. Tranchina, "A comparison among different techniques for human ERG signals processing and classification," *Physica Medica*, vol. 30, pp. 86–95, Feb. 2014. [Cited on page 19]

[73] G. Benchorin, M. Calton, M. Beaulieu, and D. Vollrath, "Assessment of Murine Retinal Function by Electroretinography," *Bio-Protocol*, vol. 7, no. 7, 2017. [Cited on page 19]

[74] T. Wright, P. Yan, and M. Easterbrook, "Machine learning to identify multifocal ERG deficits in patients taking hydroxychloroquine.," *Investigative Ophthalmology & Visual Science*, vol. 60, p. 5959, jul 2019. [Cited on page 20]

[75] A. López-Dorado, J. Pérez, M. Rodrigo, J. Miguel-Jiménez, M. Ortiz, L. de Santiago, E. López-Guillén, R. Blanco, C. Cavalliere, E. M. S. Morla, L. Boquete, and E. Garcia-Martin, "Diagnosis of multiple sclerosis using multifocal ERG data feature fusion," *Information Fusion*, vol. 76, pp. 157–167, Dec. 2021. [Cited on page 20]

[76] T. Diao, F. Kushzad, M. D. Patel, M. P. Bindiganavale, M. Wasi, M. J. Kochenderfer, and H. E. Moss, "Comparison of machine learning approaches to improve diagnosis of optic neuropathy using photopic negative response measured using a handheld device," *Frontiers in Medicine*, vol. 8, Dec. 2021. [Cited on page 20]

[77] T. Schwitzer, S. L. Cam, E. Cosker, H. Vinsard, A. Leguay, K. Angioi-Duprez, V. Laprevote, R. Ranta, R. Schwan, and V. L. Dorr, "Retinal electroretinogram features can detect depression state and treatment response in adults: A machine learning approach," *Journal of Affective Disorders*, vol. 306, pp. 208–214, June 2022. [Cited on page 21]

[78] R. Ritvo, D. Creel, D. Ph, A. S. Crandall, B. J. Freeman, and C. Ogden, "Electroretinograms in Autism : A Pilot Study of b-Wave Amplitudes," *Psychiatry: Interpersonal and Biological Processes*, no. February, pp. 229–232, 1988. [Cited on page 21]

[79] P. A. Constable, S. B. Gaigg, D. M. Bowler, H. Jägle, and D. A. Thompson, "Full-field electroretinogram in autism spectrum disorder," *Documenta Ophthalmologica*, vol. 132, pp. 83–99, Feb. 2016. [Cited on page 21]

[80] L. N. Thibos, N. Lopez-Gil, and A. Bradley, "What is a troland?," *Journal of the Optical Society of America A*, vol. 35, p. 813, Apr. 2018. [Cited on page 22]

[81] P. A. Constable, E. R. Ritvo, A. R. Ritvo, I. O. Lee, M. L. McNair, D. Stahl, J. Sowden, S. Quinn, D. H. Skuse, D. A. Thompson, and J. C. McPartland, "Light-adapted electroretinogram differences in autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 50, pp. 2874–2885, Feb. 2020. [Cited on page 22]

[82] P. A. Constable, I. O. Lee, F. Marmolejo-Ramos, D. H. Skuse, and D. A. Thompson, "The photopic negative response in autism spectrum disorder," *Clinical and Experimental Optometry*, vol. 104, pp. 841–847, Apr. 2021. [Cited on page 22]

[83] R. C. S. &. F. GmbH, "RETI-port/scan 21." [Cited on page 24]

[84] A. J. Silva, P. W. Frankland, Z. Marowitz, E. Friedman, G. Lazlo, D. Cioffi, T. Jacks, and R. Bourtchuladze, "A mouse model for the learning and memory

deficits associated with neurofibromatosis type i," *Nature Genetics*, vol. 15, pp. 281–284, Mar. 1997. [Cited on page 25]

[85] R. D. Gunkel, D. R. Bergsma, and P. Gouras, "A ganzfeld stimulator for electroretinography," *Archives of Ophthalmology*, vol. 94, pp. 669–670, Apr. 1976. [Cited on page 25]

[86] S. G. Rosolen, F. Rigaudière, J.-F. L. Gargasson, and M. G. Brigell, "Recommendations for a toxicological screening ERG procedure in laboratory animals," *Documenta Ophthalmologica*, vol. 110, pp. 57–66, Jan. 2005. [Cited on page 25]

[87] E. C. Fein, J. Gilmour, T. Machin, and L. Hendry, "Section 6.2: One-Way ANOVA Assumptions, Interpretation, and Write Up," jun 2022. [Cited on page 33]

[88] T. W. MacFarland and J. M. Yates, "Student's t-test for independent samples," in *Using R for Biostatistics*, pp. 141–240, Springer International Publishing, Oct. 2020. [Cited on page 33]

[89] J. Brownlee, *Statistical Methods for Machine Learning: Discover how to Transform Data into Knowledge with Python*. Machine Learning Mastery, 2018. [Cited on page 33]

[90] S. Lee and D. K. Lee, "What is the proper way to apply the multiple comparison test?," *Korean Journal of Anesthesiology*, vol. 71, pp. 353–360, Oct. 2018. [Cited on page 34]

[91] P. Mishra, U. Singh, C. Pandey, P. Mishra, and G. Pandey, "Application of student's t-test, analysis of variance, and covariance," *Annals of Cardiac Anaesthesia*, vol. 22, no. 4, p. 407, 2019. [Cited on page 34]

[92] B. Rosner, *Fundamentals of Biostatistics*. Cengage Learning, 2015. [Cited on page 35]

[93] D. Bzdok, N. Altman, and M. Krzywinski, "Statistics versus machine learning," *Nature Methods*, vol. 15, pp. 233–234, Apr. 2018. [Cited on page 35]

[94] D. Bzdok, "Classical statistics and statistical learning in imaging neuroscience," *Frontiers in Neuroscience*, vol. 11, Oct. 2017. [Cited on pages 35 and 49]

[95] G. Biagetti, P. Crippa, L. Falaschetti, G. Tanoni, and C. Turchetti, "A comparative study of machine learning algorithms for physiological signal classification," *Procedia Computer Science*, vol. 126, pp. 1977–1984, 2018. [Cited on page 37]

[96] Y. Xie and S. Oniga, "A review of processing methods and classification algorithm for EEG signal," *Carpathian Journal of Electronic and Computer Engineering*, vol. 13, pp. 23–29, Sept. 2020. [Cited on page 37]

[97] P. Yildirim, K. U. Birant, V. Radevski, A. Kut, and D. Birant, "Comparative analysis of ensemble learning methods for signal classification," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, IEEE, May 2018. [Cited on page 37]

[98] R. Ott and M. Longnecker, *An Introduction to Statistical Methods and Data Analysis*. Cengage Learning, 2015. [Cited on page 45]

[99] A. Gupta, P. Mishra, C. Pandey, U. Singh, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Annals of Cardiac Anaesthesia*, vol. 22, no. 1, p. 67, 2019. [Cited on page 47]

[100] M. Srivastava, "Analysis of variance and covariance: how to choose and construct models for the life sciences," *Journal of Applied Statistics*, vol. 37, pp. 1059–1060, May 2010. [Cited on page 51]

[101] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*. Packt Publishing, 2019. [Cited on page 52]

[102] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," in *Data Democracy*, pp. 83–106, Elsevier, 2020. [Cited on page 54]

[103] A. Bronshtein, "Train/Test Split and Cross Validation in Python | by Adi Bronshtein | Towards Data Science," 2017. [Cited on page 54]

Appendix A

Appendix A - Summary Results

A.1 ANOVA Results

A.1.1 A-Wave Amplitude

Table A.1: Summary table of the one-way ANOVA test for 'A-Wave Amplitude' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	1.158851e-12	1.0	0.000298	0.986292	1.158851e-12
Residual	1.864909e-07	48.0	NaN	NaN	3.885228e-09

Table A.2: Results of the HSD Test for 'A-Wave Amplitude' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.9863	-0.0	0.0	False

Table A.3: Summary table of the one-way ANOVA test for 'A-Wave Amplitude' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	2.883701e-10	1.0	0.233367	0.631235	2.883701e-10
Residual	5.931333e-08	48.0	NaN	NaN	1.235694e-09

Table A.4: Results of the HSD Test for 'A-Wave Amplitude' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0	0.6312	-0.0	0.0	False

Table A.5: Summary table of the one-way ANOVA test for 'A-Wave Amplitude' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	9.553736e-12	1.0	0.0218	0.88324	9.553736e-12
Residual	2.103614e-08	48.0	NaN	NaN	4.382530e-10

Table A.6: Results of the HSD Test for 'A-Wave Amplitude' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.8832	-0.0	0.0	False

A.1.2 A-Wave Implicit Time

Table A.7: Summary table of the one-way ANOVA test for 'A-Wave Implicit Time' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000003	1.0	5.50371	0.023154	3.123543e-06
Residual	0.000027	48.0	NaN	NaN	5.675341e-07

Table A.8: Results of the HSD Test for 'A-Wave Implicit Time' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0005	0.0232	-0.0009	-0.0001	True

Table A.9: Summary table of the one-way ANOVA test for 'A-Wave Implicit Time' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000126	1.0	1.135709	0.291892	0.000126
Residual	0.005339	48.0	NaN	NaN	0.000111

Table A.10: Results of the HSD Test for 'A-Wave Implicit Time' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0032	0.2919	-0.0092	0.0028	False

Table A.11: Summary table of the one-way ANOVA test for 'A-Wave Implicit Time' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000105	1.0	0.306385	0.582476	0.000105
Residual	0.016470	48.0	NaN	NaN	0.000343

Table A.12: Results of the HSD Test for 'A-Wave Implicit Time' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0029	0.5825	-0.0077	0.0135	False

A.1.3 B-Wave Implicit Time

Table A.13: Summary table of the one-way ANOVA test for 'B-Wave Implicit Time' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000003	1.0	0.025965	0.872661	0.000003
Residual	0.005666	48.0	NaN	NaN	0.000118

Table A.14: Results of the HSD Test for 'B-Wave Implicit Time' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0005	0.8727	-0.0067	0.0057	False

Table A.15: Summary table of the one-way ANOVA test for 'B-Wave Implicit Time' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000120	1.0	1.499733	0.226689	0.00012
Residual	0.003853	48.0	NaN	NaN	0.00008

Table A.16: Results of the HSD Test for 'B-Wave Implicit Time' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0031	0.2267	-0.0082	0.002	False

Table A.17: Summary table of the one-way ANOVA test for 'B-Wave Implicit Time' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.001203	1.0	4.501702	0.03905	0.001203
Residual	0.012825	48.0	NaN	NaN	0.000267

Table A.18: Results of the HSD Test for 'B-Wave Implicit Time' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0098	0.0391	0.0005	0.0192	True

A.1.4 B-Wave Amplitude

Table A.19: Summary table of the one-way ANOVA test for 'B-Wave Amplitude' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	1.282562e-08	1.0	1.326356	0.255159	1.282562e-08
Residual	4.641514e-07	48.0	NaN	NaN	9.669822e-09

Table A.20: Results of the HSD Test for 'B-Wave Amplitude' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.2552	-0.0001	0.0	False

Table A.21: Summary table of the one-way ANOVA test for 'B-Wave Amplitude' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	6.985010e-10	1.0	0.161802	0.689289	6.985010e-10
Residual	2.072169e-07	48.0	NaN	NaN	4.317018e-09

Table A.22: Results of the HSD Test for 'B-Wave Amplitude' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.6893	-0.0	0.0	False

Table A.23: Summary table of the one-way ANOVA test for 'B-Wave Amplitude' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	3.989554e-10	1.0	0.113842	0.737283	3.989554e-10
Residual	1.682147e-07	48.0	NaN	NaN	3.504472e-09

Table A.24: Results of the HSD Test for 'B-Wave Amplitude' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.7373	-0.0	0.0	False

A.1.5 Maximum Signal Frequency

Table A.25: Summary table of the one-way ANOVA test for 'Maximum Signal Frequency' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000113	1.0	1.401492	0.242304	0.000113
Residual	0.003872	48.0	NaN	NaN	0.000081

Table A.26: Results of the HSD Test for 'Maximum Signal Frequency' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.003	0.2423	-0.0081	0.0021	False

Table A.27: Summary table of the one-way ANOVA test for 'Maximum Signal Frequency' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	0.000024	1.0	0.519817	0.474417	0.000024
Residual	0.002219	48.0	NaN	NaN	0.000046

Table A.28: Results of the HSD Test for 'Maximum Signal Frequency' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0014	0.4744	-0.0053	0.0025	False

Table A.29: Summary table of the one-way ANOVA test for 'Maximum Signal Frequency' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	8.267099e-07	1.0	0.01639	0.898666	8.267099e-07
Residual	2.421151e-03	48.0	NaN	NaN	5.044065e-05

Table A.30: Results of the HSD Test for 'Maximum Signal Frequency' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0003	0.8987	-0.0038	0.0043	False

A.1.6 Minimum Signal Frequency

Table A.31: Summary table of the one-way ANOVA test for 'Minimum Signal Frequency' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	1.719975e-14	1.0	0.458378	0.501633	1.719975e-14
Residual	1.801108e-12	48.0	NaN	NaN	3.752308e-14

Table A.32: Results of the HSD Test for 'Minimum Signal Frequency' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.5016	-0.0	0.0	False

Table A.33: Summary table of the one-way ANOVA test for 'Minimum Signal Frequency' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	1.821183e-15	1.0	0.063695	0.801826	1.821183e-15
Residual	1.372424e-12	48.0	NaN	NaN	2.859217e-14

Table A.34: Results of the HSD Test for 'Minimum Signal Frequency' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0	0.8018	-0.0	0.0	False

Table A.35: Summary table of the one-way ANOVA test for 'Minimum Signal Frequency' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	2.673496e-14	1.0	0.781682	0.381034	2.673496e-14
Residual	1.641688e-12	48.0	NaN	NaN	3.420184e-14

Table A.36: Results of the HSD Test for 'Minimum Signal Frequency' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.381	-0.0	0.0	False

A.1.7 Maximum of Signal Derivative

Table A.37: Summary table of the one-way ANOVA test for 'Maximum of Signal Derivative' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	5.504834e-12	1.0	0.030509	0.862074	5.504834e-12
Residual	8.660674e-09	48.0	NaN	NaN	1.804307e-10

Table A.38: Results of the HSD Test for 'Maximum of Signal Derivative' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.8621	-0.0	0.0	False

Table A.39: Summary table of the one-way ANOVA test for 'Maximum of Signal Derivative' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	2.099885e-11	1.0	0.167481	0.684181	2.099885e-11
Residual	6.018270e-09	48.0	NaN	NaN	1.253806e-10

Table A.40: Results of the HSD Test for 'Maximum of Signal Derivative' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0	0.6842	-0.0	0.0	False

Table A.41: Summary table of the one-way ANOVA test for 'Maximum of Signal Derivative' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	1.496478e-12	1.0	0.086583	0.769836	1.496478e-12
Residual	8.296183e-10	48.0	NaN	NaN	1.728371e-11

Table A.42: Results of the HSD Test for 'Maximum of Signal Derivative' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0	0.7698	-0.0	0.0	False

A.1.8 Minimum of Signal Derivative

Table A.43: Summary table of the one-way ANOVA test for 'Minimum of Signal Derivative' for channel 15.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	8.157188e-14	1.0	0.00196	0.964875	8.157188e-14
Residual	1.998131e-09	48.0	NaN	NaN	4.162773e-11

Table A.44: Results of the HSD Test for 'Minimum of Signal Derivative' for channel 15.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.9649	-0.0	0.0	False

Table A.45: Summary table of the one-way ANOVA test for 'Minimum of Signal Derivative' for channel 9.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	1.312950e-12	1.0	0.045454	0.832075	1.312950e-12
Residual	1.386500e-09	48.0	NaN	NaN	2.888541e-11

Table A.46: Results of the HSD Test for 'Minimum of Signal Derivative' for channel 9.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	-0.0	0.8321	-0.0	0.0	False

Table A.47: Summary table of the one-way ANOVA test for 'Minimum of Signal Derivative' for channel 3.

	Sum of squares	DF	F	p-value	Mean of squares
C(Genotype)	2.545388e-12	1.0	0.40804	0.526003	2.545388e-12
Residual	2.994282e-10	48.0	NaN	NaN	6.238088e-12

Table A.48: Results of the HSD Test for 'Minimum of Signal Derivative' for channel 3.

Group 1	Group 2	mean_diff	p-adj	lower	upper	reject
NF1	WT	0.0	0.526	-0.0	0.0	False

A.2 Classification via Machine Learning Results

A.2.1 Logistic Regression

Table A.49: Report of the classification using LR for channel 15.

	precision	recall	f1-score	support
WT	1.0000	0.666667	0.800000	3.0
NF1	0.8750	1.000000	0.933333	7.0
accuracy	0.9000	0.900000	0.900000	0.9
macro avg	0.9375	0.833333	0.866667	10.0
weighted avg	0.9125	0.900000	0.893333	10.0

Table A.50: Report of the classification using LR for channel 9.

	precision	recall	f1-score	support
WT	0.750000	0.750000	0.750000	4.0
NF1	0.833333	0.833333	0.833333	6.0
accuracy	0.800000	0.800000	0.800000	0.8
macro avg	0.791667	0.791667	0.791667	10.0
weighted avg	0.800000	0.800000	0.800000	10.0

Table A.51: Report of the classification using LR for channel 3.

	precision	recall	f1-score	support
WT	0.80	1.000000	0.888889	4.0
NF1	1.00	0.833333	0.909091	6.0
accuracy	0.90	0.900000	0.900000	0.9
macro avg	0.90	0.916667	0.898990	10.0
weighted avg	0.92	0.900000	0.901010	10.0

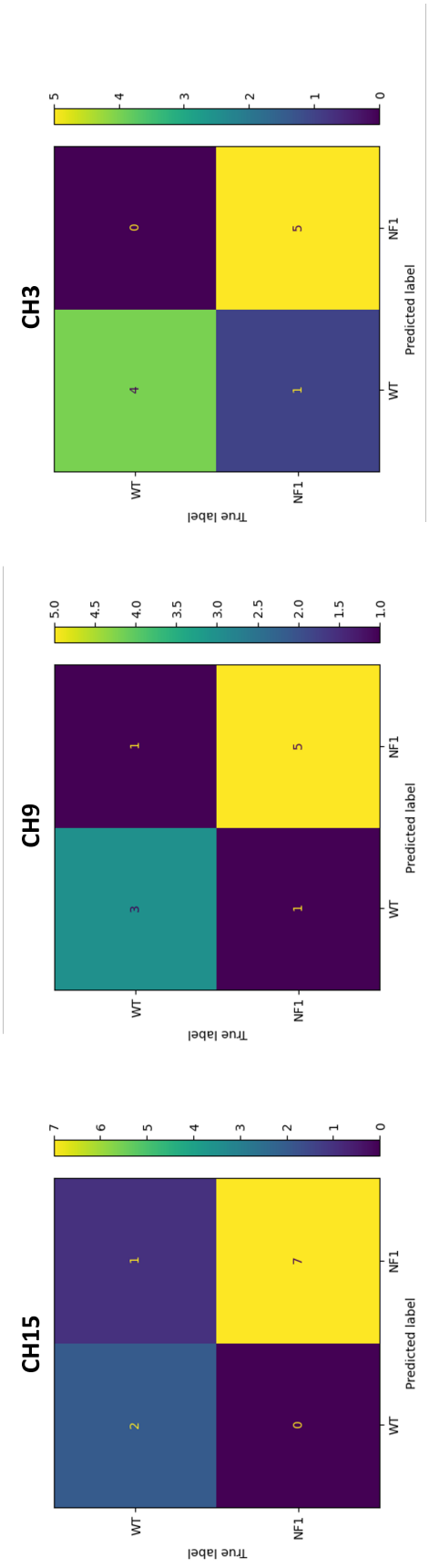


Figure A.1: Representation of the confusion matrices for channels 15, 9 and 3, for the LR classifier.

A.2.2 Support Vector Machine

Table A.52: Report of the classification using SVM for channel 15.

	precision	recall	f1-score	support
WT	0.750	1.000000	0.857143	3.0
NF1	1.000	0.857143	0.923077	7.0
accuracy	0.900	0.900000	0.900000	0.9
macro avg	0.875	0.928571	0.890110	10.0
weighted avg	0.925	0.900000	0.903297	10.0

Table A.53: Report of the classification using SVM for channel 9.

	precision	recall	f1-score	support
WT	1.000000	0.6	0.750000	5.0
NF1	0.714286	1.0	0.833333	5.0
accuracy	0.800000	0.8	0.800000	0.8
macro avg	0.857143	0.8	0.791667	10.0
weighted avg	0.857143	0.8	0.791667	10.0

Table A.54: Report of the classification using SVM for channel 3.

	precision	recall	f1-score	support
WT	1.000000	0.8	0.888889	5.0
NF1	0.833333	1.0	0.909091	5.0
accuracy	0.900000	0.9	0.900000	0.9
macro avg	0.916667	0.9	0.898990	10.0
weighted avg	0.916667	0.9	0.898990	10.0

A.2.3 Naive Bayes

Table A.55: Report of the classification using NB for channel 15.

	precision	recall	f1-score	support
WT	1.0	1.0	1.0	6.0
NF1	1.0	1.0	1.0	4.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	10.0
weighted avg	1.0	1.0	1.0	10.0

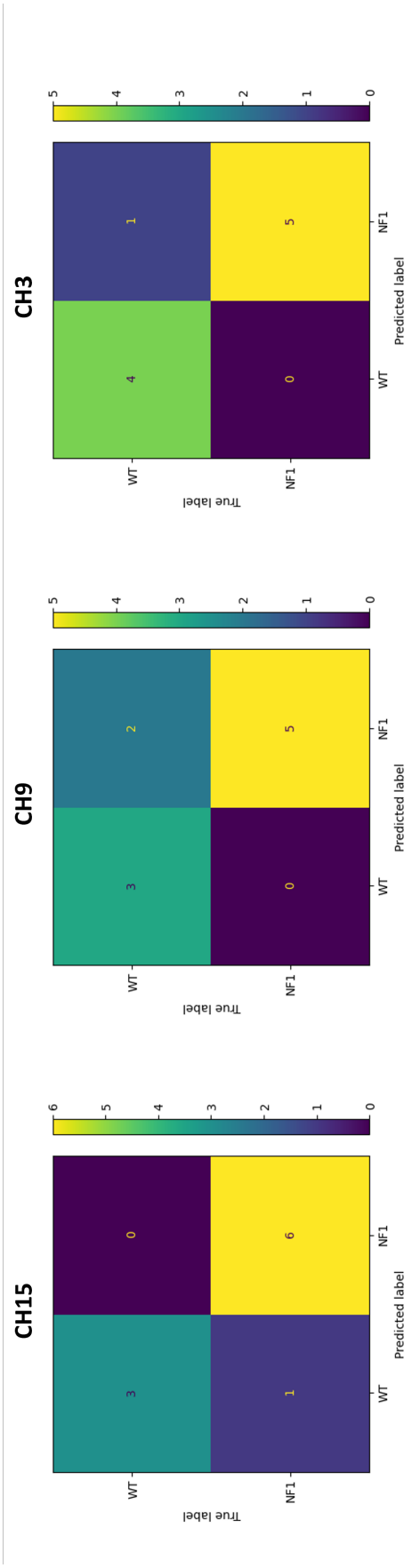


Figure A.2: Representation of the confusion matrices for channels 15, 9 and 3, for the SVM classifier.

Table A.56: Report of the classification using NB for channel 9.

	precision	recall	f1-score	support
WT	0.666667	0.666667	0.666667	3.0
NF1	0.857143	0.857143	0.857143	7.0
accuracy	0.800000	0.800000	0.800000	0.8
macro avg	0.761905	0.761905	0.761905	10.0
weighted avg	0.800000	0.800000	0.800000	10.0

Table A.57: Report of the classification using NB for channel 3.

	precision	recall	f1-score	support
WT	1.0	1.0	1.0	8.0
NF1	1.0	1.0	1.0	2.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	10.0
weighted avg	1.0	1.0	1.0	10.0

A.2.4 Random Forest

Table A.58: Report of the classification using RF for channel 15.

	precision	recall	f1-score	support
WT	1.000	0.857143	0.923077	7.0
NF1	0.750	1.000000	0.857143	3.0
accuracy	0.900	0.900000	0.900000	0.9
macro avg	0.875	0.928571	0.890110	10.0
weighted avg	0.925	0.900000	0.903297	10.0

Table A.59: Report of the classification using RF for channel 9.

	precision	recall	f1-score	support
WT	0.750000	0.750000	0.750000	4.0
NF1	0.833333	0.833333	0.833333	6.0
accuracy	0.800000	0.800000	0.800000	0.8
macro avg	0.791667	0.791667	0.791667	10.0
weighted avg	0.800000	0.800000	0.800000	10.0

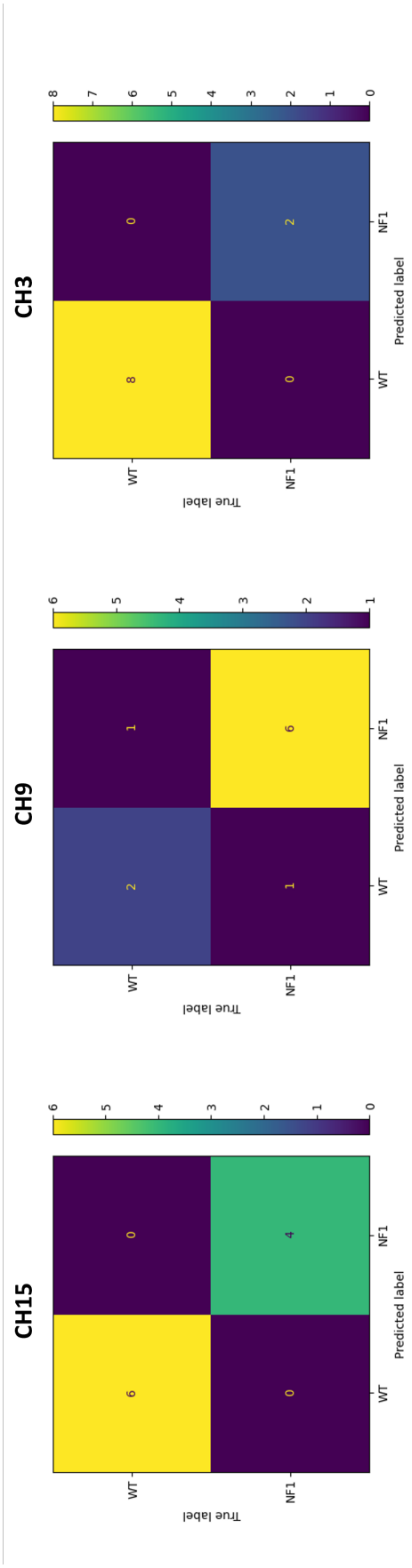


Figure A.3: Representation of the confusion matrices for channels 15, 9 and 3, for the NB classifier.

Table A.60: Report of the classification using RF for channel 3.

	precision	recall	f1-score	support
WT	1.000000	0.750	0.857143	4.0
NF1	0.857143	1.000	0.923077	6.0
accuracy	0.900000	0.900	0.900000	0.9
macro avg	0.928571	0.875	0.890110	10.0
weighted avg	0.914286	0.900	0.896703	10.0

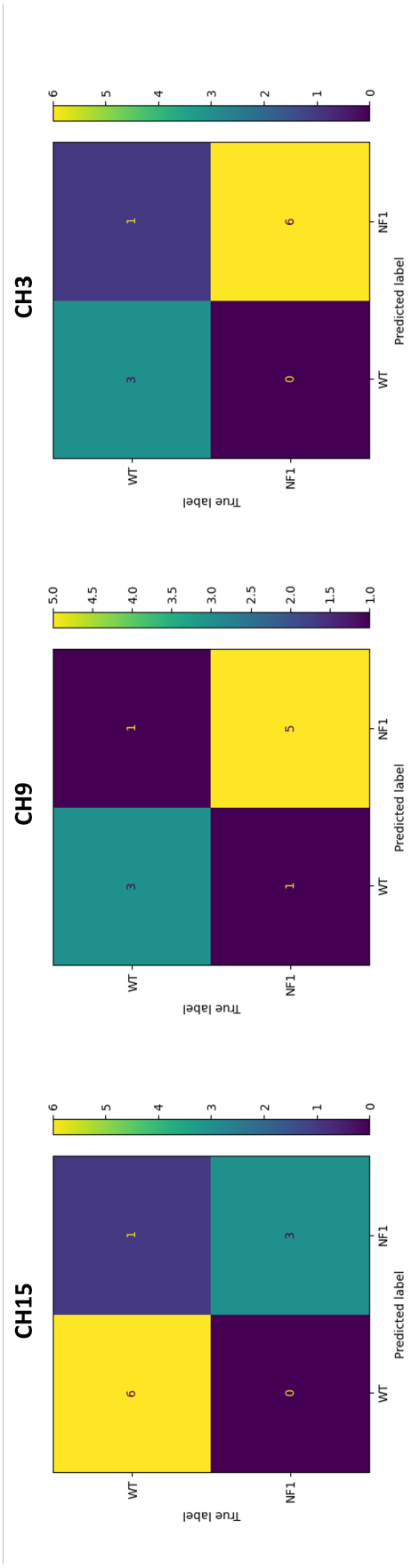


Figure A.4: Representation of the confusion matrices for channels 15, 9 and 3, for the RF classifier.