CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA

PORTO

MACHINE LEARNING APPROACHES FOR DETECTING DEPRESSION USING EEG SIGNALS

by
Eunice Monteiro de Oliveira

February 2022

CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA

PORTO

MACHINE LEARNING APPROACHES FOR DETECTING DEPRESSION USING EEG SIGNALS

Thesis presented to Escola Superior de Biotecnologia of the
Universidade Católica Portuguesa to fulfill the requirements of Master of Science degree in Biomedical Engineering

by
Eunice Monteiro de Oliveira

Supervisor: Prof. Doutor Pedro Miguel de Luís Rodrigues
Escola Superior de Biotecnologia, Universidade Católica Portuguesa

Co-Supervisor: Prof. Doutor Bruno Catarino Bispo
Universidade Federal de Santa Catarina

February 2022

Resumo

O cargo dos distúrbios neurológicos continua a crescer, estimando-se que, a nível mundial, 264 milhões de pessoas sofram de depressão atualmente. Devido ao estigma associado à doença mental e ao facto de a abordagem de diagnóstico mais comum ser tão humano-intensiva, é menos provável que os indivíduos deprimidos procurem ajuda. Além disso, os resultados do diagnóstico estão dependentes da experiência do profissional de saúde. Quando os pacientes são incorretamente diagnosticados, a procura de explicações físicas dos sintomas aumenta ainda mais o custo dos cuidados médicos, que muitas pessoas não têm capacidade financeira para suportar. Como tal, a procura de um método de diagnóstico da depressão económico, objetivo, e menos humano-intensivo torna-se crucial.

O presente estudo centrou-se no desenvolvimento de uma ferramenta capaz de detetar padrões indicativos de depressão e de discriminar automaticamente os pacientes deprimidos através da análise de sinais EEG.

Com recurso à Transformada de Wavelet Discreta 1D, foi realizada uma análise multibanda dos sinais por canal EEG. Após os processos de extração e seleção de parâmetros, os parâmetros obtidos alimentaram 25 modelos de *Machine Learning* e uma Rede Neural Convolucional (CNN). Os três classificadores com melhor desempenho foram os seguintes: Análise Discriminante Linear, Máquina de Vetor de Suporte (janela cúbica), e a CNN concebida, com uma precisão de classificação global de 94,8%, 93,9%, e 94,9%, respetivamente. Através destes três classificadores, a comparação entre sujeitos deprimidos e controlos saudáveis atingiu uma precisão de 100% em vários canais. Os resultados obtidos pelos classificadores, juntamente com uma análise através de mapas topográficos, levam a concluir que existe uma diferença na frequência das ondas cerebrais entre os dois grupos, com uma forte incidência nas regiões frontocentral, central, e parietoccipital do couro cabeludo.

Embora a análise do sinal EEG ainda não possa ser aplicada como ferramenta de diagnóstico da depressão, os resultados deste estudo continuam a ser relevantes do ponto de vista teórico.

Keywords: Depressão, Machine Learning, Rede Neural Convolucional, Sinais EEG, Transformada de Wavelet Discreta.

Abstract

The burden of neurological disorders continues to grow, as an estimated 264 million people currently suffer from depression worldwide. Due to the stigma surrounding mental illness and to the standard diagnostic approach being so human-intensive, depressed individuals are less likely to seek help. Additionally, the diagnosis' results are dependent on the doctor's experience. When patients are misdiagnosed, the search for physical explanations of symptoms further increases the medical care cost, which a lot of people are not able to support. Hence, the search for a cost-effective, objective, and less human-intensive diagnostic method of depression happens to be crucial.

The present study focused on developing a tool capable of detecting indicative patterns of depression and automatically discriminate patients with depression through EEG signal analysis.

With resource to 1D Discrete Wavelet Transform, a multiband analysis of the signals was performed per EEG channel. After the feature extraction and the feature selection processes, the obtained features fed 25 Machine Learning models and a convolutional neural network (CNN). The three classifiers with the best performance were Linear Discriminant Analysis, Cubic Support Vector Machine, and the designed CNN, with an overall classification accuracy of 94.8%, 93.9%, and 94.9%, respectively. Through these three classifiers, the comparison between depressed subjects and healthy controls reached an accuracy of 100% on several channels. The results obtained by the classifiers alongside an analysis through topographic maps lead to conclude that there is a difference in the frequency of brain waves between the two groups, with a strong incidence in the frontocentral, central, and parietooccipital regions of the scalp.

Although EEG signal analysis cannot yet be applied as a diagnostic tool for depression, the findings in this study remain relevant from a theoretical point of view.

Keywords: Convolutional Neural Network, Depression, Discrete Wavelet Transform, EEG signals, Machine Learning.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Pedro Rodrigues, who has been of invaluable help and patience during the development of this project.

I would also like to thank my parents, Iris Monteiro and Alexandre Oliveira, for investing in my education and supporting me unconditionally.

A heartfelt thanks to my grandparents, Maria Eugénia Monteiro and Armando Monteiro, for being present and understanding at the hardest times of this five-year journey.

Finally, a very special thanks to my friend André Tavares, whom support was crucial throughout this project.

Eunice Monteiro de Oliveira

Outline

List of Figures

List of Tables

Abbreviations and Symbols

adam	Adaptive Moment Estimation
ADHD	Deficit Hyperactivity Disorder
AI	Artificial Intelligence
ApEn	Approximate Entropy
AUC	Area under the ROC Curve
BCI	Brain-Computer Interface
bior	Biorthogonal
BOLD	Blood-oxygenation-level Dependent
CNN	Convolutional Neural Network
coif	Coiflets
CT	Computed Tomography
db	Daubechies
DL	Deep Learning
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (5th edition)
DST	Dexamethasone Suppression Test
DWT	Discrete Wavelet Transform
ECoG	Electrocorticography
EEG	Electroencephalography
fMRI	Functional Magnetic Resonance Imaging
HA	Hjorth Activity
HC	Hjorth Complexity
HM	Hjorth Mobility
Hz	Hertz
kNN	k-Nearest Neighbour
kurt	Kurtosis
LDA	Linear Discriminant Analysis
LEEn	Log Energy Entropy
LLE	Largest Lyapunov Exponent

ME	Mean Energy
MEG	Magnetoencephalography
mg	Milligram
ML	Machine Learning
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
NIRS	Near-infrared Spectroscopy
np	Number of Peaks
PET	Positron Emission Tomography
PMDD	Premenstrual Dysphoric Disorder
rbio	Reverse biorthogonal
ReLu	Rectified Linear Unit
RMS	Root Mean Square Value
rmsprop	Root Mean Square Propagation
ROC	Receiver Operating Characteristics
SEn	Shannon Entropy
sgdm	Stochastic Gradient Descent with Momentum
skew	Skewness
std	Standard Deviation
SVM	Support Vector Machine
sym	Symlets
μV	Microvolt

CHAPTER 1

Introduction

1.1 Motivation

The motivation behind this thesis is based on two facts concerning depression. First, depression is one of the major contributors to the overall global burden of disease, with significant impacts on health and serious social and economic consequences. Second, the lack of objective and less human-intensive, but at the same time accurate, diagnostic methods. Hence, the search for a cost-effective, objective, and less human-intensive diagnostic method of depression happens to be crucial.

The collection and analysis of EEG signals for the diagnosis of depression is an emerging field, still under study to prove its validity. This thesis aims to contribute to that direction.

1.2 Objectives and Contributions

The main objective of this thesis was to develop an algorithm capable of detecting indicative patterns of depression through EEG signal analysis. Thus, it was performed a multiband analysis of those signals and thereafter applied artificial intelligence (Machine Learning and Deep Learning) techniques to automatically discriminating patients with depression. To carry through this study, the algorithm was developed in MATLAB® 2019b.

1.3 Thesis Structure

This thesis is composed of seven chapters. The first chapter presents the motivation, objectives, and contributions behind the conducted study. The second chapter provides information to understand depression and its diagnostic methods. Chapter three addresses the main aspects of electroencephalography. Chapter four overviews several artificial intelligence techniques used in signal analysis. In

chapter five a detailed explanation of the proposed methodology is described, whereas chapter six includes the results obtained with subsequent discussion. Finally, in chapter seven conclusions are drawn and future perspectives for the study carried out are presented.

CHAPTER 2

Depression

This chapter focuses on depression by addressing its types and symptoms, causes and risks factors, epidemiology, and conventional and emerging diagnostic methods.

2.1 Types and Symptoms

Depression is a psychological disorder that affects an individual's mental and physical health. The four most common types of depression are major depressive disorder, persistent depressive disorder, bipolar disorder, and seasonal affective disorder [1]. Depression types exclusive to women include perinatal depression and premenstrual dysphoric disorder (PMDD) [1].

The core symptoms of depression are dysphoria or depressed mood, loss of interest or pleasure in nearly all activities (anhedonia), decreased memory and concentration, decreased energy, and low self-esteem [2, 3]. In addition, physical symptoms associated with depression include backaches, nonspecific musculoskeletal complaints, multiple (three or more) somatic complaints, vague complaints, and weight changes [2, 3].

2.2 Causes and Risk Factors

It is popularly said that depression results from a chemical imbalance. Despite that being true, this disorder is much more complex, as there are many other possible causes, including faulty moody regulation by the brain, genetic vulnerability, stressful life events, medications, and medical problems [4].

Depression can arise from malfunctioning of the parts of the nervous system responsible for balancing mood, behaviour, thinking, and sleep [3]. Areas that play a significant role in these

processes, therefore having a major impact on depression, are the amygdala, the thalamus, and the hippocampus [4].

As dozens of genes responsible for controlling mood differ from person to person, vulnerability to depression does too [4]. Genetically speaking, if depression runs in a person's family, that person is more vulnerable to depression. The increase in risk for the condition varies depending on the type of depression.

At some point in life, nearly everyone encounters stressful events such as the death of a loved one, an illness, the loss of a job, or a failed relationship [4]. Although almost everyone who faces these stresses does not develop a mood disorder, stress plays an important role in depression. When stress is chronic, changes in the body and brain can be long-lasting. Researchers have found that early losses and emotional trauma may leave individuals more vulnerable to depression later in life [4].

It is believed that medical illnesses or medications might be at the root of up to 15% of all depressions [4]. The following medical conditions have been associated with depression: degenerative neurological conditions, endocrine disorders, stroke, cancer, erectile dysfunction in men, some nutritional deficiencies (e.g., lack of vitamin B12), certain immune system diseases (e.g., lupus), some viruses and other infections (e.g., HIV, hepatitis, and mononucleosis) [4]. Sometimes, symptoms of depression are a side effect of certain drugs, such as steroids, birth control pills, or blood pressure medication [4].

Individuals at risk of depression include those with a prior depressive episode, a recent stressful life event, family history of depression, chronic medical conditions, cognitive impairment or dementia, anxiety, substance abuse, and multiple physical complaints, vague complaints, or unexplained physical symptoms [2].

2.3 Epidemiology

The burden of neurological disorders continues to grow with significant impacts on health and major social and economic consequences worldwide [5]. Depressive disorders are common in the general population, with a point prevalence of about 2% to 4% for major depressive disorder and about a 20% lifetime risk for the development of major depression or persistent depressive disorder [6]. Globally, an estimated 264 million people suffer from depression, with the rate of depression among women being 2 to 3 times that of men [5, 6].

Among the burdens caused by this disorder are patient suffering, family distress and conflict, impaired cognitive development of young children in cases of perinatal depression, and the strikingly increased risk of suicide [6].

The effects of depression on the consumption of health care are astounding. When patients are misdiagnosed, the search for physical explanations of symptoms causes medical utilization rates to increase, further increasing health care costs [6]. Costs of depression in the United States of America have been estimated at $210.5 billion per year, with 45-47% of this value being relative to direct

medical care [7]. A part between 48-50% of these costs are attributed to the workplace (missed days from work and reduced productivity while at work) and about 5% of the total expenditures are related to suicide [7].

2.4 Diagnostic Methods

Considering the epidemiology of depression, it is plausible to state that its diagnosis in the early curable stages is critical and might save lives. Conventional and emerging diagnostic methods of depression are overviewed in this subchapter.

2.4.1 Interview-based Methods

As with most psychiatric disorders, the diagnosis of depression is fundamentally clinical [6]. The most used diagnostic method of depression is a scale-based interview conducted by a psychologist or a psychiatrist along with a mental status examination [6, 8]. The standard diagnostic approach is the criteria in "Diagnostic and Statistical Manual of Mental Disorders (5th edition)" (DSM-5), last updated in 2013 [6]. The clinical test "Mini-Mental State Examination" (MMSE) is also commonly applied [8].

Evidence suggests that such a diagnostic method is comparable in sensitivity and specificity to radiologic and laboratory tests [8]. However, that approach is human-intensive, and the results are dependent on the doctor's experience [9].

2.4.2 Laboratory Methods

The dexamethasone suppression test (DST) has been used as a diagnostic method for depression [9]. This laboratory method assesses the negative feedback of dexamethasone, a cortisol-like synthetic hormone, on pituitary corticotropin release. An administration of a low dose (1-2 mg) of dexamethasone should reduce corticotropin levels and lead to decreased cortisol levels in healthy individuals, but in many depressed patients, cortisol levels remain elevated [9]. This test can differentiate severe melancholic depression, manic behaviour, or acute psychosis from chronic psychosis or dysthymia. However, it has proven to be inconvenient for patients and lacks good clinical performance for depression assessment [9]. As for now, DST cannot be broadly used in depression assessment, but it may still have value for some patients when used with discretion and supported by an abundance of other clinical information [9].

Genomic, proteomic, and metabolic profiling are beginning to emerge in psychiatry to identify biomarkers with application to depression diagnosis [9]. Although promising, there is still no sufficient evaluation of these methods to fully validate their clinical performance and reliability [9].

2.4.3 Diagnostic Imaging Techniques

Structural and functional brain alterations are common in patients suffering from depression [10]. Because of that, diagnostic imaging techniques have been used to study the core aspects of depression during the last three decades. Particularly, magnetic resonance imaging (MRI) has taken an important part in deciphering the etiology and pathogenesis of depression and other psychiatric disorders [10].

Structural MRI is used to assess anatomical alterations in the brain, which are typically denoted as volume differences [10]. Functional brain imaging is generally combined with anatomic imaging modalities, such as MRI or computed tomography (CT), to visualize the activation of specific brain regions [10]. Blood-oxygenation-level dependent (BOLD) functional MRI (fMRI) is the most used fMRI technique to assess functional brain activity [10]. Patients with depression present distinct structural and functional brain alterations [10].

Given the biological complexity of this disorder, two patients never show the same structural or functional brain patterns. Therefore, a combination of structural and functional alterations should be considered when developing diagnostic tools for depression [10].

Currently, diagnostic imaging modalities are not being used to diagnose depression in the clinic [10]. Bearing in mind that different neurological disorders can produce similar brain alterations, a physician is unable to determine whether a specific alteration in a brain scan is related to depression alone, as it may be associated with other biological processes or behaviours, concurrent or past diseases, or imaging artifacts [10]. Hence, diagnostic imaging modalities cannot be used to diagnose a specific neurological disorder [10].

Researchers are attempting to develop new technologies to overcome the current limits of diagnostic imaging modalities in depression assessment, so they can be used to diagnose depression in the clinic [10].

2.4.4 Brain Signal Acquisition Techniques

The collection and analysis of brain signals for the diagnosis of neurological disorders can open a new path for an objective depression detection tool [8].

Over the past years, various technologies have been developed to measure brain activity. Some of them measure the variation of the electrical activities related to the different states of the brain, while others measure different parameters [11]. The available modalities are classified based on their invasiveness. Invasive techniques include electrocorticography (ECoG) and intracranial electroencephalography. Noninvasive techniques include electroencephalography (EEG), near-infrared spectroscopy (NIRS), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and magnetoencephalography (MEG) [11]. Unlike the noninvasive techniques, invasive techniques are not currently used in clinical applications [11].

CHAPTER 3

Electroencephalography

With the advance of technology, brain science has become a popular research field in the past decades. Electroencephalography (EEG) is broadly applied to analyze the electrical activities of the brain. This technique has many applications for human-computer interaction and neurological disease diagnosis [8]. In the present study, EEG was used for depression detection.

3.1 EEG Signal Acquisition

The EEG signal is a recording of the spontaneous, rhythmic, electrical activity of brain neurons [8]. These biomedical signals are recorded from electrodes strategically placed on the scalp that measure the amplitude of the electrical impulses generated by brain neurons [12, 13]. A typical EEG signal has an amplitude range of about 10 μV to 100 μV and a frequency range of 1 Hz to 100 Hz [12].

There are two types of EEG recordings: monopolar and bipolar. The monopolar recording picks up the voltage difference between an active electrode on the scalp and a reference electrode on the ear lobe, whereas bipolar recording gives the voltage difference between two scalp electrodes [12].

The 10-20 International System of Electrodes is commonly used for EEG electrode placement and for correlating external skull locations to underlying cortical areas [14]. Among the electrode systems reported in the literature, for research purposes, the most used are 128-electrode and 256-electrode EEG systems [8].

The nomenclature of the electrodes identifies its location. According to the region where the electrode is placed on the scalp's surface, it will have a stipulated nomenclature: Fp (frontopolar), AF (anterior frontal), F (frontal), C (central), T (temporal), P (parietal), O (occipital), FC (frontocentral), FT (frontotemporal), CP (centroparietal), TP (temporal-posterior temporal), PO (parietooccipital) [15]. Additionally, the left hemisphere of the brain is indicated by odd numbers, the middle is indicated

by the letter Z, and the right hemisphere is indicated by even numbers. The electrodes' disposition and their nomenclature are represented in Figure 3.1.

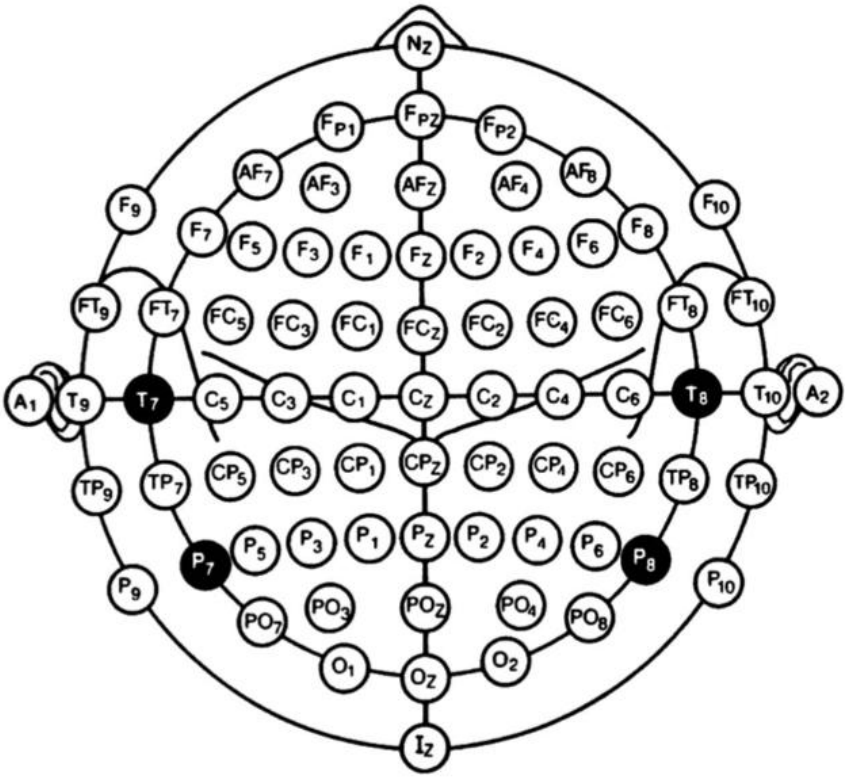

Figure 3.1: Electrodes' disposition on the scalp's surface and their corresponding nomenclature (reproduced from [15]]).

3.2 EEG Signal Characterization

The nature of EEG signals is nonlinear, highly non-Gaussian, random, noncorrelated, and non-stationary [12, 13]. EEG signals are characterized by five rhythms: delta waves, theta waves, alpha waves, beta waves, and gamma waves. Each rhythm provides information about a person's state of mind and can help draw conclusions about their health [13].

The five rhythms along with their frequency range, amplitude, and respective state of mind are presented in Table 3.1. An EEG recording and the corresponding rhythms are displayed in Figure 3.2.

Table 3.1: The five brain rhythms along with their frequency range, amplitude, and respective state of mind (adapted from [13]).

Rhythm	**Frequency Range (Hz)**	**Amplitude (μV)**	**State of Mind**
Delta (δ)	0.1-4	High amplitude (20-200)	Deep sleep
Theta (θ)	4-8	More than 20	Emotional stress, drowsiness, and sleep in adults
Alpha (α)	8-13	30-50	Relaxed awareness
Beta (β)	13-30	5-30	Active thinking, active attention, alert
Gamma (γ)	30-100	Less than 5	Mechanism of consciousness

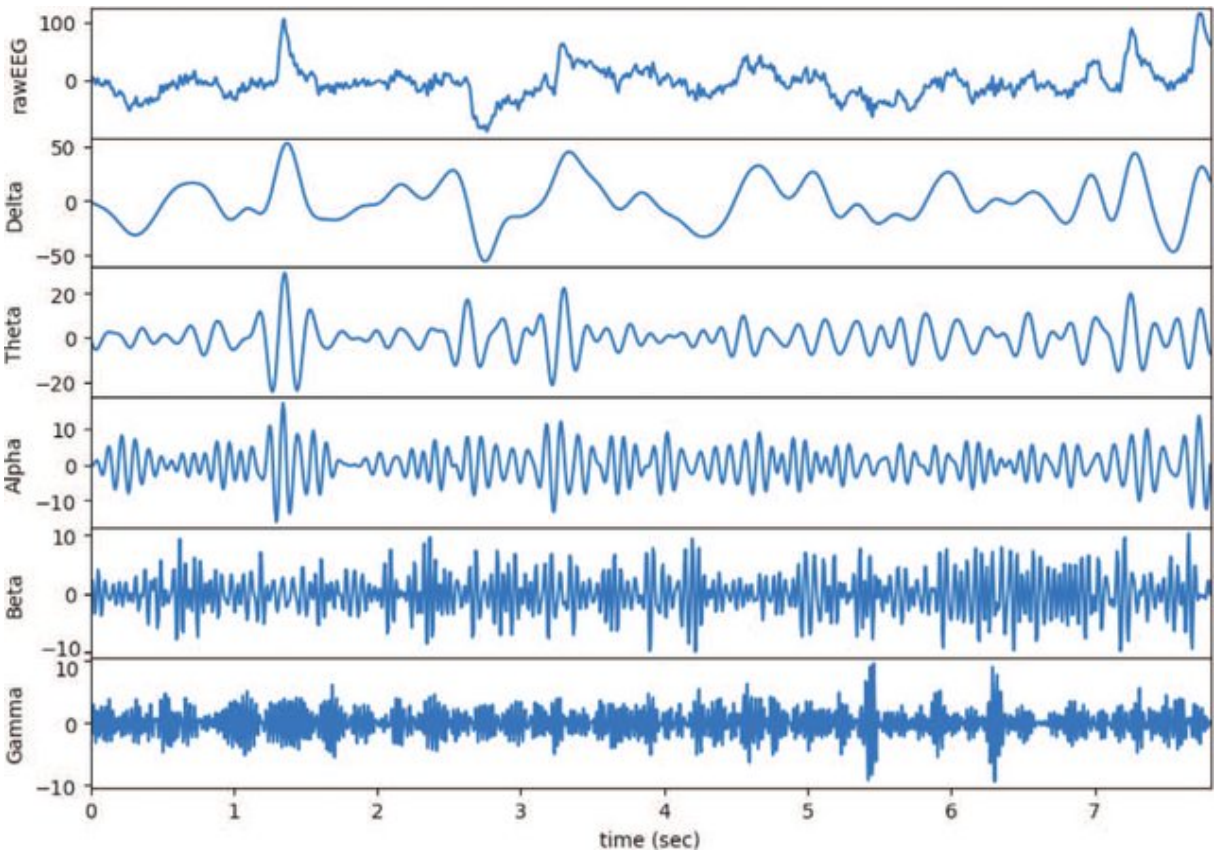

Figure 3.2: EEG recording and its five corresponding rhythms [16]

Delta waves are low-frequency waves that are found predominantly in infants and deep sleep stages of healthy adults. Their presence in awake adults indicates brain damage due to a tumor,

inflammation, or vascular blockage [17]. Their suppression can lead to an inability to rejuvenate the body and revitalize the brain, and poor sleep [18].

Theta activity exists in healthy infants and children just as during drowsiness and sleep in adults. The presence of high theta activity in awake adults may indicate cerebral dysfunction [17]. Its suppression can lead to anxiety, poor emotional awareness, and stress [18].

Alpha waves occur in healthy relaxed adults and disappear during sleep or when concentrating on a specific task. Their activity is maximal over the occipital region of the scalp [17]. Their suppression can cause anxiety, high stress, and insomnia [18].

Beta waves replace alpha waves during attention to tasks or stimuli. Beta activity is predominant over the frontocentral region of the scalp [17]. The prominence of these waves causes an inability to relax, high arousal, stress, and anxiety, whereas its suppression can lead to attention deficit hyperactivity disorder (ADHD), depression, daydreaming, and poor cognition [18].

Gamma activity has the highest frequency. These waves are responsible for cognitive functioning, learning, memory, and information processing [18]. The prominence of gamma activity leads to high arousal, stress, and anxiety, whereas its suppression can lead to ADHD, depression, and learning disabilities [18].

3.3 EEG Signal Analysis

Apart from its complex nature, the features of EEG signals are subject to individual differences such as the individual himself, age, and mental state [13, 19]. Additionally, EEG signals are highly noisy and susceptible to be distorted by artificial interference, which makes it difficult to extract effective information related to specific tasks [19]. Hence, understanding its behaviour and dynamics implies various linear and nonlinear signal processing methods [12].

The interpretation of the signal involves a spectral analysis, a temporal analysis, and a spatial analysis [13]. The spectral analysis of the signal determines the dominant frequencies in the EEG. The temporal analysis of the EEG keeps a record of normal and abnormal wave shapes in the signal, as well as the presence and absence of these rhythms. The spatial analysis estimates the distribution of these rhythms over the different brain regions.

Generally, there are four main steps in the process of extracting relevant information from EEG recordings [19, 20]: first, signal preprocessing; second, feature extraction; third, selection of the best features for the problem at hand; in last, signal classification based on the selected features.

3.4 Electroencephalography vs Other Methods

Research on EEG can be used to understand the mechanism underlying brain activity and the human cognitive process, in the diagnosis of neurological diseases and in the field of Brain-Computer Interface (BCI) [8]. In comparison with other methods used in the clinic (CT, PET, MRI, fMRI, MEG,

and NIRS), EEG has a higher temporal resolution, a simpler method to operate with, lower purchase and maintenance costs, and less bulky equipment [8]. On top of that, this method of obtaining physiological data is noninvasive and does not require exposure to ionizing radiation [8, 11].

CHAPTER 4

Machine Learning Techniques

The application of artificial intelligence (AI) in healthcare is a domain of great interest nowadays. Most AI work involves machine learning (ML), which is the field of study that provides computers the ability to learn without being explicitly programmed [21]. With the advancements of ML, deep learning (DL) emerged as its subfield, implementing ML by mainly using neural networks for learning and prediction of data [21].

This chapter overviews a variety of ML algorithms used in signal classification.

4.1 X-ROC

The X-ROC is a binary classifier that uses the squared distance between mean distributions to find a good threshold to discriminate two classes and the ROC (Receiver Operating Characteristic) to analyze the classification result [22]. This classifier is applied with a leave-one-out cross-validation.

The ROC curve demonstrates the performance of the classifier and represents a plot describing the classifier's true positive rate (Sensitivity) against its false-positive rate (1-Specificity), for different values of a parameter [23]. The area under the ROC curve (AUC) is a single number that represents the performance of the method [22].

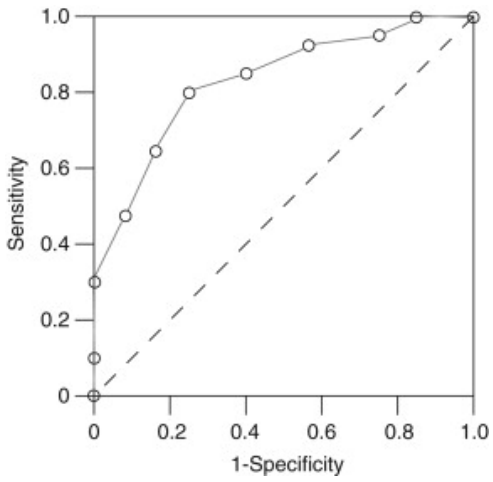

Figure 4.1: An example of a ROC curve (reproduced from [24]).

4.2 Decision Trees

Decision tree algorithms are widely used to classify biomedical signals. A decision tree has a tree-like structure composed of one node called root, which has no incoming edges, test nodes (nodes with outgoing edges), and terminal nodes (nodes with one incoming edge and no outgoing edges) [25]. Each node is labeled with its testing attribute, and its branches are labeled with their corresponding values [25]. To each terminal node is assigned a class representing the most appropriate target. A simple illustrative decision tree is presented in Figure 4.2.

In decision trees, the training set is structured in a top-down recursive manner. In each iteration, the algorithm chooses the best way to split the training set through a discrete function of the input attribute values [25]. After selecting the most appropriate split, each node further subdivides the training set into smaller subsets. The branching of the tree continues until a stopping criterion is satisfied [25].

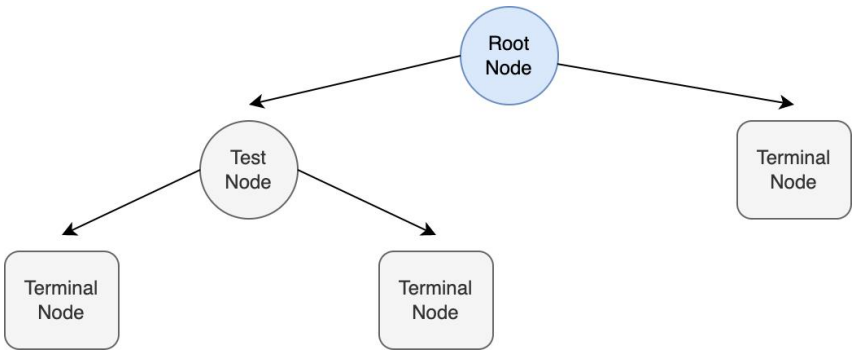

Figure 4.2: An illustrative example of a Decision Tree.

4.3 Discriminant Analysis

Discriminant analysis algorithms are probabilistic parametric classifiers. They use labelled training data to build a predictive model of group membership which can then be applied to new data [26]. Classification of new data can be determined with a linear or a quadratic function that will assign each test data set to one of the predefined groups (class labels). The graph in Figure 4.3 displays both functions separating the data into two groups.

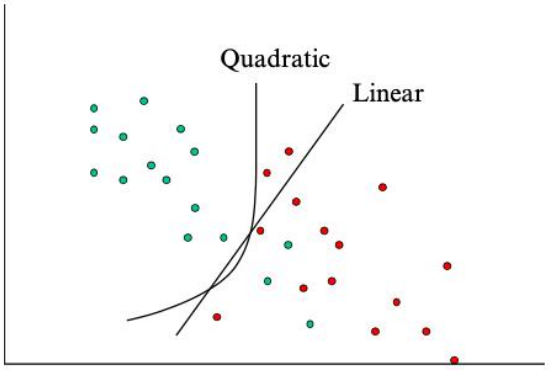

Figure 4.3: An example of the Discriminant Analysis classification process (reproduced from [26]).

4.4 Logistic Regression

The logistic regression model is a statistical method for binary classification that can be generalized to multiclass classification [27]. This classifier measures the relationship between a dichotomous dependent variable and one or more (categorical or continuous) independent variables by using probability scores as the predicted values of the dependent variable [28]. The independent variables do not need to be normally distributed, nor linearly related, nor of equal variance within each group.

In this classification algorithm, log-odds in favor of one of the classes are defined and maximized via a weight vector [29]. Each weighted feature vector is mapped to a value in the range from 0 to 1 by the sigmoid-shaped logistic function [30]. That value indicates the probability of an example to belong to a certain class. The vertical axis stands for the probability for a given classification, whereas the horizontal axis is the value of the independent variable x. The formula of the logistic regression model is the following [31]:

$$Logistic function = \frac{1}{1+e^{-x}} \tag{4.1}$$

4.5 Support Vector Machines

A support vector machine (SVM) is a statistical learning method of classification based on the concept of support vectors and margin maximization. The algorithm aims to project a function in a multidimensional space that can separate training data according to predefined class labels [32].

Considering a training set with two predefined class labels and two available measurements per training case, SVM derives from the idea that these measurements can be regarded as a two-dimensional space where each training case is represented by a point [32]. For the two-dimensional space, a line is drawn to separate the training cases. Despite the infinite number of lines that could in an equally efficient manner separate the cases, the best line is the one that maximizes the margin between the two classes, i.e., the line with a maximum distance to the closest points from each of the two classes [32]. After computing the separation line, new data can be automatically assigned to any of the two classes depending on their location relative to the line [32]. Such a concept can be extended to a higher-dimensional space, in which the separation line becomes a plane in the three-dimensional case or a hyperplane for higher dimensions. The closest points of both classes to the hyperplane are named support vectors since they define its margin and correspondingly its location and orientation [32]. When the training data are not separable, which is usually the case, SVM kernel methods are performed [33]. An example of the SVM classification process is given in Figure 4.4.

Recently, SVMs have been used in the context of precision psychiatry, particularly for applications that involve predicting the diagnosis/prognosis of neurological diseases such as Alzheimer's disease, schizophrenia, and depression [33].

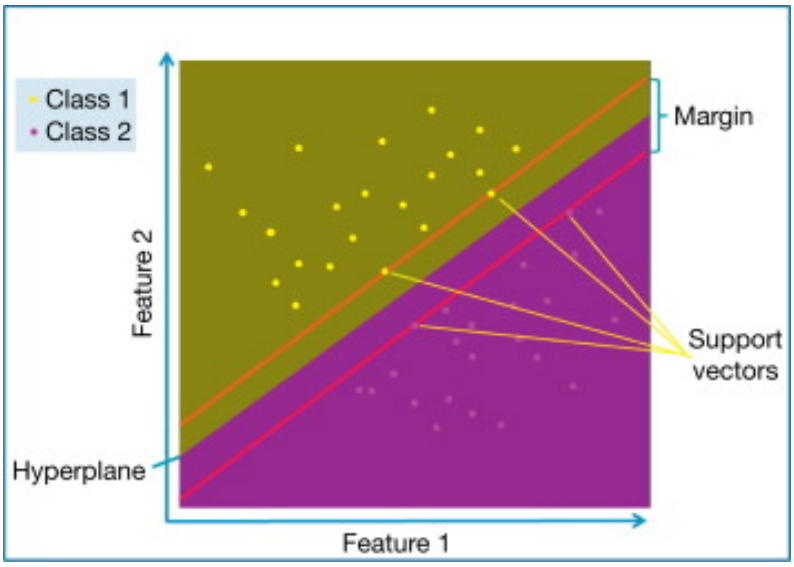

Figure 4.4: An example of the SVM classification process (reproduced from [32]).

4.6 *k*-Nearest Neighbours

The k-nearest neighbour (*k*NN) classifier is widely used in biomedical signal analysis. This nonparametric algorithm relies on the assumption that observations with similar characteristics will tend to have similar outcomes. Thus, it classifies a given test set by comparing with training sets that are similar [34].

Each training set is defined by a set of variables and signifies a point plotted in a multidimensional space [34]. Given a new point (test set), the aim is to find out the k training sets that are the closest to it. The closeness of the sets is defined with a distance metric, such as Euclidean distance, Mahalanobis distance, and city-block distance [35]. The number of neighbours k is chosen by cross-validation or by testing the quality of the classifier on another test set [35].

The test set is assigned to the class that has the most training sets out of the k training sets. Test sets for which there is no majority are classified randomly to one of the classes or left unclassified [35].

The classification process of *k*NN is illustrated in Figure 4.5, in which the two classes are depicted by a square and a circle, and the new point to be classified is represented by a triangle. In this example, the class of the test set is determined by identifying the 5 nearest neighbours and is classified as a square.

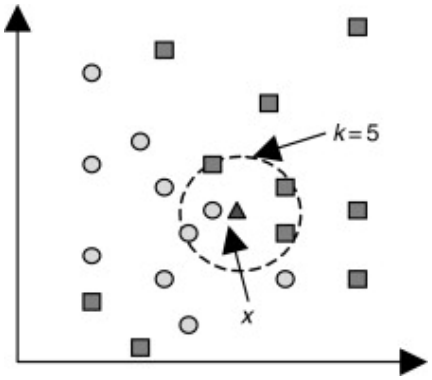

Figure 4.5: An example of the *k*NN classification process (reproduced from [36]).

4.7 Naïve Bayes

The naive Bayes classifier relies on the assumption that each feature makes an independent and equal contribution to the target class. In other words, it assumes that each feature is independent and does not interact with the others so that each feature independently and equally contributes to the probability of a data set to belong to a certain class [37].

As any other classifier, this algorithm must accomplish the task of assigning a class to a training set based on its feature values. To do that so, it processes the training set to approximate the probability of a class y_i for a given set of feature values (X_1, X_2, X_3), which is expressed as [37]:

$$P(y_i|X_1,X_2,X_3) \propto P(X_1|y_i)P(X_2|y_i)P(X_3|y_i)P(y_i) \tag{4.2}$$

In Eq. 5.1, individual $P(X_j|y_i)$, where j=1,2, or 3, can be calculated based on the assumption of the distributions of the features [37]. The training set is separated by class to calculate the parameters regarding the feature distributions. Then, these parameters are calculated for each class to allow the calculation of $P(X_j|y_i)$. Lastly, the algorithm estimates the probability of a new set with known feature values to belong to a certain class. The class assigned is the one with the highest probability, i.e., the one that leads to the largest value of $P(X_1|y_i)P(X_2|y_i)P(X_3|y_i)P(y_i)$ [37].

4.8 Ensembles

Ensemble learning intelligently aggregates predictions of multiple classifiers to obtain more accurate and robust classification results. An ensemble algorithm trains a set of base classifiers as ensemble members and performs classification by combining their predictions into a single output, hence creating the ensemble classifier [38]. Generally, this classifier has better classification performance than any other ensemble member.

There are three main steps in ensemble learning strategy for data classification [38]. First, the input data is split into a training set and a test set through a partitioning tool. Second, the training set is given as input to n classification algorithms that run simultaneously to build n independent classification models. In last, a voter tool accesses the *n* models and performs classification by assigning to each instance of the test set the class predicted by the majority of the *n* models generated.

Identification of optimal ways to combine the ensemble members is the key to good classification performance [30, 38]. An ensemble algorithm yields better results if there are significant differences or diversity among those models [30]. Common types of ensembles include Bayes optimal classifier, Bayesian parameter averaging, Bootstrap aggregating (Bagging), Boosting, Stacking, and Bucket of models [30].

4.9 Convolutional Neural Networks

Convolutional neural network (CNN) is a type of deep neural network, whose mechanism is inspired by the human neurons' connectivity pattern. This DL tool was originally designed to operate on 2D data, yet it was modified to also be implemented for 1D and 3D data classification.

An advantage of 1D-CNNs is their low computational complexity since the only operation with a significant computation cost is the sequence of 1D convolutions, which are simply linear weighted sums of two 1D arrays [39].

The configuration of a 1D-CNN is made up of an input layer, multiple hidden convolutional, pooling, and fully-connected layers, an output layer, a kernel size, a subsampling factor, and the choice of pooling and activation functions [39]. The input layer receives the 1D signal and the output layer is a fully-connected layer whose number of neurons is equal to the number of classes to be assigned [39]. Convolutional layers process the input data and perform 1D convolutions, pooling layers reduce the scale of feature vectors, and fully connected layers classify the extracted features [39]. The architecture of a typical 1D-CNN is presented in Figure 4.6.

The convolution operation is performed between convolutional layers [40]. The activation function, which typically is the rectified linear unit (ReLu) function, gives the results of those convolution calculations [40, 41]. After convolutional and pooling blocks, the previously extracted features are fed to a set of fully connected layers to perform classification. In multiclass classification is commonly implemented a softmax layer, which makes use of the softmax function to determine the probability of each class label over the number of classes [41].

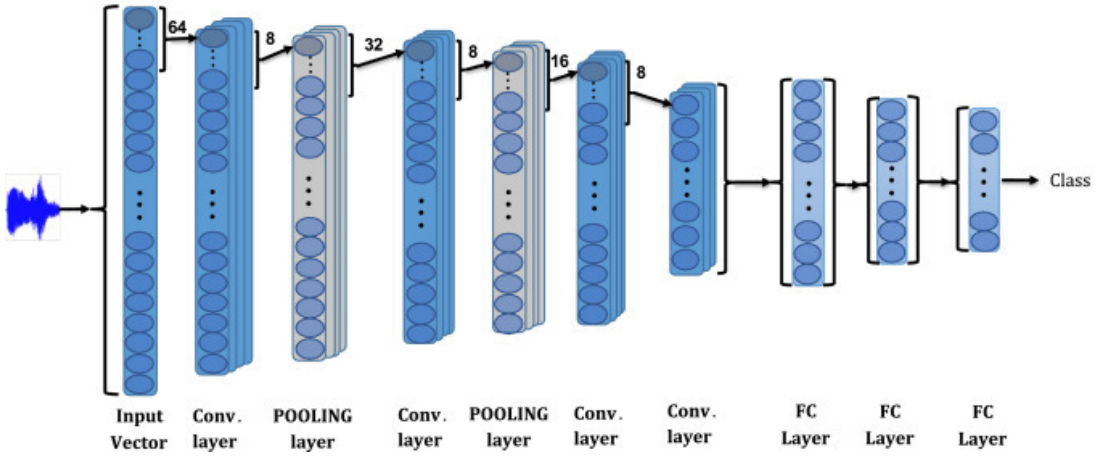

Figure 4.6: A sample 1D-CNN architecture (reproduced from [42]).

CHAPTER 5

Methodology

5.1 Database Characterization

The EEG recordings analyzed were obtained from a publicly available database [43]. This database comprises EEG signals from 34 individuals, of which 25 are healthy control subjects and 9 are patients diagnosed with depression. Every recording was acquired at a sampling rate of 500 Hz by means of 67 electrodes on the scalp of every individual.

Only the first 63 channels were considered, as the last four (64 to 67) correspond to ECG recordings. Thus, 63 EEG channels have been analyzed per subject.

5.2 Methodology

To carry through this study, the methodology was implemented by using MATLAB® 2019b. The developed algorithm is summarized in Figure 5.1.

The EEG signals were firstly loaded from the database to MATLAB® and thereafter normalized according to the following equation [22]:

$$x(n) = \frac{x(n)}{\sum_n^N x(n)} \tag{5.1}$$

where N represents the signals' size.

Secondly, a Butterworth filter with cutoff frequencies of 2-50 Hz was added to remove any possible artifact or noise from the signals. Then, the signals were split into segments of 5s (2500 samples).

After preprocessing, all signals per segments were decomposed through 1D Discrete Wavelet Transform. The X-ROC classifier was used to select the mother wavelet with the best performance.

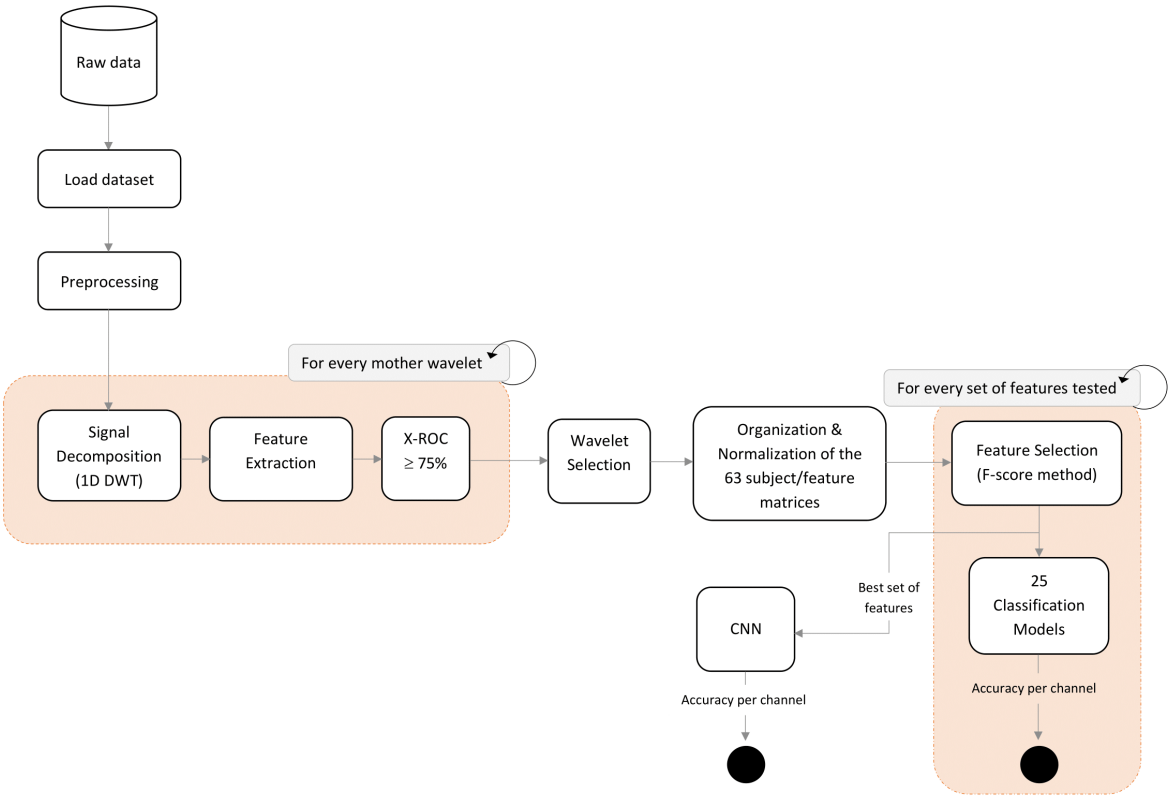

Figure 5.1: Diagram of the methodology implemented in this study.

Data organization and normalization took place prior to feature selection. The data was organized in a set of 63 subject/feature matrices. Each matrix held all the data associated to a channel, so that an analysis per channel could be conducted.

Feature selection was performed in each subject/feature matrix. The resulting matrices were fed individually to 25 automatically trained classification models to determine the best combination of selected features. A classification accuracy was assigned to each matrix, thus indicating the accuracy per channel.

Finally, the matrices containing the best combination of selected features were fed individually to a CNN to perform classification, from which an accuracy per channel was obtained.

5.2.1 1D Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is a time-frequency analysis method used to decompose a given signal into several sets, where each set is a time series of coefficients that describe the time evolution of the signal in the corresponding decomposition level, i.e., frequency band [44].

The DWT of a signal is defined based on approximation (low frequency) coefficients and detail (high frequency) coefficients [45]. Across a filter bank structure, the signal is passed through a low pass filter and a high pass filter, followed by downsampling by a factor of two. The low pass signal is again filtered by both filters and sub-sampled by a factor of two to obtain the next level approximation and detail coefficients [45]. This process is repeated for a multiband decomposition, in which each level corresponds to a specific frequency sub-band.

The process is performed by means of an analysis scale function $\widetilde{\phi}_1$(n) and an analysis wavelet function $\widetilde{\varphi}_1$(n), which are the impulse responses of the low and high pass filters, respectively [40]. Defining the following recursive formulas [40]:

$$\widetilde{\phi}_{i+1}(n) = \widetilde{\phi}_i(n/2) * \widetilde{\phi}_i(n) \tag{5.2}$$

$$\widetilde{\varphi}_{i+1}(n) = \widetilde{\phi}_i(n) * \widetilde{\varphi}_i(n/2) \tag{5.3}$$

the equivalent filter of the *m*th sub-band is given by [40]:

$$h_m(n) = \begin{cases} \widetilde{\phi}_S(n), m = 0 \\ \widetilde{\varphi}_{S+1-m}(n), m = 1,2,...,S \end{cases} \tag{5.4}$$

and the signal of the *m*th sub-band of the signal $x(n)$ is defined as [40]:

$$x_m(n) = \begin{cases} \sum_{k=-\infty}^{\infty} x(k) h_m(2^S n - k), m = 0 \\ \sum_{k=-\infty}^{\infty} x(k) h_m(2^{S-m+1} n - k), m = 1,2,...,S \end{cases} \tag{5.5}$$

In this study, the DWT was applied to the EEG segments of every subject with the purpose of decomposing them into the conventional EEG sub-bands (ρ, θ, α, β, and γ). The decomposition level was chosen up to six (S=6). The signals of the sub-bands relate to the EEG sub-bands as represented in Figure 5.2.

5.2.2 Feature Extraction

The choice of suitable features that can best represent the characteristics of EEG signals is the key for efficient EEG classification [46]. DWT coefficients from the six frequency sub-bands construct the feature matrix of one EEG segment [46].

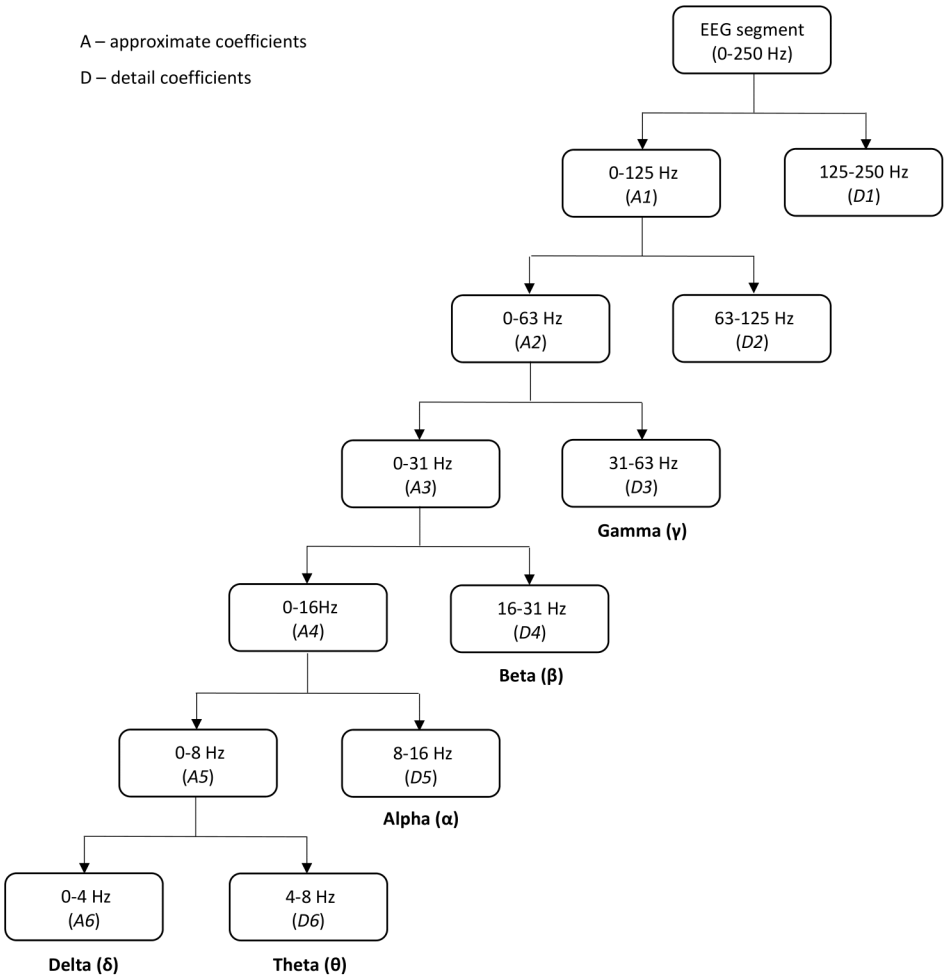

Figure 5.2: 1D DWT 6-level decomposition and their corresponding frequency sub-bands.

Time-domain features are the most intuitive EEG features. The time-domain linear features extracted in this study include the number of peaks, standard deviation, skewness, kurtosis, mean energy, Hjorth activity, Hjorth mobility, and Hjorth complexity.

Considering EEG signals contain some of the characteristics of the nonlinear dynamics system, analyzing them based on the nonlinear dynamics theory has come up to research [8]. The nonlinear features extracted in this study include Lyapunov exponent, Shannon entropy, log energy entropy, and approximate entropy.

For each of the 34 subjects, the twelve features mentioned were calculated for the six sub-bands of each EEG segment of all 63 channels and used to discriminate, by sub-band and channel, within the study group Depressed vs Control.

These features are briefly described below, where the signal $x(n)$ has size N and mean $\overline{x}$.

5.2.2.1 Linear Features

5.2.2.1.1 Number of Peaks

The number of peaks (np) is the number of values that are higher than their root mean square value (RMS), which is calculated by [47]:

$$RMS = \sqrt{\frac{\sum_{n=1}^{N} |x(n)|^2}{N}} \tag{5.6}$$

5.2.2.1.2 Standard Deviation

The standard deviation (std) is a measure of how far the signal fluctuates from its mean value [48]. This statistical parameter is calculated by [47]:

$$std = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (x(n) - \bar{x})^2} \tag{5.7}$$

5.2.2.1.3 Skewness

Skewness ($skew$) measures the asymmetry in a statistical distribution around the signal's mean value and is defined by the following equations [47]:

$$skew = \frac{M_3}{M_2} * \sqrt{M_2} \tag{5.8}$$

where

$$M_k = \frac{1}{N} \sum_{n=1}^{N} (x(n) - \bar{x})^k \tag{5.9}$$

5.2.2.1.4 Kurtosis

Kurtosis ($kurt$) measures the shape of a statistical distribution compared with the normal distribution and is defined as [47]:

$$kurt = \frac{M_4}{M_2 M_2} \tag{5.10}$$

where M is defined in (5.9).

5.2.2.1.5 Mean Energy

The mean energy (ME) measures the mean of the energy of the sub-bands, which is defined as [40]:

$$E = \sum_{n=1}^{N} |x(n)|^2 \tag{5.11}$$

5.2.2.1.6 Hjorth Parameters

The Hjorth parameters are indicators of statistical properties originally developed by Bo Hjorth for EEG signal analysis [49]. He presented a set of three parameters based on time-domain properties: activity, mobility, and complexity.

Hjorth activity (HA), also referred to as variance, gives a measure of the squared standard deviation of the amplitude of the signal [50].

Hjorth mobility (HM), which represents an estimate of the mean frequency, is proportional to the standard deviation of the power spectrum [8, 47].

Hjorth complexity (HC), which represents an estimate of the bandwidth of the signal, compares the similarity of the shape of the signal with a pure sine wave [47, 50].

These parameters are calculated based on the following derivations [50]:

$$HA = \sigma^2_{x(n)} \tag{5.12}$$

$$HM = \frac{\sigma_d}{\sigma_{x(n)}} \tag{5.13}$$

$$HM = \frac{\frac{\sigma_{dd}}{\sigma_d}}{\frac{\sigma_d}{\sigma_{x(n)}}} \tag{5.14}$$

where $\sigma^2_{x(n)}$ is the variance of the signal $x(n)$; σ_d is the standard deviation of the first derivate of $x(n)$; σ_{dd} is the standard deviation of the second derivate of $x(n)$. An advantage of using Hjorth parameters is that their calculation only involves the variance of the signal, making the cost of their computation very low [51].

5.2.2.2 Nonlinear features

Most of the studies on neurological impairments analyze EEGs with the help of conventional linear techniques, such as spectral domain analysis or waveform investigations [52]. However, because EEG is a complex and irregular signal with nonlinear behaviors and without a linear relation between

cause and effect, the complex dynamic variations in an EEG time series cannot be described by linear techniques [52].

Complexity estimators such as largest Lyapunov exponents, correlation dimension, mutual information, fractal dimension, Shannon entropy, sample entropy, approximate entropy, log energy entropy, Lempel-Ziv complexity, and Kolmogorov complexity have been proven to be more suitable techniques to comprehend the underlying dynamics of brain activity with unknown parameters [52]. It must be noted that these nonlinear measures reflect the complexity with different approaches. For instance, while correlation dimension is related to dimensional complexity, Lyapunov exponents and entropies focus on determining the irregularity and predictability of the signals [52].

The concept of phase space is essential to the analysis of nonlinear dynamics [53]. In a system composed of *m* variables, the phase space is *m*-dimensional. Each state of the system matches to a point in the phase space whose *m*-coordinates are the values assumed by the variables that compose this specific state [53]. If the system is observed for a long period, the sequence of points in the phase space forms a trajectory. This trajectory fills a subspace of the phase space named the system's attractor [53]. The process of representing the phase space of a dynamic system from a signal is defined as attractor's reconstruction [40]. The attractor's reconstruction in the phase space can be held by the technique of delay coordinates, which is formulated by [40]:

$$x_i = [x(i), x(i+\tau), ..., x(i+(m-1)\tau] \tag{5.15}$$

where τ is the delay.

The sequence of vectors x_i, i =1,2,...,M, where $M = N-(m-1)\tau$, compose the reconstructed attractor [40].

5.2.2.2.1 Largest Lyapunov Exponent

The Lyapunov exponents quantify the exponential divergence or convergence of initially nearby trajectories in phase space, therefore characterizing the predictability of a dynamic system [54].

A *m*-dimensional dynamical system has m Lyapunov exponents. However, in most applications (including this one), the largest Lyapunov exponent (*LLE*) is computed instead of all exponents [54]. A positive *LLE* indicates that divergence along initially nearby trajectories grows in time, hence the chaos of the nonlinear system [54].

This measure can be estimated by, for each state x_i, finding the state x_j that satisfies $min_j \|x_i - x_j\|$ such that $|i-j| > T_m$, where T_m is the mean period of the signal $x(n)$. Defining the following equation [40]:

$$\lambda(i) = \frac{1}{M+2} \sum_{k=1}^{M} \frac{1}{kT_s} \ln \frac{\|x_{i+k} - x_{j+k}\|}{\|x_i - x_j\|} \tag{5.16}$$

where T_s is the sampling period, *LLE* is given by the slope of the best linear approximation of $\lambda(i)$.

5.2.2.2.2 Shannon Entropy

Entropy is a measure that describes the amount of disorder and uncertainty in a system [40]. Shannon entropy is a concept introduced by Claude E. Shannon, in which given partial information about a system, a measure of the uncertainty of occurrence of a certain event is proposed [55].

Regarding EEG analysis, this entropy measures the predictability of future amplitude values based on the probability distribution of amplitude values previously observed in the signal [56]. In other words, it quantifies the probability density function of values [56].

Shannon entropy (*SEn*) can be estimated by [40]:

$$SEn = -\sum_{n=1}^{N} |x(n)|^2 \log[|x(n)|^2] \tag{5.17}$$

5.2.2.2.3 Log Energy Entropy

This feature has been frequently reported in the literature for EEG analysis. Similar to Shannon entropy, log energy entropy (*LEEn*) is defined by the following equation [40]:

$$LEEn = \sum_{n=1}^{N} \log[|x(n)|^2] \tag{5.18}$$

5.2.2.2.4 Approximate Entropy

Approximate entropy provides a general understanding of the regularity of a time series [57, 58]. It assigns a non-negative number to quantify the complexity or irregularity in the time series, with larger numbers corresponding to more complexity/irregularity in the data [57, 58].

This entropy discriminates series for which clear feature recognition is difficult and detects changes in underlying episodic behaviour not reflected in peak occurrences or amplitudes [58].

Approximate entropy (*ApEn*) can be estimated by [40]:

$$ApEn(m,r,N) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log[C_i^m(r)] - \frac{1}{N-m} \sum_{i=1}^{N-m} \log[C_i^{m+1}(r)] \tag{5.19}$$

where

$$c_i^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} \theta(r - \|x_i - x_j\|) \tag{5.20}$$

is the probability that point x_i on the attractor is separated from the other points by a distance less than or equal to r.

5.2.3 Mother Wavelet and Wavelet Families

The values obtained for each feature depend on the mother wavelet ($\widetilde{\varphi}_1(n)$) used in the decomposition. To find the mother wavelet that results in features with the highest discriminating ability for the group of study Depressed vs Control, there were tested 52 mother wavelets. The wavelet families evaluated include Biorthogonal (`bior`), Reverse biorthogonal (`rbio`), Coiflets (`coif`), Symlets (`sym`) and Daubechies (`db`). The tested wavelets, all contained in these families, are presented in Table 5.1.

Table 5.1: The fifty-two tested mother wavelets and their corresponding families

Wavelet family	**Mother wavelets**
Biorthogonal (`bior`)	bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8
Reverse biorthogonal (`rbio`)	rbio1.3, rbio1.5, rbio2.2, rbio2.4, rbio2.6, rbio2.8, rbio3.1, rbio3.3, rbio3.5, rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8
Coiflets (`coif`)	coif1, coif2, coif3, coif4, coif5
Symlets (`sym`)	sym2, sym3, sym4, sym5, sym6, sym7, sym8, sym9, sym10
Daubechies (`db`)	db1, db2, db3, db4, db5, db6, db7, db8, db9, db10

The features' values calculated from the EEG segments were separated by wavelet and subsequently by feature. For each mother wavelet analysis, the data was classified by feature using the leave-one-out cross-validation ROC tool for the binary comparison Depressed vs Control. The selection was based on the number of accuracies above or equal to 75% that each mother wavelet signal analysis presented per feature. Thus, the higher the number of accuracies above or equal to 75% per feature, the better the wavelet's performance. The mother wavelet selected to decompose all segments in each sub-band, from each channel, and all subjects, was the reverse biorthogonal 3.1 (rbio3.1).

The data obtained from the chosen wavelet was organized in a set of 63 matrices, where each matrix contains all the information regarding a channel. For a specific channel, each matrix contains the features' values of the twelve features in each sub-band of every subject. The values of every matrix were normalized according to (5.1).

5.2.4 Feature Selection and Classification

A proper representation of data from all features is crucial for efficient data classification. Some of the original features are not beneficial to the classification task, meaning they are irrelevant/redundant or noise in the distribution of the dataset [59]. Therefore, to reduce the computation cost of the classifier and increase classification performance, the process of feature selection should be applied [59].

The F-score feature selection method measures the distinction between two classes with real values [60]. It begins with computing the F-score values of each feature in the dataset. Then, to select the features from the whole dataset, the threshold value is obtained by calculating the average value of F-scores of all features. If the F-score value of a feature is higher than the threshold value, that feature is added to the feature space. Elsewise, that feature is removed from feature space.

Given the training vectors if the number of positive and negative instances are n^+ and n^-, respectively, the F-score $F(i)$ of the ith feature is defined by the following equation [60]:

$$F(i) = \frac{(\bar{x}_i^+ - \bar{x}_i)^2 + (\bar{x}_i^- - \bar{x}_i)^2}{\frac{1}{n^+-1}\sum_{k=1}^{n^+}((\bar{x}_{k,i}^{(+)} - \bar{x}_i^{(+)})^2) + \frac{1}{n^--1}\sum_{k=1}^{n^-}((\bar{x}_{k,i}^{(-)} - \bar{x}_i^{(-)})^2)} \tag{5.21}$$

where $\bar{x}_i$, $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$) are the averages of the ith feature of the whole, positive, and negative sets, respectively; $\bar{x}_{k,i}^{(+)}$ is the ith feature of the kth positive instance and $\bar{x}_{k,i}^{(-)}$) is the ith feature of the kth negative instance.

The equation's numerator indicates the discrimination between the positive and negative sets, and the equation's denominator indicates the discrimination within each of those sets [60]. According to the method, the higher the F-score, the more discriminative a given feature is likely to be [60].

The F-score method was applied to each of the 63 matrices to select the best 1000 features per matrix. The process was repeated to select the best 5, 10, 33, 82, 164 and 328 features (i.e., 0.15%, 0.3%, 1%, 2.5%, 5% and 10% of all features). Thereafter, to find the most discriminative set of selected features, the matrices individually fed 25 classification models that were trained automatically with resource to Classification Learner, a MATLAB® app. A leave-one-out cross-validation was employed in every classification task.

The models are related to the following classifiers: decision trees, discriminant analysis, SVMs, logistic regression, *k*NNs, ensembles, and naïve Bayes. All of them are listed in Table 5.2.

The process described above was repeated for 1%, 2.5%, 5%, and 10% of total selected features and 5, 10, and 1000 selected features.

For each model, a classification accuracy was assigned to each matrix, thus indicating the accuracy per channel. Based on the number of accuracies achieved in each channel classification, equal or above 90%, the set of features with the highest discriminating ability for the binary comparison Depressed vs Control was proven to be constituted by the set of 327 features.

The 63 matrices holding 327 features individually fed a CNN to discriminate between the study group per channel. Training was realized using a leave-one-out cross-validation and three different training algorithms: stochastic gradient descent with momentum (sgdm), adaptive moment estimation (adam) and root mean square propagation (rmsprop). A classification accuracy was assigned to each matrix, thus indicating the accuracy per channel.

The architecture of the CNN developed in this study consisted of eight layers: a 1D signal input layer, two convolutional layers, a ReLu layer, two fully connected layers, a softmax layer, and an

Table 5.2: The twenty-five trained classification models and their corresponding classifiers.

Classifier	**Classification Model**
Decision trees	Fine Tree
	Medium Tree
	Coarse Tree
Discriminant Analysis	Linear Discriminant
	Quadratic Discriminant
Support Vector Machines (SVMs)	Linear SVM
	Quadratic SVM
	Cubic SVM
	Fine Gaussian SVM
	Medium Gaussian SVM
	Coarse Gaussian SVM
Logistic Regression	
K-nearest Neighbours (KNNs)	Fine KNN
	Medium KNN
	Coarse KNN
	Cosine KNN
	Cubic KNN
	Weighted KNN
Ensembles	Boosted Trees
	Bagged Trees
	Subspace Discriminant
	Subspace KNN
	RUSBoosted Trees
Naïve Bayes	Gaussian Naïve Baye
	Kernel Naïve Baye

output (classification) layer. The input data was convolved with a kernel size of [2 1] and a stride size of 12.

CHAPTER 6

Results and Discussion

Out of the 25 classic Machine Learning trained models, for the best set of selected features, the two classification models with the highest discriminating ability for the binary comparison Depressed vs Control were Linear Discriminant and Cubic SVM, with an overall classification accuracy of 94.8% and 93.9%, respectively. The accuracies obtained by the other classification models are presented in Appendix A.

Regarding the Linear Discriminant model, the channels given the lowest classification accuracy (above 80% but below 90%) were F2, FC1, T7, C5, TP7, CP3, CP4, and CP6. The channels classified with an accuracy of 100% were AF4, F4, FC4, C3, C1, CP2, PO5, PO3, POZ, and O2.

Regarding the Cubic SVM, the channels given the lowest classification accuracy (above 85% but below 90%) were F2, T7, C5, C6, CP3, CP4, CP6, and O1. The channels classified with an accuracy of 100% were AF4, F6, C3, C1, P7, and POZ.

As for the designed CNN, the network trained with the root mean square propagation (rmsprop) technique had the highest discriminating ability for the binary comparison Depressed vs Control, with an overall classification accuracy of 94.9%. The accuracies obtained by means of the other training techniques are presented in Appendix A. The channels classified with the lowest classification accuracy (above 80% but below 90%) were F2, T7, C6, T8, TP7, CP4, and CP6. The channels classified with an accuracy of 100% were F7, F4, FC4, C3, C1, CZ, C2, CP2, P5, P2, POZ, PO4, O1, and OZ. The accuracies obtained by means of the other training techniques are presented in Appendix A.

The results of the CNN and the two best classification models are indicated on the topographic maps represented in Figure 6.1. Each topographic map shows the results obtained for the 63 channels.

The topographic maps show that there are strong differences between Depressed and Control in the right frontal lobe, left central lobe, central/posterior parietal lobe, and central occipital lobe. The results of the electrodes placed on these regions suggest that there is a difference in the frequency of

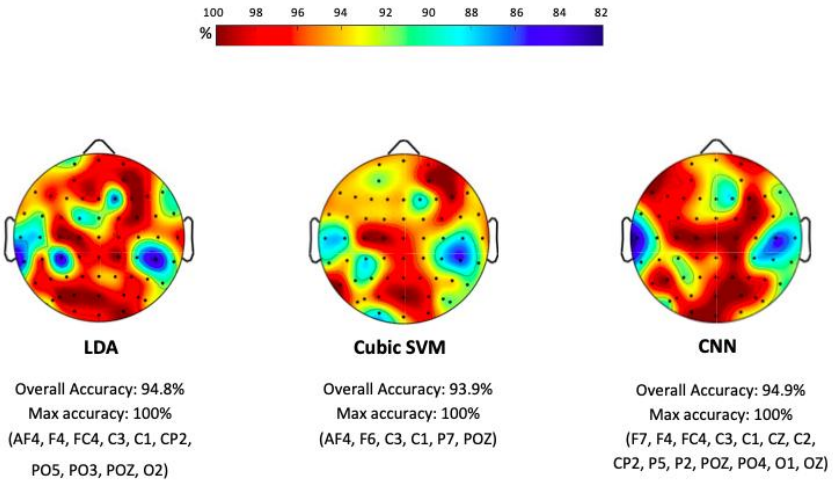

Figure 6.1: Topographic maps of the results that were obtained by means of three classifiers: Linear Discriminant Analysis (LDA), Cubic SVM, and CNN, and the respective channels classified with an accuracy of 100%.

brain waves between depressed subjects and healthy controls. The classifier that best distinguished between these two groups was CNN (94.9%).

Regarding these three classifiers, the features that had the most weight in each classification task are listed in Table 6.1, Table 6.2, and Table 6.3. These features were not the only ones to be selected by the F-score method, yet they were the ones that were selected the most in each binary classification per channel.

Table 6.1: The most relevant features selected by the F-score method within each sub-band of the channels classified with an accuracy of 100% by LDA.

Channels	**Delta (δ)**	**Theta (θ)**	**Alpha (α)**	**Beta (β)**	**Gamma (γ)**
AF4	LLE	HM	np	np	np
F4	skew	HM	np	np	np
FC4	np, HM, skew	SEn, LLE, kurt	SEn	np	HM
C3	HC, HM	np, LLE	HC, HM	np	HM
C1	HM	ApEn	np	np	np, HM
CP2	ApEn	ApEn, HM, skew	ApEn, np	np	HC
PO5	ApEn, SEn	np	HM	np	HM
PO3	ApEn	HC	np	np	HM
POZ	ApEn	ApEn	np	np	HM
O2	skew	skew	np	np	HM

Table 6.2: The most relevant features selected by the F-score method within each sub-band of the channels classified with an accuracy of 100% by Cubic SVM.

Channels	**Delta (δ)**	**Theta (θ)**	**Alpha (α)**	**Beta (β)**	**Gamma (γ)**
AF4	LLE	HM	np	np	np
F7	HM	HC	np	np	HM
C3	HC, HM	np, LLE	HC, HM	np	HM
C1	HM	ApEn	np	np	np, HM
P7	skew	skew	HM	np	HM
POZ	ApEn	ApEn	np	np	HM

Table 6.3: The most relevant features selected by the F-score method within each sub-band of the channels classified with an accuracy of 100% by CNN.

Channels	**Delta (δ)**	**Theta (θ)**	**Alpha (α)**	**Beta (β)**	**Gamma (γ)**
F7	HM	HM	np	np	HM
F4	skew	HM	np	np	np
FC4	np, HM, skew	SEn, LLE, kurt	SEn	np	HM
C3	HC, HM	np, LLE	HC, HM	np	HM
C1	HM	ApEn	np	np	np, HM
CZ	ApEn, np, skew	np	SEn	ApEn	skew
CP2	ApEn	ApEn, HM, skew	ApEn, np	np	HC
P5	HC, LLE, skew	kurt, SEn	HM	np	HC
P2	np	LLE	np	np	HC, HM
POZ	ApEn	ApEn	np	np	HM
PO4	LLE	HM	np	np	HM
O1	LLE	np	np	np	HM
OZ	LEEn	LEEn	LEEn	np	HM

Overall, the most selected features for classification were approximate entropy, number of peaks, Hjorth complexity, Hjorth mobility, Lyapunov exponent, and skewness.

In addition, for the channels indicated, the best set of features was mostly selected out of features from gamma and beta sub-bands. This indicates that the features selected from these sub-bands had more discriminative power in the binary classification per channel, which can be explained by the relationship between these sub-bands and the cognitive dysfunction of depressed subjects reported in other studies [61, 62].

CHAPTER 7

Conclusion and Future Perspectives

Epidemiological studies show that depression is a major contributor to the overall global burden of disease and a leading cause of suicide. Considering this is plausible to state that the diagnosis of this neurological disorder in the early curable stages is critical and might save lives.

In comparison with other diagnostic methods, the collection and analysis of EEG signals for the diagnosis of depression open a new path for a cost-effective, objective, less human-intensive depression detection tool. By means of signal processing methods, the brain activity of a depressed subject can be distinguished from that of a healthy subject.

With resources to 1D Discrete Wavelet Transform and several AI techniques, the present study analyzed EEG signals from depressed subjects and healthy controls with the purpose of evaluating the differences between these groups and automatically discriminate the patients with onset of depression.

The results obtained by the applied classifiers suggest that there is a difference in the frequency of brain waves between depressed subjects and healthy controls, with a strong incidence in the frontocentral, central, and parietooccipital regions of the scalp. The classifier that best distinguished between the two groups was CNN, with an overall classification accuracy of 94.9%.

The features with the most weight in the classification process were approximate entropy, number of peaks, Hjorth complexity, Hjorth mobility, Lyapunov exponent, and skewness. The best set of features was mostly selected out of features from gamma and beta sub-bands, which leads to assuming that the features selected from these sub-bands had more discriminative power in the classification process.

Even though the small size of the database compromises the conclusions drawn, it is possible to infer that the techniques implemented in the developed algorithm are useful for depression assessment. EEG signal analysis cannot yet be applied as a diagnostic tool for depression, nonetheless, the findings in this study remain relevant from a theoretical point of view.

Further work must be done to validate and improve the developed algorithm. Other analyses can be conducted such as the search for other features that, coupled with the ones chosen in this study, could have better results; the use of a larger database to allow a more realistic classification; the development of a more robust algorithm that could differentiate between types of depression; and the design of a graphical user interface.

APPENDIX A

Accuracies of All Binary Classifications Per Classifier

Table A.1: Accuracies of the binary classifications using Decision Trees and Discriminant Analysis.

Channels	Fine Tree	Medium Tree	Coarse Tree	Linear Discriminant	Quadratic Discriminant
1	64.7	64.7	64.7	91.2	FAILED
2	70.6	70.6	70.6	97.1	FAILED
3	88.2	88.2	88.2	97.1	FAILED
4	82.4	82.4	82.4	97.1	FAILED
5	47.1	47.1	47.1	100	FAILED
6	70.6	70.6	70.6	94.1	FAILED
7	76.5	76.5	76.5	97.1	FAILED
8	64.7	64.7	64.7	94.1	FAILED
9	70.6	70.6	70.6	97.1	FAILED
10	50.0	50.0	50.0	97.1	FAILED
11	70.6	70.6	70.6	85.3	FAILED
12	58.8	58.8	58.8	100	FAILED
13	73.5	73.5	73.5	97.1	FAILED
14	85.3	85.3	85.3	91.2	FAILED
15	85.3	85.3	85.3	94.1	FAILED
16	67.6	67.6	67.6	91.2	FAILED
17	70.6	70.6	70.6	97.1	FAILED

Continued on next page

Table A.1 – continued from previous page

Channels	Fine Tree	Medium Tree	Coarse Tree	Linear Discriminant	Quadratic Discriminant
18	52.9	52.9	52.9	88.2	FAILED
19	79.4	79.4	79.4	91.2	FAILED
20	79.4	79.4	79.4	97.1	FAILED
21	64.7	64.7	64.7	100	FAILED
22	97.1	97.1	97.1	94.1	FAILED
23	76.5	76.5	76.5	91.2	FAILED
24	70.6	70.6	70.6	88.2	FAILED
25	82.4	82.4	82.4	88.2	FAILED
26	73.5	73.5	73.5	100	FAILED
27	85.3	85.3	85.3	100	FAILED
28	79.4	79.4	79.4	94.1	FAILED
29	67.6	67.6	67.6	97.1	FAILED
30	82.4	82.4	82.4	94.1	FAILED
31	85.3	85.3	85.3	94.1	FAILED
32	91.2	91.2	91.2	97.1	FAILED
33	91.2	91.2	91.2	85.3	FAILED
34	85.3	85.3	85.3	97.1	FAILED
35	70.6	70.6	70.6	82.4	FAILED
36	61.8	61.8	61.8	97.1	FAILED
37	64.7	64.7	64.7	97.1	FAILED
38	82.4	82.4	82.4	100	FAILED
39	58.8	58.8	58.8	88.2	FAILED
40	70.6	70.6	70.6	82.4	FAILED
41	91.2	91.2	91.2	91.2	FAILED
42	58.8	58.8	58.8	97.1	FAILED
43	55.9	55.9	55.9	97.1	FAILED
44	61.8	61.8	61.8	94.1	FAILED
45	67.6	67.6	67.6	97.1	FAILED
46	73.5	73.5	73.5	94.1	FAILED
47	67.6	67.6	67.6	94.1	FAILED
48	82.4	82.4	82.4	97.1	FAILED
49	58.8	58.8	58.8	94.1	FAILED

Continued on next page

Table A.1 – continued from previous page

Channels	Fine Tree	Medium Tree	Coarse Tree	Linear Discriminant	Quadratic Discriminant
50	82.4	82.4	82.4	91.2	FAILED
51	88.2	88.2	88.2	94.1	FAILED
52	64.7	64.7	64.7	100	FAILED
53	73.5	73.5	73.5	100	FAILED
54	82.4	82.4	82.4	100	FAILED
55	61.8	61.8	61.8	97.1	FAILED
56	73.5	73.5	73.5	97.1	FAILED
57	85.3	85.3	85.3	94.1	FAILED
58	73.5	73.5	73.5	97.1	FAILED
59	58.8	58.8	58.8	97.1	FAILED
60	79.4	79.4	79.4	100	FAILED
61	35.3	35.3	35.3	97.1	FAILED
62	50.0	50.0	50.0	97.1	FAILED
63	58.8	58.8	58.8	94.1	FAILED
Overall Accuracy	72.0	72.0	72.0	94.8	Non applicable

Table A.2: Accuracies of the binary classifications using SVM.

Channels	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
1	82.4	91.2	94.1	73.5	91.2	73.5
2	85.3	94.1	94.1	73.5	79.4	73.5
3	79.4	97.1	97.1	73.5	82.4	73.5
4	76.5	91.2	91.2	73.5	88.2	73.5
5	79.4	100	100	73.5	91.2	73.5
6	88.2	97.1	94.1	73.5	82.4	73.5
7	94.1	94.1	94.1	73.5	85.3	73.5
8	91.2	94.1	94.1	73.5	85.3	73.5
9	94.1	94.1	94.1	73.5	91.2	73.5
10	88.2	91.2	94.1	73.5	88.2	73.5

Continued on next page

Table A.2 – continued from previous page

Channels	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
11	82.4	88.2	88.2	73.5	88.2	73.5
12	85.3	94.1	94.1	73.5	82.4	73.5
13	88.2	97.1	100	73.5	82.4	73.5
14	88.2	94.1	94.1	73.5	91.2	73.5
15	88.2	97.1	94.1	73.5	88.2	73.5
16	88.2	97.1	94.1	73.5	82.4	73.5
17	91.2	91.2	94.1	73.5	82.4	73.5
18	88.2	91.2	94.1	73.5	82.4	73.5
19	91.2	94.1	94.1	73.5	85.3	73.5
20	88.2	94.1	94.1	73.5	88.2	73.5
21	85.3	94.1	97.1	73.5	85.3	73.5
22	85.3	94.1	94.1	73.5	88.2	73.5
23	88.2	94.1	94.1	73.5	85.3	73.5
24	85.3	88.2	88.2	73.5	82.4	73.5
25	85.3	88.2	88.2	73.5	82.4	73.5
26	88.2	100	100	73.5	94.1	73.5
27	88.2	100	100	73.5	97.1	73.5
28	94.1	88.2	97.1	73.5	82.4	73.5
29	88.2	94.1	94.1	73.5	79.4	73.5
30	85.3	88.2	91.2	73.5	82.4	73.5
31	82.4	91.2	88.2	73.5	88.2	73.5
32	82.4	91.2	94.1	73.5	91.2	73.5
33	82.4	91.2	91.2	73.5	79.4	73.5
34	91.2	94.1	94.1	73.5	82.4	73.5
35	85.3	88.2	88.2	73.5	76.5	73.5
36	79.4	94.1	94.1	73.5	88.2	73.5
37	85.3	88.2	97.1	73.5	85.3	73.5
38	85.3	94.1	94.1	73.5	88.2	73.5
39	79.4	82.4	88.2	73.5	88.2	73.5
40	82.4	82.4	85.3	73.5	82.4	73.5
41	73.5	88.2	91.2	73.5	76.5	73.5

Continued on next page

Table A.2 – continued from previous page

Channels	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
42	88.2	97.1	100	73.5	88.2	73.5
43	79.4	94.1	91.2	73.5	79.4	73.5
44	82.4	91.2	91.2	73.5	76.5	73.5
45	85.3	94.1	97.1	73.5	76.5	73.5
46	91.2	91.2	97.1	73.5	85.3	73.5
47	85.3	91.2	97.1	73.5	85.3	73.5
48	91.2	94.1	94.1	73.5	91.2	73.5
49	94.1	94.1	94.1	73.5	94.1	73.5
50	91.2	88.2	91.2	73.5	88.2	73.5
51	88.2	88.2	94.1	73.5	85.3	73.5
52	85.3	94.1	94.1	73.5	79.4	73.5
53	73.5	97.1	97.1	73.5	82.4	73.5
54	82.4	91.2	100	73.5	76.5	73.5
55	82.4	88.2	97.1	73.5	76.5	73.5
56	88.2	91.2	91.2	73.5	85.3	73.5
57	88.2	85.3	91.2	73.5	85.3	73.5
58	85.3	85.3	88.2	73.5	79.4	73.5
59	82.4	94.1	94.1	73.5	88.2	73.5
60	82.4	94.1	97.1	73.5	76.5	73.5
61	82.4	91.2	97.1	73.5	79.4	73.5
62	88.2	94.1	94.1	73.5	82.4	73.5
63	91.2	91.2	91.2	73.5	70.6	73.5
Overall Accuracy	85.9	92.3	93.9	73.5	84.4	73.5

Table A.3: Accuracies of the binary classifications using kNN.

Channels	Fine kNN	Medium kNN	Coarse kNN	Cosine Gaussian kNN	Cubic Gaussian kNN	Weighted Gaussian SVM
1	91.2	91.2	73.5	85.3	88.2	85.3
2	91.2	73.5	73.5	85.3	82.4	85.3
3	82.4	97.1	73.5	94.1	94.1	94.1
4	88.2	82.4	73.5	82.4	88.2	88.2
5	88.2	85.3	73.5	94.1	91.2	100
6	85.3	79.4	73.5	91.2	82.4	88.2
7	82.4	73.5	73.5	94.1	73.5	82.4
8	85.3	76.5	73.5	88.2	79.4	79.4
9	91.2	76.5	73.5	91.2	76.5	76.5
10	91.2	79.4	73.5	91.2	79.4	85.3
11	91.2	79.4	73.5	85.3	76.5	82.4
12	91.2	82.4	73.5	94.1	79.4	85.3
13	88.2	91.2	73.5	91.2	88.2	91.2
14	88.2	82.4	73.5	79.4	85.3	88.2
15	88.2	79.4	73.5	88.2	85.3	88.2
16	94.1	79.4	73.5	94.1	88.2	85.3
17	88.2	88.2	73.5	94.1	82.4	85.3
18	91.2	76.5	73.5	85.3	79.4	82.4
19	94.1	79.4	73.5	88.2	82.4	85.3
20	97.1	79.4	73.5	85.3	82.4	88.2
21	88.2	76.5	73.5	94.1	82.4	85.3
22	91.2	79.4	73.5	94.1	79.4	85.3
23	94.1	73.5	73.5	85.3	79.4	91.2
24	85.3	82.4	73.5	82.4	85.3	85.3
25	88.2	79.4	73.5	76.5	79.4	82.4
26	97.1	85.3	73.5	88.2	85.3	91.2
27	100	91.2	73.5	97.1	94.1	100
28	91.2	76.5	73.5	94.1	79.4	88.2
29	82.4	76.5	73.5	88.2	82.4	82.3
30	94.1	82.4	73.5	94.1	82.4	85.3
31	94.1	76.5	73.5	85.3	85.3	88.2

Continued on next page

Table A.3 – continued from previous page

Channels	Fine kNN	Medium kNN	Coarse kNN	Cosine Gaussian kNN	Cubic Gaussian kNN	Weighted Gaussian SVM
32	85.3	82.4	73.5	88.2	85.3	88.2
33	91.2	82.4	73.5	91.2	82.4	88.2
34	91.2	73.5	73.5	88.2	76.5	79.4
35	88.2	76.5	73.5	79.4	76.5	76.5
36	91.2	88.2	73.5	82.4	85.3	88.2
37	97.1	82.4	73.5	88.2	82.4	82.4
38	91.2	82.4	73.5	91.2	82.4	88.2
39	91.2	82.4	73.5	88.2	82.4	88.2
40	91.2	79.4	73.5	82.4	79.4	88.2
41	79.4	82.4	73.5	82.4	82.4	85.3
42	91.2	79.4	73.5	91.2	82.4	91.2
43	91.2	73.5	73.5	70.6	76.5	82.4
44	79.4	76.5	73.5	79.4	76.5	79.4
45	94.1	76.5	73.5	94.1	76.5	82.4
46	88.2	82.4	73.5	79.4	82.4	82.4
47	97.1	82.4	73.5	88.2	79.4	88.2
48	88.2	85.3	73.5	94.1	88.2	88.2
49	91.2	85.3	73.5	94.1	82.4	85.3
50	97.1	85.3	73.5	85.3	85.3	88.2
51	97.1	85.3	73.5	79.4	85.3	82.4
52	85.3	82.4	73.5	82.4	82.4	82.4
53	91.2	79.4	73.5	94.1	85.3	88.2
54	97.1	76.5	73.5	85.3	79.4	85.3
55	88.2	79.4	73.5	88.2	79.4	82.4
56	85.3	76.5	73.5	94.1	79.4	79.4
57	82.4	82.4	73.5	82.4	79.4	85.3
58	76.5	79.4	73.5	73.5	79.4	82.4
59	94.1	82.4	73.5	94.1	88.2	91.2
60	88.2	76.5	73.5	85.3	82.4	82.4
61	85.3	79.4	73.5	88.2	76.5	82.4
62	82.4	73.5	73.5	82.4	73.5	76.5

Continued on next page

Table A.3 – continued from previous page

Channels	Fine kNN	Medium kNN	Coarse kNN	Cosine Gaussian kNN	Cubic Gaussian kNN	Weighted Gaussian SVM
63	76.5	73.5	73.5	76.5	73.5	73.5
Overall Accuracy	89.5	80.6	73.5	87.4	82.0	85.5

Table A.4: Accuracies of the binary classifications using Ensembles.

Channels	Boosted Trees	Bagged Trees	Subspace Discriminant	Subspace kNN	RUSBoosted Trees
1	73.5	79.4	91.2	91.2	79.4
2	73.5	67.6	94.1	91.2	73.5
3	73.5	79.4	91.2	88.2	88.4
4	73.5	70.6	88.2	91.2	85.3
5	73.5	73.5	97.1	97.1	70.6
6	73.5	67.6	100	85.3	85.3
7	73.5	67.6	94.1	88.2	67.6
8	73.5	73.5	91.2	85.3	76.5
9	73.5	79.4	85.3	82.4	73.5
10	73.5	85.3	88.2	88.2	79.4
11	73.5	82.4	97.1	91.2	73.5
12	73.5	70.6	91.2	91.2	70.6
13	73.5	79.4	91.2	85.3	70.6
14	73.5	76.5	94.1	88.2	73.5
15	73.5	70.6	94.1	88.2	70.6
16	73.5	88.2	91.2	91.2	79.4
17	73.5	79.4	94.1	85.3	70.6
18	73.5	73.5	94.1	94.1	82.4
19	73.5	70.6	88.2	88.2	76.5
20	73.5	85.3	94.1	97.1	79.4
21	73.5	76.5	94.1	91.2	79.4
22	73.5	73.5	94.1	91.2	67.6
23	73.5	73.5	94.1	94.1	70.6

Continued on next page

Table A.4 – continued from previous page

Channels	Boosted Trees	Bagged Trees	Subspace Discriminant	Subspace kNN	RUSBoosted Trees
24	73.5	88.2	88.2	85.3	76.5
25	73.5	79.4	85.3	88.2	82.4
26	73.5	79.4	97.1	97.1	82.4
27	73.5	79.4	94.1	100	91.2
28	73.5	70.6	91.2	88.2	73.5
29	73.5	76.5	94.1	88.2	91.2
30	73.5	79.4	88.2	88.2	82.4
31	73.5	73.5	88.2	94.1	79.4
32	73.5	82.4	88.2	88.2	88.2
33	73.5	61.8	91.2	91.2	88.2
34	73.5	76.5	88.2	88.2	88.2
35	73.5	82.4	94.1	85.3	70.6
36	73.5	73.5	91.2	94.1	70.6
37	73.5	82.4	94.1	97.1	79.4
38	73.5	79.4	91.2	91.2	79.4
39	73.5	82.4	85.3	91.2	73.5
40	73.5	79.4	85.3	91.2	79.4
41	73.5	85.3	85.3	85.3	58.8
42	73.5	70.6	88.2	91.2	76.5
43	73.5	82.4	88.2	88.2	67.6
44	73.5	70.6	88.2	85.3	76.5
45	73.5	76.5	88.2	91.2	79.4
46	73.5	85.3	91.2	94.1	73.5
47	73.5	73.5	97.1	91.2	79.4
48	73.5	76.5	94.1	88.2	79.4
49	73.5	85.3	94.1	91.2	82.4
50	73.5	82.4	91.2	97.1	82.4
51	73.5	79.4	91.2	94.2	82.4
52	73.5	67.6	91.2	82.4	76.5
53	73.5	73.5	91.2	94.1	82.4
54	73.5	79.4	91.2	94.1	85.3
55	73.5	79.4	91.2	85.3	79.4

Continued on next page

Table A.4 – continued from previous page

Channels	Boosted Trees	Bagged Trees	Subspace Discriminant	Subspace kNN	RUSBoosted Trees
56	73.5	76.5	97.1	85.3	91.2
57	73.5	73.5	94.1	85.3	97.1
58	73.5	73.5	88.2	76.5	73.5
59	73.5	70.6	97.1	91.2	82.4
60	73.5	70.6	88.2	88.2	67.6
61	73.5	70.6	91.2	82.4	91.2
62	73.5	73.5	94.1	79.4	70.6
63	73.5	70.6	91.2	73.5	88.2
Overall Accuracy	89.5	80.6	73.5	87.4	82.0

Table A.5: Accuracies of the binary classifications using Ensembles.

Channels	Logistic Regression	Gaussian Naive Bayes	Kernel Naive Bayes	CNN (adam)	CNN (rmsprop)	CNN (sgdm)
1	79.4	88.2	82.4	91.2	91.2	94.1
2	73.5	58.8	85.3	94.1	91.2	94.1
3	79.4	91.2	79.4	100	97.1	98.0
4	85.3	82.4	76.5	94.1	91.2	97.1
5	88.2	91.2	79.4	100	97.1	94.1
6	94.1	88.2	88.2	97.1	97.1	100
7	91.2	91.2	94.1	94.1	91.2	97.1
8	91.2	94.1	91.2	97.1	94.1	97.1
9	85.3	91.2	94.1	94.1	88.2	94.1
10	91.2	94.1	88.2	97.1	91.2	91.2
11	85.3	91.2	82.4	88.2	88.2	88.2
12	79.4	91.2	85.3	100	94.1	100
13	70.6	91.2	88.2	97.1	97.1	94.1
14	82.4	91.2	88.2	91.2	91.2	91.2
15	76.5	88.2	88.2	94.1	88.2	91.2
16	82.4	88.2	88.2	97.1	94.1	97.1

Continued on next page

Table A.5 – continued from previous page

Channels	Logistic Regression	Gaussian Naive Bayes	Kernel Naive Bayes	CNN (adam)	CNN (rmsprop)	CNN (sgdm)
17	88.2	97.1	91.2	97.1	97.1	97.1
18	82.4	94.1	88.2	94.1	94.1	94.1
19	73.5	94.1	91.2	94.1	94.1	94.1
20	85.3	94.1	88.2	97.1	97.1	97.1
21	82.4	85.3	85.3	94.1	94.1	100
22	76.5	88.2	85.3	94.1	97.1	94.1
23	79.4	82.4	88.2	94.1	91.2	91.2
24	82.4	85.3	85.3	82.2	82.4	82.4
25	88.2	85.3	85.3	88.2	85.3	91.2
26	94.1	97.1	88.2	100	97.1	100
27	85.3	97.1	88.2	100	100	100
28	82.4	100	94.1	100	100	100
29	94.1	88.2	88.2	97.1	97.1	100
30	79.4	82.4	85.3	97.1	94.1	94.1
31	85.3	82.4	82.4	88.2	91.2	85.3
32	91.2	88.2	82.4	88.2	88.2	88.2
33	76.5	88.2	82.4	88.2	88.2	88.2
34	70.6	91.2	91.2	94.1	88.2	94.1
35	79.4	94.1	85.3	88.2	91.2	91.2
36	91.2	85.3	79.4	91.2	97.1	97.1
37	82.4	88.2	85.3	88.2	88.2	94.1
38	94.1	85.3	85.3	100	94.1	100
39	85.3	85.3	79.4	88.2	88.2	88.2
40	73.5	79.4	82.4	82.4	85.3	88.2
41	88.2	82.4	73.5	82.4	88.2	91.2
42	85.3	76.5	88.2	97.1	91.2	97.1
43	88.2	76.5	79.4	94.1	94.1	100
44	85.3	76.5	82.4	94.1	94.1	91.2
45	85.3	82.4	85.3	88.2	85.3	94.1
46	67.6	88.2	91.2	97.1	97.1	97.1
47	88.2	88.2	85.3	94.1	94.1	100

Continued on next page

Table A.5 – continued from previous page

Channels	Logistic Regression	Gaussian Naive Bayes	Kernel Naive Bayes	CNN (adam)	CNN (rmsprop)	CNN (sgdm)
48	91.2	94.1	91.2	97.1	97.1	97.1
49	94.1	91.2	94.1	97.1	97.1	97.1
50	79.4	85.3	91.2	91.2	91.2	91.2
51	94.1	91.2	88.2	91.2	91.2	91.2
52	91.2	79.4	85.3	94.1	94.1	97.1
53	97.1	88.2	73.5	94.1	94.1	94.1
54	85.3	91.2	82.4	100	97.1	100
55	88.2	91.2	82.4	97.1	97.1	100
56	85.3	91.2	88.2	97.1	97.1	97.1
57	88.2	88.2	88.2	97.1	97.1	94.1
58	82.4	88.2	85.3	97.1	100	100
59	85.3	88.2	82.4	100	97.1	100
60	85.3	85.3	82.4	94.1	94.1	94.1
61	76.5	91.2	82.4	94.1	94.1	94.1
62	91.2	85.3	88.2	97.1	94.1	94.1
63	79.4	88.2	91.2	88.2	88.2	97.1
Overall Accuracy	84.5	87.9	85.9	94.1	93.1	94.9

Bibliography

[1] B. Merz. "Six common depression types". In: *Harvard Health Publishing* (2020). Retrieved October 15, 2021, from Harvard Health Publishing website: https://www.health.harvard.edu/mind-and-mood/six-common-depression-types.

[2] C. W. M. Ng, C. H. How, and Y. P. Ng. "Major depression in primary care: making the diagnosis". In: *Singapore Medical Journal* 57(11) (2016), pp. 591–597. DOI: https://doi.org/10.11622/smedj.2016174. URL: http://www.smj.org.sg/article/major-depression-primary-care-making-diagnosis.

[3] "Depression". In: *National Institute of Mental Health* (2018). Retrieved October 15, 2021, from National Institute of Mental Health website: https://www.nimh.nih.gov/health/topics/depression,

[4] "What causes depression?" In: *Harvard Health Publishing* (2019). Retrieved October 15, 2021, from Harvard Health Publishing website: https://www.health.harvard.edu/mind-and-mood/what-causes-depression.

[5] "Mental disorders". In: *World Health Organization* (2019). Retrieved October 15, 2021, from World Health Organization website: https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

[6] L. S. Goldman, N. H. Nielsen, H. C. Champion, and A. M. A. the Council on Scientific Affairs. "Awareness, diagnosis, and treatment of depression". In: *Journal of General Internal Medicine* 14.9 (1999), pp. 569–580. ISSN: 1525-1497. DOI: 10.1046/j.1525-1497.1999.03478.x. URL: https://doi.org/10.1046/j.1525-1497.1999.03478.x.

[7] "Quantifying the Cost of Depression". In: *American Psychiatric Association* (2021). Retrieved October 15, 2021, from Center for Workplace Mental Health website: https://www.workplacementalhealth.org/MentalHealthTopics/Depression/Quantifying-the-Cost-of-Depression.

[8] H. Cai, J. Han, Y. Chen, X. Sha, et al. "A Pervasive Approach to EEG-Based Depression Detection". In: *Complexity* (2018), pp. 1–13. DOI: 10.1155/2018/5238028. URL: https://doi.org/10.1155/2018/5238028.

[9] K. M. Smith, P. F. Renshaw, and J. Bilello. "The diagnosis of depression: current and emerging methods". In: *Comprehensive Psychiatry* 54.1 (2013), pp. 1–6. ISSN: 0010-440X. DOI: https://doi.org/10.1016/j.comppsych.2012.06.006. URL: https://www.sciencedirect.com/science/article/pii/S0010440X12001186.

[10] C. Zhuo, G. Li, X. Lin, D. Jiang, Y. Xu, H. Tian, W. Wang, and X. Song. "The rise and fall of MRI studies in major depressive disorder". In: *Translational Psychiatry* 9.1 (2019), p. 335. ISSN: 2158-3188. DOI: 10.1038/s41398-019-0680-6. URL: https://doi.org/10.1038/s41398-019-0680-6.

[11] D. Bandara and K. Kiguchi. "Brain signal acquisition methods in BCIs to estimate human motion intention – a survey". In: *2018 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*. 2018, pp. 1–7. DOI: 10.1109/MHS.2018.8887072.

[12] D. P. Subha, P. K. Joseph, R. Acharya U, and C. M. Lim. "EEG Signal Analysis: A Survey". In: *Journal of Medical Systems* 34.2 (2010), pp. 195–212. ISSN: 1573-689X. DOI: 10.1007/s10916-008-9231-z. URL: https://doi.org/10.1007/s10916-008-9231-z.

[13] S. Vaid, P. Singh, and C. Kaur. "EEG Signal Analysis for BCI Interface: A Review". In: *2015 Fifth International Conference on Advanced Computing Communication Technologies*. 2015, pp. 143–147. DOI: 10.1109/ACCT.2015.72.

[14] U. Herwig, P. Satrapi, and C. Schönfeldt-Lecuona. "Using the International 10-20 EEG System for Positioning of Transcranial Magnetic Stimulation". In: *Brain Topography* 16.2 (2003), pp. 95–99. ISSN: 1573-6792. DOI: 10.1023/B:BRAT.0000006333.93597.9d. URL: https://doi.org/10.1023/B:BRAT.0000006333.93597.9d.

[15] J. N. Acharya, A. Hani, J. Cheek, P. Thirumala, and T. N. Tsuchida. "American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature". In: *Journal of Clinical Neurophysiology* 33.4 (2016), pp. 308–311. ISSN: 0736-0258. URL: https://journals.lww.com/clinicalneurophys/Fulltext/2016/08000/American_Clinical_Neurophysiology_Society.4.aspx.

[16] N. Bajaj. "Wavelets for EEG Analysis". In: *Wavelet Theory* (2021). DOI: 10.5772/intechopen.94398.

[17] J. Moini and P. Piran. "Chapter 6 - Cerebral cortex". In: *Functional and Clinical Neuroanatomy*. Ed. by J. Moini and P. Piran. Academic Press, 2020, pp. 177–240. ISBN: 978-0-12-817424-1. DOI: https://doi.org/10.1016/B978-0-12-817424-1.00006-9. URL: https://www.sciencedirect.com/science/article/pii/B9780128174241000069.

[18] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra. "Chapter 3 - Technical Aspects of Brain Rhythms and Speech Parameters". In: *Introduction to EEG- and Speech-Based Emotion Recognition*. Ed. by P. A. Abhang, B. W. Gawali, and S. C. Mehrotra. Academic Press, 2016,

pp. 51–79. ISBN: 978-0-12-804490-2. DOI: https://doi.org/10.1016/B978-0-12-804490-2.00003-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128044902000038.

[19] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu. "A review on transfer learning in EEG signal analysis". In: *Neurocomputing* 421 (2021), pp. 1–14. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2020.09.017. URL: https://www.sciencedirect.com/science/article/pii/S0925231220314223.

[20] H. U. Amin, A. S. Malik, R. F. Ahmad, N. Badruddin, N. Kamel, M. Hussain, and W.-T. Chooi. "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques". In: *Australasian Physical & Engineering Sciences in Medicine* 38.1 (2015), pp. 139–149. ISSN: 1879-5447. DOI: 10.1007/s13246-015-0333-x. URL: https://doi.org/10.1007/s13246-015-0333-x.

[21] J. Latif, C. Xiao, A. Imran, and S. Tu. "Medical Imaging using Machine Learning and Deep Learning Algorithms: A Review". In: *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. 2019, pp. 1–5. DOI: 10.1109/ICOMET.2019.8673502.

[22] P. M. Rodrigues, B. C. Bispo, C. Garrett, D. Alves, J. P. Teixeira, and D. Freitas. "Lacsogram: A New EEG Tool to Diagnose Alzheimer's Disease". In: *IEEE Journal of Biomedical and Health Informatics* 25.9 (2021), pp. 3384–3395. DOI: 10.1109/JBHI.2021.3069789.

[23] E. Trucco, A. McNeil, S. McGrory, L. Ballerini, M. R. K. Mookiah, S. Hogg, A. Doney, and T. MacGillivray. "Chapter 9 - Validation". In: *Computational Retinal Image Analysis*. Ed. by E. Trucco, T. MacGillivray, and Y. Xu. The Elsevier and MICCAI Society Book Series. Academic Press, 2019, pp. 157–170. ISBN: 978-0-08-102816-2. DOI: https://doi.org/10.1016/B978-0-08-102816-2.00009-5.

[24] S. Grimnes and Ørjan G Martinsen. "Chapter 9 - Data and Models". In: *Bioimpedance and Bioelectricity Basics (Third Edition)*. Ed. by S. Grimnes and Ørjan G Martinsen. Third Edition. Oxford: Academic Press, 2015, pp. 329–404. ISBN: 978-0-12-411470-8. DOI: https://doi.org/10.1016/B978-0-12-411470-8.00009-X.

[25] L. Rokach and O. Maimon. "Decision Trees". In: *Data Mining and Knowledge Discovery Handbook*. Ed. by O. Maimon and L. Rokach. Boston, MA: Springer US, 2005, pp. 165–192. ISBN: 978-0-387-25465-4. DOI: 10.1007/0-387-25465-X_9.

[26] J. Hallinan. "Chapter 2 - Data mining for microbiologists". In: *Systems Biology of Bacteria*. Ed. by C. Harwood and A. Wipat. Vol. 39. Methods in Microbiology. Academic Press, 2012, pp. 27–79. DOI: https://doi.org/10.1016/B978-0-08-099387-4.00002-8.

[27] A. Subasi. "Chapter 3 - Machine learning techniques". In: *Practical Machine Learning for Data Analysis Using Python*. Ed. by A. Subasi. Academic Press, 2020, pp. 91–202. ISBN: 978-0-12-821379-7. DOI: https://doi.org/10.1016/B978-0-12-821379-7.00003-5.

[28] K. Roy, S. Kar, and R. N. Das. "Chapter 6 - Selected Statistical Methods in QSAR". In: *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Ed. by K. Roy, S. Kar, and R. N. Das. Boston: Academic Press, 2015, pp. 191–229. ISBN: 978-0-12-801505-6. DOI: https://doi.org/10.1016/B978-0-12-801505-6.00006-5.

[29] K. Yeturu. "Chapter 3 - Machine learning algorithms, applications, and practices in data science". In: *Principles and Methods for Data Science*. Ed. by A. S. Srinivasa Rao and C. Rao. Vol. 43. Handbook of Statistics. Elsevier, 2020, pp. 81–206. DOI: https://doi.org/10.1016/bs.host.2020.01.002.

[30] V. Gudivada, M. Irfan, E. Fathi, and D. Rao. "Chapter 5 - Cognitive Analytics: Going Beyond Big Data Analytics and Machine Learning". In: *Cognitive Computing: Theory and Applications*. Ed. by V. N. Gudivada, V. V. Raghavan, V. Govindaraju, and C. Rao. Vol. 35. Handbook of Statistics. Elsevier, 2016, pp. 169–205. DOI: https://doi.org/10.1016/bs.host.2016.07.010.

[31] R. Sinnott, H. Duan, and Y. Sun. "Chapter 15 - A Case Study in Big Data Analytics: Exploring Twitter Sentiment Analysis and the Weather". In: *Big Data*. Ed. by R. Buyya, R. N. Calheiros, and A. V. Dastjerdi. Morgan Kaufmann, 2016, pp. 357–388. ISBN: 978-0-12-805394-2. DOI: https://doi.org/10.1016/B978-0-12-805394-2.00015-5.

[32] J. Dukart. "Basic Concepts of Image Classification Algorithms Applied to Study Neurodegenerative Diseases". In: *Brain Mapping*. Ed. by A. W. Toga. Waltham: Academic Press, 2015, pp. 641–646. ISBN: 978-0-12-397316-0. DOI: https://doi.org/10.1016/B978-0-12-397025-1.00072-5.

[33] D. A. Pisner and D. M. Schnyer. "Chapter 6 - Support vector machine". In: *Machine Learning*. Ed. by A. Mechelli and S. Vieira. Academic Press, 2020, pp. 101–121. ISBN: 978-0-12-815739-8. DOI: https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

[34] A. Subasi, K. Khateeb, T. Brahimi, and A. Sarirete. "Chapter 5 - Human activity recognition using machine learning methods in a smart healthcare environment". In: *Innovation in Health Informatics*. Ed. by M. D. Lytras and A. Sarirete. Next Gen Tech Driven Personalized MedSmart Healthcare. Academic Press, 2020, pp. 123–144. ISBN: 978-0-12-819043-2. DOI: https://doi.org/10.1016/B978-0-12-819043-2.00005-8.

[35] R. Neath and M. Johnson. "Discrimination and Classification". In: *International Encyclopedia of Education (Third Edition)*. Ed. by P. Peterson, E. Baker, and B. McGaw. Third Edition. Oxford: Elsevier, 2010, pp. 135–141. ISBN: 978-0-08-044894-7. DOI: https://doi.org/10.1016/B978-0-08-044894-7.01312-9.

[36] Q. Chaudhry, J. Chrétien, M. Craciun, G. Guo, F. Lemke, J.-A. Müller, D. Neagu, N. Piclin, M. Pintore, and P. Trundle. "Chapter 4 - Algorithms for (Q)SAR model building". In: *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*. Ed. by E. BENFENATI. Amsterdam: Elsevier, 2007, pp. 111–147. ISBN: 978-0-444-52710-3. DOI: https://doi.org/10.1016/B978-044452710-3/50006-9.

[37] S. Misra and H. Li. "Chapter 9 - Noninvasive fracture characterization based on the classification of sonic wave travel times". In: *Machine Learning for Subsurface Characterization*. Ed. by S. Misra, H. Li, and J. He. Gulf Professional Publishing, 2020, pp. 243–287. ISBN: 978-0-12-817736-5. DOI: https://doi.org/10.1016/B978-0-12-817736-5.00009-0.

[38] D. Talia, P. Trunfio, and F. Marozzo. "Chapter 1 - Introduction to Data Mining". In: *Data Analysis in the Cloud*. Ed. by D. Talia, P. Trunfio, and F. Marozzo. Computer Science Reviews and Trends. Boston: Elsevier, 2016, pp. 1–25. ISBN: 978-0-12-802881-0. DOI: https://doi.org/10.1016/B978-0-12-802881-0.00001-9.

[39] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman. "1D convolutional neural networks and applications: A survey". In: *Mechanical Systems and Signal Processing* 151 (2021), p. 107398. ISSN: 0888-3270. DOI: https://doi.org/10.1016/j.ymssp.2020.107398.

[40] D. Li, J. Zhang, Q. Zhang, and X. Wei. "Classification of ECG signals based on 1D convolution neural network". In: *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. 2017, pp. 1–6. DOI: 10.1109/HealthCom.2017.8210784.

[41] A. Tavares, E. Di Lorenzo, B. Peeters, G. Coppotelli, and N. Silvestre. "Damage Detection in Lightweight Structures Using Artificial Intelligence Techniques". In: *Experimental Techniques* 45.3 (2021), pp. 389–410. ISSN: 1747-1567. DOI: 10.1007/s40799-020-00421-5.

[42] S. Abdoli, P. Cardinal, and A. Lameiras Koerich. "End-to-end environmental sound classification using a 1D convolutional neural network". In: *Expert Systems with Applications* 136 (2019), pp. 252–263. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2019.06.040.

[43] J. F. Cavanagh, A. Napolitano, C. Wu, and A. Mueen. "The Patient Repository for EEG Data + Computational Tools (PRED+CT)". In: *Frontiers in Neuroinformatics* 11 (2017). ISSN: 1662-5196. DOI: 10.3389/fninf.2017.00067.

[44] M. Hosseinzadeh. "4 - Robust control applications in biomedical engineering: Control of depth of hypnosis". In: *Control Applications for Biomedical Engineering Systems*. Ed. by A. T. Azar. Academic Press, 2020, pp. 89–125. ISBN: 978-0-12-817461-6. DOI: https://doi.org/10.1016/B978-0-12-817461-6.00004-4.

[45] R. J. Martis, U. R. Acharya, and L. C. Min. "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform". In: *Biomedical Signal Processing and Control* 8.5 (2013), pp. 437–448. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2013.01.005.

[46] D. Chen, S. Wan, J. Xiang, and F. S. Bao. "A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG". In: *PLOS ONE* 12.3 (Mar. 2017), pp. 1–21. DOI: 10.1371/journal.pone.0173138.

[47] S. Abbaspour, M. Lindén, H. Gholamhosseini, A. Naber, and M. Ortiz-Catalan. "Evaluation of surface EMG-based recognition algorithms for decoding hand movements". In: *Medical & Biological Engineering & Computing* 58.1 (2020), pp. 83–100. ISSN: 1741-0444. DOI: 10.1007/s11517-019-02073-z.

[48] S. W. Smith. *The scientist and engineer's Guide to Digital Signal Processing*. California Technical Pub., 1997.

[49] T. Cecchin, R. Ranta, L. Koessler, O. Caspary, H. Vespignani, and L. Maillard. "Seizure lateralization in scalp EEG using Hjorth parameters". In: *Clinical Neurophysiology* 121.3 (2010), pp. 290–300. ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2009.10.033.

[50] M. Mouzé-Amady and F. Horwat. "Evaluation of Hjorth parameters in forearm surface EMG analysis during an occupational repetitive task". In: *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control* 101.2 (1996), pp. 181–183. ISSN: 0924-980X. DOI: https://doi.org/10.1016/0924-980X(96)00316-5.

[51] C. Grover and N. Turk. "Rolling Element Bearing Fault Diagnosis using Empirical Mode Decomposition and Hjorth Parameters". In: *Procedia Computer Science* 167 (2020). International Conference on Computational Intelligence and Data Science, pp. 1484–1494. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.03.359.

[52] S. Akdemir Akar, S. Kara, S. Agambayev, and V. Bilgiç. "Nonlinear analysis of EEGs of patients with major depression during different emotional states". In: *Computers in Biology and Medicine* 67 (2015), pp. 49–60. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2015.09.019.

[53] J. Jeong. "Nonlinear dynamics of EEG in Alzheimers disease". In: *Drug Development Research* 56.2 (2002), 57–66. DOI: 10.1002/ddr.10061.

[54] B. Hosseinifard, M. H. Moradi, and R. Rostami. "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal". In: *Computer Methods and Programs in Biomedicine* 109.3 (2013), pp. 339–345. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2012.10.008.

[55] C. G. Pernalete, J. van Baten, J. C. Urbina, and J. F. Arévalo. "A molecular reconstruction feed characterization and CAPE OPEN implementation strategy to develop a tool for modeling HDT reactors for light petroleum cuts". In: *12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering*. Ed. by

K. V. Gernaey, J. K. Huusom, and R. Gani. Vol. 37. Computer Aided Chemical Engineering. Elsevier, 2015, pp. 359–364. DOI: https://doi.org/10.1016/B978-0-444-63578-5.50055-4.

[56] B. Bein. "Entropy". In: *Best Practice Research Clinical Anaesthesiology* 20.1 (2006). Monitoring Consciousness, pp. 101–109. ISSN: 1521-6896. DOI: https://doi.org/10.1016/j.bpa.2005.07.009.

[57] Q. Yuan, W. Zhou, S. Li, and D. Cai. "Epileptic EEG classification based on extreme learning machine and nonlinear features". In: *Epilepsy Research* 96.1 (2011), pp. 29–38. ISSN: 0920-1211. DOI: https://doi.org/10.1016/j.eplepsyres.2011.04.013.

[58] D. Abásolo, R. Hornero, P. Espino, J. Poza, C. I. Sánchez, and R. de la Rosa. "Analysis of regularity in the EEG background activity of Alzheimer's disease patients with Approximate Entropy". In: *Clinical Neurophysiology* 116.8 (2005), pp. 1826–1834. ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2005.04.001.

[59] K. Polat and S. Güneş. "A new feature selection method on classification of medical datasets: Kernel F-score feature selection". In: *Expert Systems with Applications* 36.7 (2009), pp. 10367–10373. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2009.01.041.

[60] S. Güneş, K. Polat, and Şebnem Yosunkaya. "Multi-class f-score feature selection approach to classification of obstructive sleep apnea syndrome". In: *Expert Systems with Applications* 37.2 (2010), pp. 998–1004. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2009.05.075.

[61] P. J. Fitzgerald and B. O. Watson. "Gamma oscillations as a biomarker for major depression: an emerging topic". In: *Translational Psychiatry* 8.1 (2018), p. 177. ISSN: 2158-3188. DOI: 10.1038/s41398-018-0239-y.

[62] J. J. Newson and T. C. Thiagarajan. "EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies". In: *Frontiers in Human Neuroscience* 12 (2019). ISSN: 1662-5161. DOI: 10.3389/fnhum.2018.00521.