DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

A NOVEL HYBRID FEATURE EXTRACTION METHOD FOR CLASSIFICATION OF CHEMOSENSORY EEG SIGNALS

by
Begüm KARA GÜLAY

February, 2023
İZMİR

A NOVEL HYBRID FEATURE EXTRACTION METHOD FOR CLASSIFICATION OF CHEMOSENSORY EEG SIGNALS

A Thesis Submitted to the
Graduate School of Natural And Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics

by
Begüm KARA GÜLAY

February, 2023
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**A NOVEL HYBRID FEATURE EXTRACTION METHOD FOR CLASSIFICATION OF CHEMOSENSORY EEG SIGNALS**" completed by **BEGÜM KARA GÜLAY** under supervision of **PROF. DR. NESLIHAN DEMİREL** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

.....

Prof. Dr. Neslihan DEMİREL

Supervisor

.....	.....
Thesis Committee Member	Thesis Committee Member
.....	.....
Examining Committee Member	Examining Committee Member

Prof. Dr. Okan FISTIKOĞLU
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

First of all, I am very grateful to my supervisor, Prof. Dr. Neslihan DEMİREL, for her valuable advice, guidance, continuous support, and patience during my academic life and Ph.D. study. Her personality, positivity, and openness to innovation have always inspired me.

I want to thank Prof. Dr. Aylin ALIN and Assoc. Prof. Dr. Çağdaş GÜDÜCÜ, who are on the thesis meeting committee and shared their knowledge and suggestions about my work and guided me with their contributions.

I would also like to thank Dr. Alper VAHAPLAR for his unwavering support and patience throughout my work.

I also take this opportunity to express gratitude to all of the Department of Statistics members for their help and support.

I am so grateful to my dear family, my husband, and my sweet little girl for their invaluable support. Especially my spouse and colleague, Dr. Emrah GÜLAY, who motivated me and gave me strong support. I would like to thank them for their patience and encouragement. Without their love and support, I would not have been able to complete this thesis.

Begüm KARA GÜLAY

A NOVEL HYBRID FEATURE EXTRACTION METHOD FOR CLASSIFICATION OF CHEMOSENSORY EEG SIGNALS

ABSTRACT

The main goal of this study is to develop a more accurate way of diagnosing Parkinson's Disease in its early stages using chemosensory Electroencephalography (EEG) signals, which are often difficult to study. We propose a hybrid feature extraction method called EEMD_VAR that combines Ensemble Empirical Mode Decomposition (EEMD) and Vector Auto-regressive (VAR) analysis. This method is intended to avoid the arbitrary selection of features and automatically determine the number of features. The pre-processed EEG signals are decomposed using EEMD, and the resulting Intrinsic Mode Functions (IMFs) are used as independent variables in the VAR model. The coefficients of the VAR model are then used as features in various supervised classification algorithms. The performance of the EEMD_VAR method is compared to that of the Auto-Regressive (AR) model and Hjorth parameters. The proposed method is found to achieve a maximum classification accuracy of 100% using Artificial Neural Networks (ANN) in the C2 electrode, whereas the AR method and Hjorth parameters only achieve maximum accuracy of 72%. The other performance metrics also support the effectiveness of the proposed method. Furthermore, the higher results from electrodes on the right side of the brain suggest that the right side of the brain may be more sensitive to olfactory stimuli.

Keywords: Classification, feature extraction, electroencephalogram, Parkinson's disease, chemosensory.

KEMODUYSAL EEG SİNYALLERİNİN SINIFLANDIRILMASI İÇİN YENİ BİR HİBRİT ÖZNİTELİK ÇIKARMA YÖNTEMİ

ÖZ

Bu çalışmanın temel amacı, genellikle incelenmesi zor olan kemosensör elektroensefalografi (EEG) sinyallerini kullanarak Parkinson hastalığını erken evrelerinde teşhis etmenin daha doğru bir yolunu geliştirmektir. Bu amaçla bu çalışmada Topluluk Ampirik Mod Ayrıştırma ve Vektör Otoregresif Modeli birleştiren EEMD_VAR adlı hibrit bir özellik çıkarma yöntemi önerilmektedir. Bu yöntem, özniteliklerin rastgele seçilmesini önlemek ve öznitelik sayısını otomatik olarak belirlemek için tasarlanmıştır. Önceden işlenmiş EEG sinyalleri, Topluluk Ampirik Mod Ayrıştırma kullanılarak ayrıştırılır ve ortaya çıkan İçsel Mod Fonksiyonları, Vektör Otoregresif modelinde bağımsız değişkenler olarak kullanılır. VAR modelinin katsayıları daha sonra çeşitli denetimli sınıflandırma algoritmalarında özellik olarak kullanılır. EEMD_VAR yönteminin performansı, Otoregresif modeli ve Hjorth parametreleri ile karşılaştırılır. Önerilen yöntemin, C2 elektrotunda yapay sinir ağları kullanarak maksimum yüzde 100'lük bir sınıflandırma doğruluğu elde ettiği, Otoregresif yöntemi ve Hjorth parametrelerinin ise yalnızca yüzde 72'lik bir maksimum doğruluk elde ettiği bulunmuştur. Diğer performans ölçütleri de önerilen yöntemin etkinliğini desteklemektedir. Ayrıca, beynin sağ tarafındaki elektrotlardan alınan daha yüksek sonuçlar, beynin sağ tarafının koku alma uyaranlarına karşı daha duyarlı olabileceği görülmektedir.

Anahtar kelimeler: Sınıflandırma, öznitelik çıkarımı, elektroensefalografi, Parkinson hastalığı, kemosensör.

CONTENTS

Page

Ph.D. THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT iv
ÖZ v
LIST OF FIGURES viii
LIST OF TABLES x
LIST OF ABBREVIATIONS xi
LIST OF SYMBOLS xiii

CHAPTER ONE – INTRODUCTION 1

1.1 Brain-Computer Interface 2
1.2 Literature Overview 2
1.2.1 Electroencephalography 2
1.2.2 Parkinson's Disease 5
1.2.3 Feature Extraction and Classification 5
1.3 Motivation 8
1.4 Contributions 8
1.5 Outline of the Thesis 8

CHAPTER TWO – BACKGROUND 10

2.1 Emprical Mode Decomposition 10
2.2 Ensemble Empirical Mode Decomposition 15
2.3 Vector Autoregressive Model 18
2.4 Autoregressive Model and Hjorth Parameters 19
2.4.1 AR Model 20
2.4.2 Hjorth Parameters 21
2.5 Feature Reduction 22

2.6 Classification 24
2.6.1 K- Nearest Neighbors 25
2.6.2 Support Vector Machines 26
2.6.3 Decision Tree 29
2.6.4 Bagging 30
2.6.5 Extreme Gradient Boosting 31
2.6.6 Logistic Regression 33
2.6.7 Artificial Neural Networks 34
2.7 Performance Evaluation 38

CHAPTER THREE – A NOVEL HYBRID FEATURE EXTRACTION METHOD 43

3.1 Data Acquisition 43
3.2 Pre-Processing 44
3.3 Proposed Method 47

CHAPTER FOUR – RESULTS AND DISCUSSIONS 50

4.1 Findings 53
4.2 Comparative Analysis and Discussions 58

CHAPTER FIVE – CONCLUSIONS AND FUTURE WORKS 64

REFERENCES 67

LIST OF FIGURES

Page

Figure 1.1 An electroencephalogram recording sketch 3
Figure 2.1 The lower and upper envelopes defined as local maximum and local minimum and the mean value of the upper and lower envelopes..... 11
Figure 2.2 Data and h_1 12
Figure 2.3 Repeated sifting steps with h_1 and m_2 13
Figure 2.4 Original data and the residue r_1 14
Figure 2.5 A synthetic signal that illustrates the mode mixing problem 16
Figure 2.6 Flowchart of EEMD algorithm 17
Figure 2.7 Visualization of k-Nearest Neighbors 26
Figure 2.8 Basics of a support vector machine 27
Figure 2.9 Kernel trick in SVM 28
Figure 2.10 The components of a decision tree 29
Figure 2.11 An illustration for the concept of bagging 31
Figure 2.12 An illustration of the concept of boosting 32
Figure 2.13 Sigmoid function 33
Figure 2.14 Logistic regression for classification 34
Figure 2.15 Block diagram of an artificial neural network 36
Figure 2.16 A single layer neural network 37
Figure 2.17 Schematic representation of the leave-one-out cross-validation 39
Figure 2.18 The area under the curve 41
Figure 3.1 The general process of the proposed procedure 43
Figure 3.2 The EEG cap layout for 62 electrodes 46
Figure 3.3 The EEMD_VAR Algorithm 49
Figure 4.1 The EEG signal for a control subject and its IMF components from data through the EEMD method 51
Figure 4.2 The EEG signal for a subject with PD and its IMF components from data through the EEMD method 52
Figure 4.3 Accuracy results for 160 scenarios using LR in C2 53

Figure 4.4 The sensitivity of the results with the highest accuracy for the EEMD_VAR method 56
Figure 4.5 The specificity of the results with the highest accuracy for the EEMD_VAR method 57
Figure 4.6 The area under the curve of the results with the highest accuracy for the EEMD_VAR method 57
Figure 4.7 The sensitivities obtained using the Hjorth parameters and the AR model60
Figure 4.8 The specificity obtained using the Hjorth parameters and the AR model 61
Figure 4.9 The AUC values obtained using the Hjorth parameters and the AR model62

LIST OF TABLES

Page

Table 2.1 Structure of a confusion matrix 39
Table 4.1 The maximum accuracy values obtained using the 160 scenarios for the EEMD_VAR method 55
Table 4.2 The accuracy values obtained using the Hjorth parameters 58
Table 4.3 The accuracy values obtained using the AR model 59

LIST OF ABBREVIATIONS

BCI	: Brain-Computer Interface
CNS	: Central Nervous System
EEG	: Electroencephalography
PD	: Parkinson's Disease
ERP	: Event-Related Potentials
FFT	: Fast Fourier Transform
AR	: Auto-Regressive
NN	: Neural Network
KNN	: K-Nearest Neighbours
EWT	: Empirical Wavelet Transform
IMF	: Intrinsic Mode Functions
EEMD	: Ensemble Empirical Mode Decomposition
VAR	: Vector Auto-regressive
ANN	: Artificial Neural Networks
EMD	: Empirical Mode Decomposition
SVM	: Support Vector Machine
AD	: Alzheimer's Disease
PCA	: Principal Component Analysis
DT	: Decision Tree
XGBoost	: Extreme Gradient Boosting
LR	: Logistic Regression
LOOCV	: Leave-One-Out Cross-Validation
TP	: True Positive
TN	: True Negative
FP	: False Positive
FN	: False Negative
ACC	: Accuracy
SEN	: Sensitivity
SPF	: Specificity

AUC	: Area Under The Curve
TPR	: True Positive Rate
FPR	: False Positive Rate
ROC	: Receiver Operating Characteristic
DEU	: Dokuz Eylul University
ROI	: Region of Interest

LIST OF SYMBOLS

t	: Time
x_t	: The original signal
m	: The mean value of the upper and lower envelopes
h	: The component of EMD
n	: The number of sifting in EMD
r	: The residual of the signal x_t in EMD
k	: The number of IMFs obtained via EMD
IMF_j	: The j^{th} IMF obtained via EMD
N	: The ensemble number of EEMD
x_t^i	: The noisy signal of the i^{th} trial
w_t^i	: The i^{th} added white noise
IMF_j^i	: The j^{th} IMF of the i^{th} trial in EEMD
$\overline{IMF}_j$	: The mean of j^{th} IMFs obtained via EMD over an ensemble of trials
p	: The lag order
$\boldsymbol{Y_t}$	: The vector of endogenous variables
$\boldsymbol{Y_{t-p}}$	: The p time period earlier value of $\boldsymbol{Y_t}$
$\boldsymbol{\beta_0}$	: The vector of intercept of the models
$\boldsymbol{\beta_p}$	: The matrices of p lag coefficients
$\boldsymbol{\epsilon_t}$	: The vector of error terms
D	: The delay operator
$x_t^{'}$	: The first derivative of x_t
ms	: Millisecond
kHz	: Kilohertz
Hz	: Hertz
μV	: Microvolt
$\tilde{x}_t$	: The first difference of x_t
φ_{tr}	: The training set
$\tilde{x}_t^{\varphi_{tr}}$	: The first difference of x_t in the training set

$f_t{}^{\varphi_{tr}}$: IMF of $\tilde{x}_t^{\varphi_{tr}}$

n_{f_t} : The number of IMFs of $\tilde{x}_t^{\varphi_{tr}}$

$\tilde{x}_{t_p}^{\varphi_{tr}}$: The p time period earlier value of $\tilde{x}_t^{\varphi_{tr}}$

$f_{t_{n_{f_t}}}^{\varphi_{tr}}$: The ${n_{f_t}}^{th}$ IMF of $\tilde{x}_t^{\varphi_{tr}}$

$\delta_{VAR}{}^{\varphi_{tr}}$: The autoregressive coefficients of VAR model

ϑ_C : The vector of components that are obtained via PCA

y : The labels of subjects(healthy-PD)

$y^{'}$: The predicted label

C : The classifier used in process

CHAPTER ONE
INTRODUCTION

Human body function is frequently associated with electrical and chemical signals. Such signals contain information that is not immediately clear but is hidden in the structure of the signal. Before the signals can be interpreted, this information must be extracted in some way.The signals from the biological systems contain information about the properties of those systems, and this information is beneficial in understanding and diagnosing various medical conditions. Sometimes, it is easy to obtain information from a signal and only manual effort is needed, like looking at the signal on a printout or computer screen. However, the complexity of the signal can be significant, so using specialized techniques in biomedical signal processing is often necessary to extract useful information that is hidden within the signal (Sörnmo & Laguna, 2005).

The brain consists of billions of cells, half of which are neurons and the other half of which help and facilitate neuron activity. These neurons are connected by synapses that are gateways for inhibitory or excitatory activity. Synaptic activity produces subtle electrical impulses called postsynaptic potentials. Each synchronous firing of thousands of neurons produces an electric field strong enough to propagate through tissues, bones, and skull. Finally, it can be measured on the head surface.

This thesis focuses on the processing of electrical signals representing brain activity. Some signals are produced naturally and reflect ongoing activity, while others are only produced in response to external stimuli. Different methods of processing are needed for the characteristics of these signals. A single waveform can directly indicate a specific medical diagnosis in some signals, but for others, lots of waveform configurations must be analyzed before a meaningful interpretation can be made (Sörnmo & Laguna, 2005).

1.1 Brain-Computer Interface

A BCI is a system that allows the brain to directly communicate with a computer. It measures the activity of the central nervous system (CNS) and converts that activity into signals that can be used to control a device.BCI systems create a connection between the human brain and the external environment, allowing people to communicate without using traditional methods. They can detect and decode thoughts, enabling people with disabilities to express themselves in various ways, such as through writing or silent speech. BCI systems only require brain signals to perform a set of commands and do not require muscle movement. They can also be used to detect diseases such as abnormal brain structure, seizure disorders, sleep disturbances, and neurodegenerative diseases. At present, BCI applications use techniques like electroencephalography (EEG), functional magnetic resonance imaging, and electrocorticography. Among these techniques, EEG is generally preferred for BCI systems because it is easy to obtain and apply in practice. Analyzing and classifying EEG signals is an active area of research in the field of BCI.

1.2 Literature Overview

With the increasing use of EEG in the field of BCI, it has played an active role in the diagnosis of many neurodegenerative diseases. In this section, we provide a literature overview related to EEG studies, Parkinson's Disease (PD) disease and diagnosis, and information extraction studies from EEG signals.

1.2.1 Electroencephalography

EEG is a method of recording electrical activity from the scalp. It can be used to monitor changes in brain activity and is often used to evaluate the neurological condition of a patient. It is portable, non-invasive, relatively inexpensive, and

provides signals with a high temporal resolution (Sergeeva et al., 2022).

Richard Caton (1842-1926), a medical lecturer at Liverpool, was considered the pioneer in the field of evoked potentials. He observed that ”feeble currents of varying direction pass through the multiplier when the electrodes are placed on two points of the external surface.” This marked the beginning of the field of EEG, although it was not officially invented until 1929 by Hans Berger, a German psychiatrist (Sur & Sinha, 2009).

Many neurological disorders can be identified through brain rhythms, which can be detected even by visually inspecting the EEG signal. EEG is used clinically to identify seizures, monitor the depth of anesthesia, and locate areas of damage due to head injuries, strokes, tumors, and other conditions. There is ongoing research in this area and the most crucial prospect of all this research is based on the recording of EEG signals.

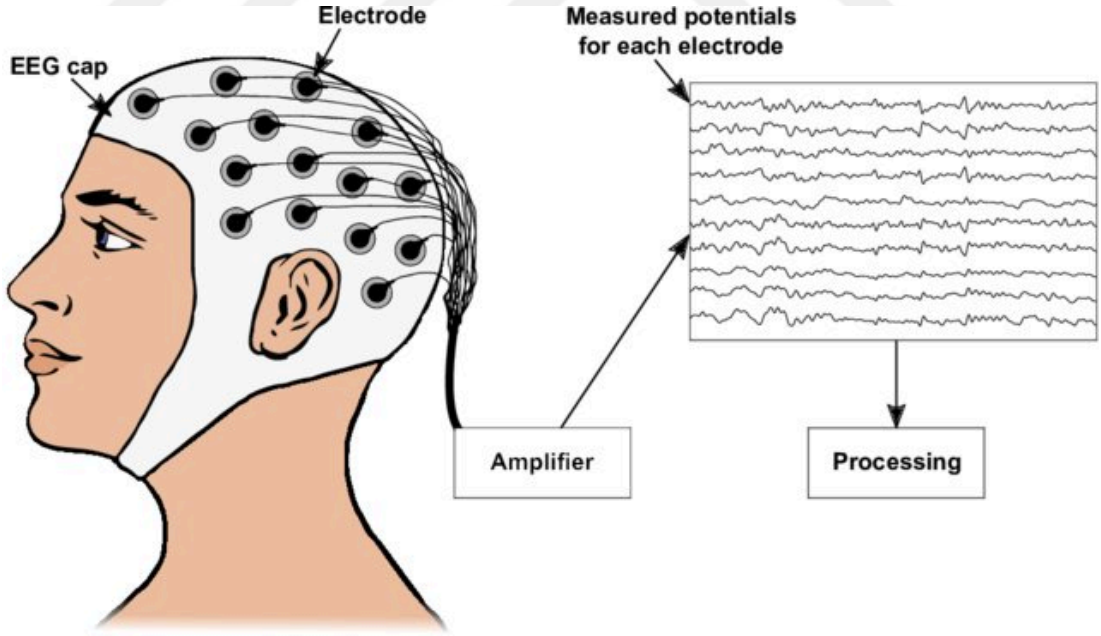


Figure 1.1 An electroencephalogram recording sketch (Nagel, 2019)

The cerebral cortex is a key part of the CNS and is responsible for vital functions like sensation, learning, voluntary movement, speech, and perception. Different regions of the cortex are involved in processing these functions. Most EEG signals originate in

the cerebral cortex. Scientists have developed various methods for recording brain signals, including the 10-10 international standard system for EEG recording. This is an invasive medical procedure that involves placing electrodes on the scalp according to a set of standardized rules and recording the brain's spontaneous electrical activity over a period of time. The electrodes are attached to an elastic cap, allowing data to be collected from the identical scalp position for all subjects. Since the voltage fluctuations measured at the electrodes are insignificant, the recorded data are digitized and sent to an amplifier shown in Figure 1.1. The amplified data can be shown as a series of voltage values. By monitoring the time course of electrical activity in the brain, it is possible to interpret which areas of the cortex are involved in processing information at a given time.

The recording process can introduce errors or artifacts into the EEG data at many points. If these artifacts are present during further processing, it can lead to poor results. EEG signals recorded from electrodes may contain artifacts that are caused by various factors and must be removed. Before starting to use the data, we often need certain pre-processing: cleaning noise, removing artifacts and malfunctioning channels, filtering low and high frequencies, and applying baseline assignment. Pre-processing of signals helps to remove the outliers and makes it easier to perform further signal processing tasks (Khatwani & Tiwari, 2013). Signal processing techniques can be used to address a wide range of issues in EEG signal analysis, including data compression, detection and classification, noise reduction, signal separation, and feature extraction. The analysis of EEG signals is important for both research and medical purposes, including diagnosis and treatment (Gomathi et al., 2013).

Event-related potentials (ERPs) are changes in the electrical activity of the brain in response to specific events. While ongoing EEG activity is being recorded from electrodes placed on the scalp, the participant is presented with a stimulus such as written or spoken words, pictures, or sounds. Each time a stimulus is presented, it evokes a response by the brain, which is continuously recorded on the computer. By analyzing the brain's response to different types of events, we can gain insight into how the brain processes different kinds of information (Beres, 2017).

1.2.2 Parkinson's Disease

PD is a progressive neurological disorder that is characterized by a decrease in dopamine levels, which leads to problems with body function. In addition to these findings in regarding movement disorders, patients may have different complaints. In the literature, there are numerous studies on PD. Some studies demonstrate that a large majority of patients with PD have olfactory dysfunction (Doty, 2012, Doty et al., 1988). Psychophysical and electrophysiological data are evaluated for the sense of smell. Olfactory performance is affected by many parameters such as the humidity and temperature of the environment, olfactory sensitivity, gender, and mood. The effect of these parameters in psychophysical tests is more prominent than in electrophysiological tests. Therefore, electrophysiological evaluation of the sense of smell is more objective. One of the most common electrophysiological tests is EEG, which is a recording of brain activity. EEG is useful in diagnosing and classifying PD (Maria et al., 2018). People with PD often have abnormalities in their EEG signals during olfactory tasks, such as smelling a specific odor. These abnormalities can include changes in power and coherence in different frequency bands, as well as alterations in the timing and amplitude of ERPs. These findings suggest that EEG could be used as a biomarker for olfactory dysfunction in PD, and could help in the early detection and management of this symptom. More research is needed to fully understand the mechanisms behind olfactory dysfunction in PD and to develop more effective treatments.

EEG can be used for clinical evaluation since it is simple to use, non-invasive, and capable of differentiating the type and severity of PD at a lower cost than other neuroimaging techniques (Aljalal et al., 2022).

1.2.3 Feature Extraction and Classification

Feature extraction is a process in machine learning and data analysis where relevant features or characteristics are extracted from raw data. These features are

typically selected for their ability to represent important information about the data and are often used as input for a machine learning algorithm. In the context of EEG, feature extraction is the process of identifying and extracting relevant features from raw EEG data. This can be a challenging task because of the high dimensionality and complexity of EEG signals, which are often contaminated by various artifacts. These features can be used as input to machine learning algorithms for tasks such as BCI, sleep stage classification, and seizure detection. There is ongoing research into feature extraction techniques for EEG, with the goal of developing new methods that can improve the performance of machine learning models on EEG data.

When EEG records are used for a long time, substantial data are collected. To extract information from this extensive data, automated processes that analyze and classify the data are required. Visual assessment of EEG signals is not a satisfying method because there are no standard criteria for evaluation, and it is a time-consuming process that frequently results in errors. Recently, several methods for extracting features from EEG recordings and classifying them using machine learning algorithms have been proposed. Various researchers have examined specifically feature extraction and machine learning methods that have been applied to EEG analysis (Aljalal et al., 2022, Aydemir, 2017, Albaqami et al., 2021, Boonyakitanont et al., 2020, Rahman & Fattah, 2020). For instance, conventional feature extraction methods like Fast Fourier Transform (FFT) and Auto-Regressive (AR) methods have been reviewed and compared (Al-Fahoum & Al-Fraihat, 2014). Uyulan & Erguzel (2017) were applied feature extraction procedures, and the classification of the patterns based on the time-frequency spectrum features was carried out using a neural network (NN). A method to automatically detect PD using vowels has been proposed by Tuncer et al. (2020) and the features extracted by single value decomposition have been classified with K-Nearest Neighbours (KNN). Sadiq et al. (2020) implemented multiscale principal component analysis for noise removal and presented a methodology for the classification of motor-imagery EEG signals by two-dimensional modeling of modes acquired with the Empirical Wavelet Transform (EWT) and extraction of a single feature only. A data-adaptive EWT-based signal

decomposition method was proposed by Sadiq et al. (2019) for feature extraction to improve the classification accuracy. A multileveled feature generator was proposed and a highly accurate EEG classification approach is presented (Tuncer et al., 2021). A robust and simple automated multivariate EWT algorithm was proposed for the key challenges in motor imagery based on EEG (Sadiq et al., 2019). In another work (Leoni et al., 2021), ERP was classified as an automatic, easily interpretable classification model capable of assisting the neuroscientist in the recognition of visual and auditory stimuli. Uktveris & Jusas (2017) evaluated different feature extraction methods and found that the FFT energy map behaves best in the four-class classification of motor imagery tasks.

At present, there is no cure for PD, but it is possible to delay the onset of symptoms. Therefore, early diagnosis is critical for people with PD to have a better quality of life. However, the number of studies evaluating the sense of smell to diagnose PD at an early stage or determine whether an individual is at risk of PD is limited and generally not recent. It is seen that the monitoring period of patients with PD is generally the period after the medication is implemented (Guducu, 2015). Most of the studies on smell are related to recognizing olfactory stimuli from the EEG response and separating noisy stimuli (Saha et al., 2014). Additionally, the findings of Leoni et al. (2021) and Iannilli et al. (2017) support the idea that the reduction in olfactory perception in PD is different from the dysfunctions seen in other conditions that cause a decrease in the ability to smell. This olfactory dysfunction is so reliable that it can be used as a marker of PD (Tremblay et al., 2017). De Oliveira et al. (2020) proposed a biomarker of PD using EEG signals, photic stimulation, partial directed coherence for feature extraction, and machine learning techniques to classify participants into three distinct groups: PD patients who were taking medication, PD patients who were not taking medication, and healthy subjects. Another study (Guducu et al., 2015) revealed that electrophysiological measurements suggest that there is a significant change in entropy among time windows for olfactory stimulation in healthy subjects, but not in patients with PD.

1.3 Motivation

Since the loss of smell can be an early sign of PD and there is currently no cure for the disease, this study aims to achieve high accuracy in detecting PD using the classification of chemosensory EEG signals and machine learning techniques. With this motivation, we propose a new approach for extracting features from olfactory EEG signals in the time domain. No one, to the best of our knowledge, has studied olfactory EEG signals for this purpose. The primary goal of this thesis is to avoid arbitrary selection of features and to determine the number of features to use.

1.4 Contributions

In this thesis, a new hybrid approach called EEMD_VAR combines the two methods. First, several Intrinsic Mode Functions (IMFs) of data are decomposed using the Ensemble Empirical Mode Decomposition (EEMD) method. Second, IMFs are added to the Vector Auto-Regressive (VAR) model as regressors to calculate auto-regressive coefficients, which are used as features. Then, the features feed the classification algorithms to classify EEG signals.

The proposed hybrid feature extraction technique (EEMD_VAR) prevents the arbitrary selection of features.

The proposed approach achieved 100% classification accuracy using Artificial Neural Networks (ANN). The performance of EEMD_VAR is verified using nine different channels along with seven classification algorithms. The results prove the success of the proposed method.

1.5 Outline of the Thesis

The rest of this thesis is organized as follows:

In Chapter 2, some background information about the methods used in this thesis. Chapter 2 starts by introducing the mathematical expression of the methods that we used throughout this thesis. Then, some information about the feature reduction is provided. Next, the classification methods are introduced that we used to classify the extracted features. At the end of chapter 2, the performance evaluation methods have been described.

In Chapter 3, the data set is described that we used for evaluating the proposed model. Then, the pre-processing steps in the data analysis process are given. In the last section of this chapter, we proposed a novel hybrid feature extraction method. To construct this combined model, we used EEMD and VAR methods together.

Chapter 4 provides the classification performances with the necessary discussion. We compared the performance of the proposed method with the AR and Hjorth methods.

In Chapter 5, we present a general conclusion based on the results obtained in this thesis and suggest some potential future directions for this research.

CHAPTER TWO
BACKGROUND

This chapter provides background information on important topics related to the content of this thesis. First, we present the details of the mathematical methods used. The next subsection provides details of the methods used in this study. After that, we provide details on the metrics we used to evaluate the performance of the proposed approach.

2.1 Emprical Mode Decomposition

Empirical Mode Decomposition (EMD) is a technique that can be used to detect synchronization in EEG signals. EMD is an iterative process for analyzing nonlinear, non-stationary time series and is particularly well suited for analyzing EEG data (Huang et al., 1998). It decomposes the signal into a finite number of IMFs based on the local, characteristic time scale of the data. The first IMF component obtained through EMD is the component with the highest frequency in the original signal.

Each IMF represents a simple oscillation within the signal. The amplitudes and frequencies of IMFs are not constant and may change over time. According to Huang, IMFs of the same length as the original signal have two important characteristics:

1. The number of extremum and the number of zero crossings must be the same or differ by at most one;
2. The mean value of the envelope defined by the local maximum and the envelope defined by the local minimum must be zero at any point.

EMD is an iterative procedure that extracts IMFs by first identifying the local maximum and minimum and using cubic splines to construct upper and lower envelopes that connect all of the local maximum and minimum, respectively. The

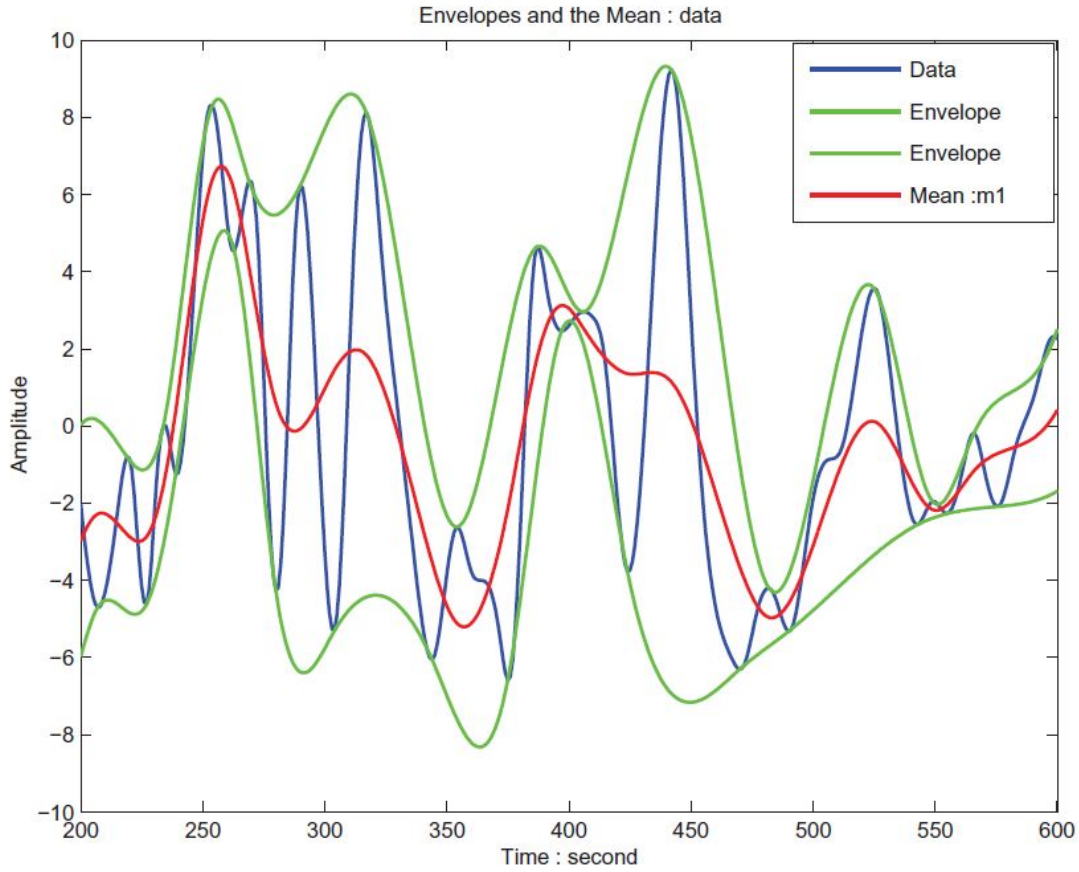


Figure 2.1 The lower and upper envelopes defined as local maximum and local minimum and the mean value of the upper and lower envelopes (Huang, 2014)

mean of these envelopes m_1 is then calculated and subtracted from the original signal x_t, as shown in Figure 2.1. The first component h_1, shown in Figure 2.2, is computed by Equation 2.1.

$$h_1 = x_t - m_1 + 58 \tag{2.1}$$

If h_1 satisfies the two criteria mentioned above, then it is considered an IMF. If it does not meet these criteria, the process of sifting must be applied again until an IMF is obtained. In subsequent iterations of the sifting process, h_1 is treated as the data being analyzed; then,

$$h_{11} = h_1 - m_{11}. \tag{2.2}$$

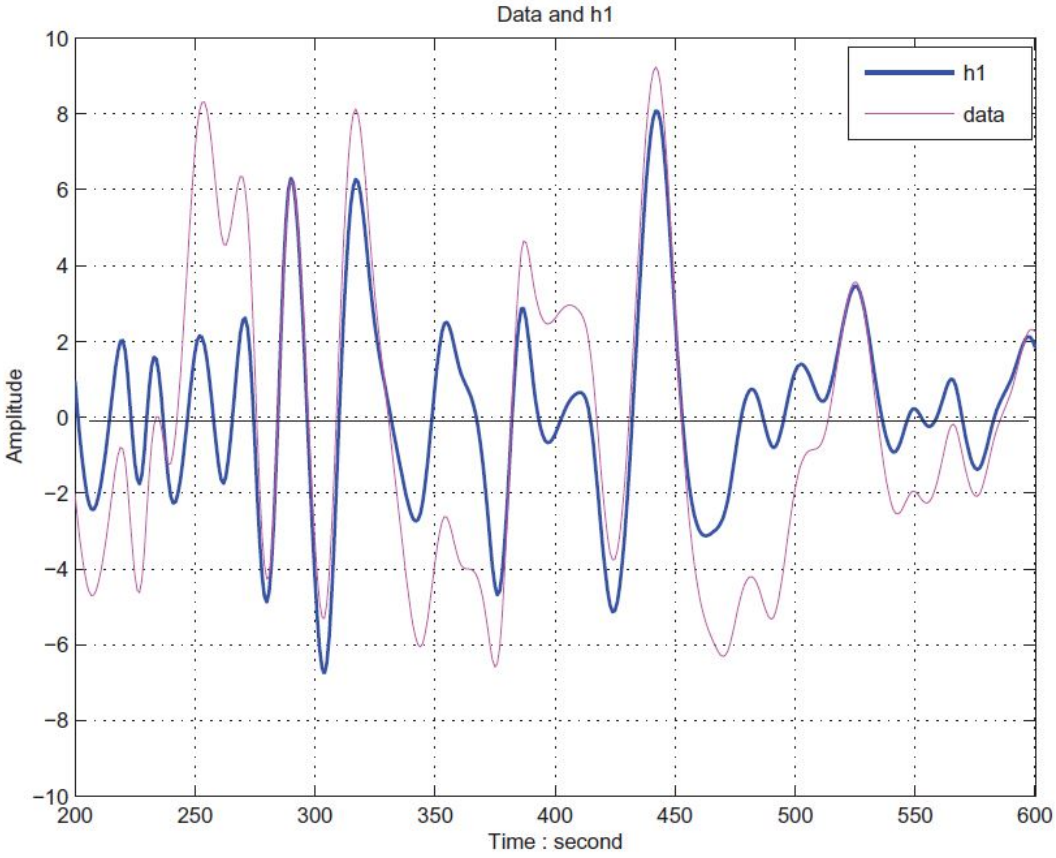


Figure 2.2 Data and h_1 (Huang, 2014)

After repeated sifting in this manner,shown in Figure 2.3, up to n times, h_{1n} becomes an IMF; that is,

$$h_{1n} = h_{1(n-1)} - m_{1n}, \tag{2.3}$$

where h_{1n} represents the function h_1 in Equation 2.1, m_{1n} is the mean of the upper and lower envelopes of h_{1n}. In Equation 2.3, the indices of h and m represent the IMF component and the iteration of the sifting process, respectively. The first IMF component, h_{1n}, contains the finest scale and is equivalent to the shortest period component of the signal, meaning that it is associated with the highest frequency components of the raw signal. Then, h_{1n} is designates as

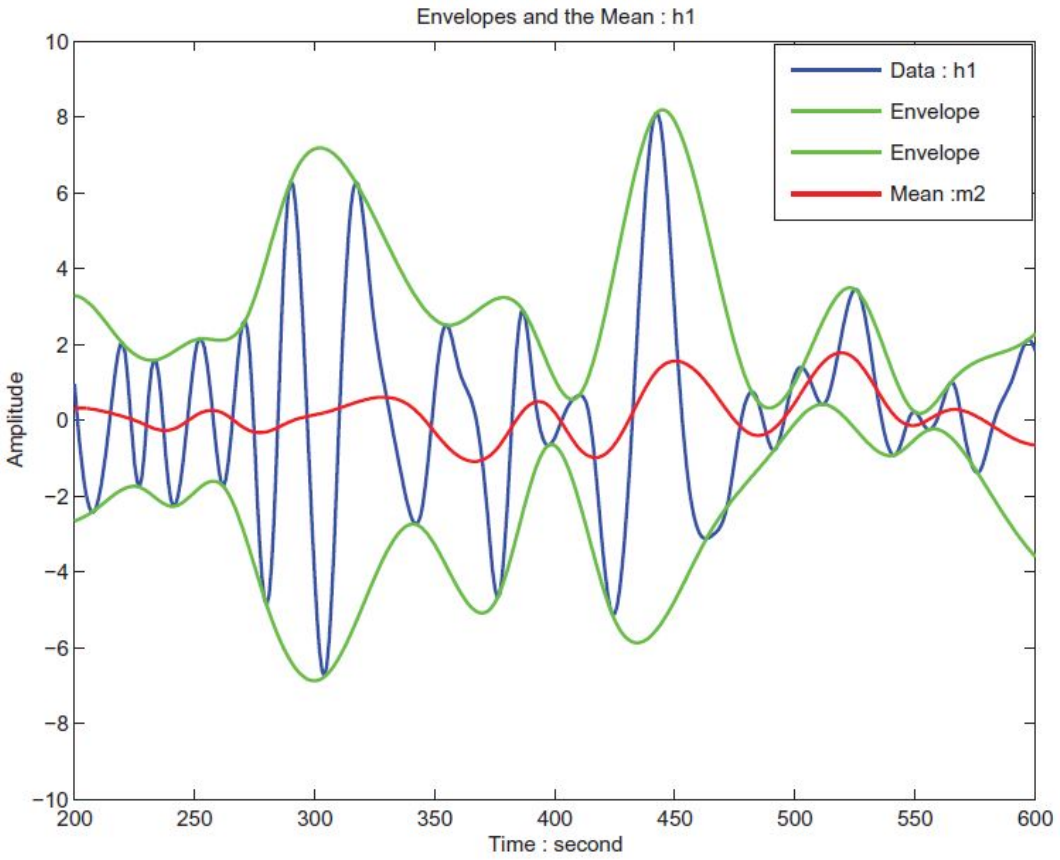


Figure 2.3 Repeated sifting steps with h_1 and m_2(Huang, 2014)

$$IMF_1 \equiv h_{1n} \tag{2.4}$$

the first IMF component from the data.

Here, a stopping criterion must be selected. Two different criteria can be used as the criterion for stopping the sifting process:

1. The first criterion was proposed by Huang et al. (1998). The normalized squared difference (SD_n) between two successful sifting is defined as follows:

$$SD_n = \frac{\sum_{t=0}^{T} |h_{n-1}(t) - h_n(t)|^2}{\sum_{t=0}^{T} h_{n-1}^2(t)} \tag{2.5}$$

If SD_n is smaller than a predetermined value, the sifting process will be stopped.

2. The second criterion is based on the so-called S-number (Huang et al., 1999). The number S is pre-selected. The elimination process stops only when the number of zero-crossings and extreme points and the number of successions S are the same and are equal or differ by at most a single unit. Huang et al. (2003) tested different values of S and developed an empirical guide for determining the optimal range of S values and found that the best range for S is between 4 and 8 for optimal sifting.

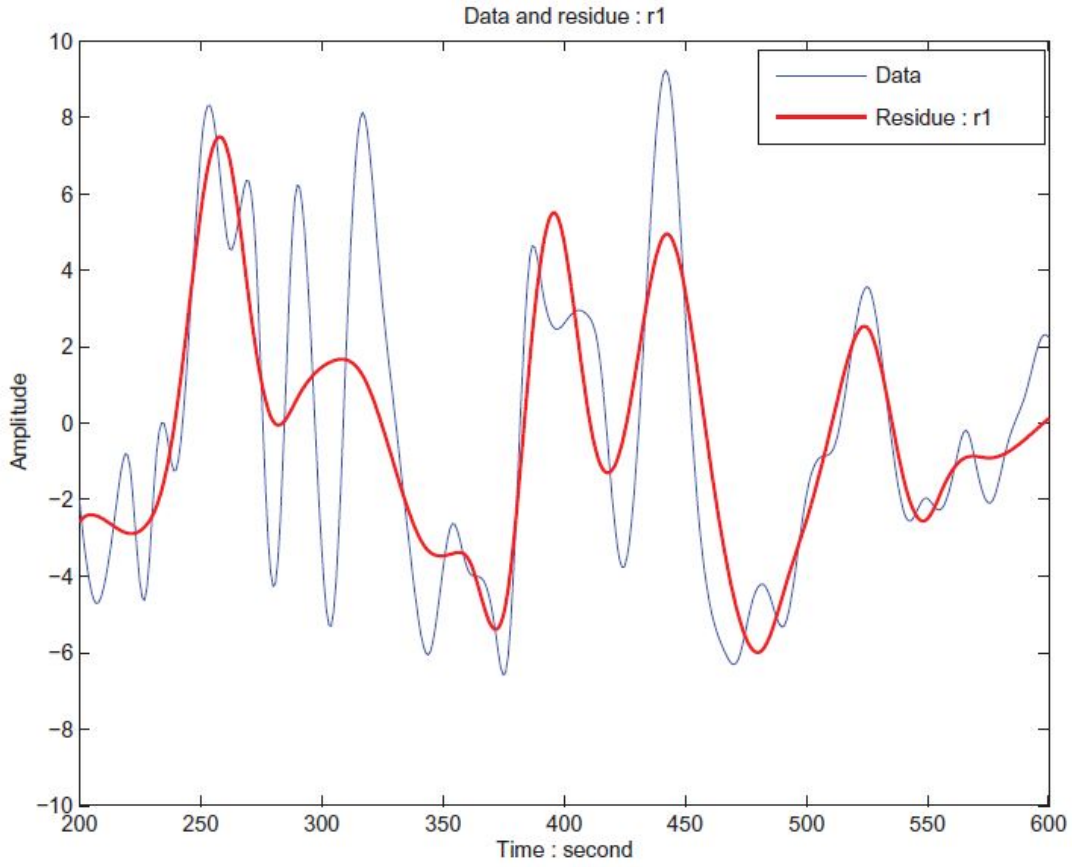


Figure 2.4 Original data and the residue r_1 (Huang, 2014)

Once the stopping criterion has been chosen, IMF_1 is obtained as the first IMF component. IMF_1 contains the finest scale or shortest period component of the signal. By subtracting the first component IMF_1 from the data, its residual signal is denoted by r_1, shown in Figure 2.4.

$$r_1 = x_t - IMF_1 \tag{2.6}$$

Since r_1 still contains longer-term variations in the data, it is treated as new data and undergoes the same sifting process as described above. This procedure can be repeated with all subsequent r_j's and

$$\begin{aligned} r_2 &= r_1 - IMF_2 \\ &\vdots \\ r_k &= r_{k-1} - IMF_k \end{aligned} \tag{2.7}$$

The iteration can be terminated either IMF_k or the residue r_k becomes too small or since r_k is a monotonic function, which implies that no more IMF can be extracted. At the end of the decomposition, x_t can, hence, be represented in terms of k IMFs:

$$x_t = \sum_{j=1}^{k} IMF_j + r_k \tag{2.8}$$

2.2 Ensemble Empirical Mode Decomposition

As discussed in the previous section, EMD decomposes the signal based on extremum information, so changes in extremum and values due to noise can significantly affect the results.

As shown in Figure 2.5, some signals have at least two discernible oscillations mixed in one component because the sifting process starts by identifying the maximum and minimum values in the data. As a result, the first IMF's initial estimate is a mixture of both low-frequency fundamental and high-frequency intermittent waves, as shown in Figure 2.5(d). This is known as mod mixing and is one of the main drawbacks of EMD.

Mode mixing occurs when a single IMF contains waves of widely disparate scales or when a signal of a similar scale is found in different IMF components. This can

happen since the EMD is sensitive to noise in the data. If the noise in the data changes the distribution of extremum in the corresponding noise-free data, or if it contains a different realization of noise, the decompositions of the noisy and noise-free data can be significantly different. This means that the results of the EMD may be sensitive to noise in the data.

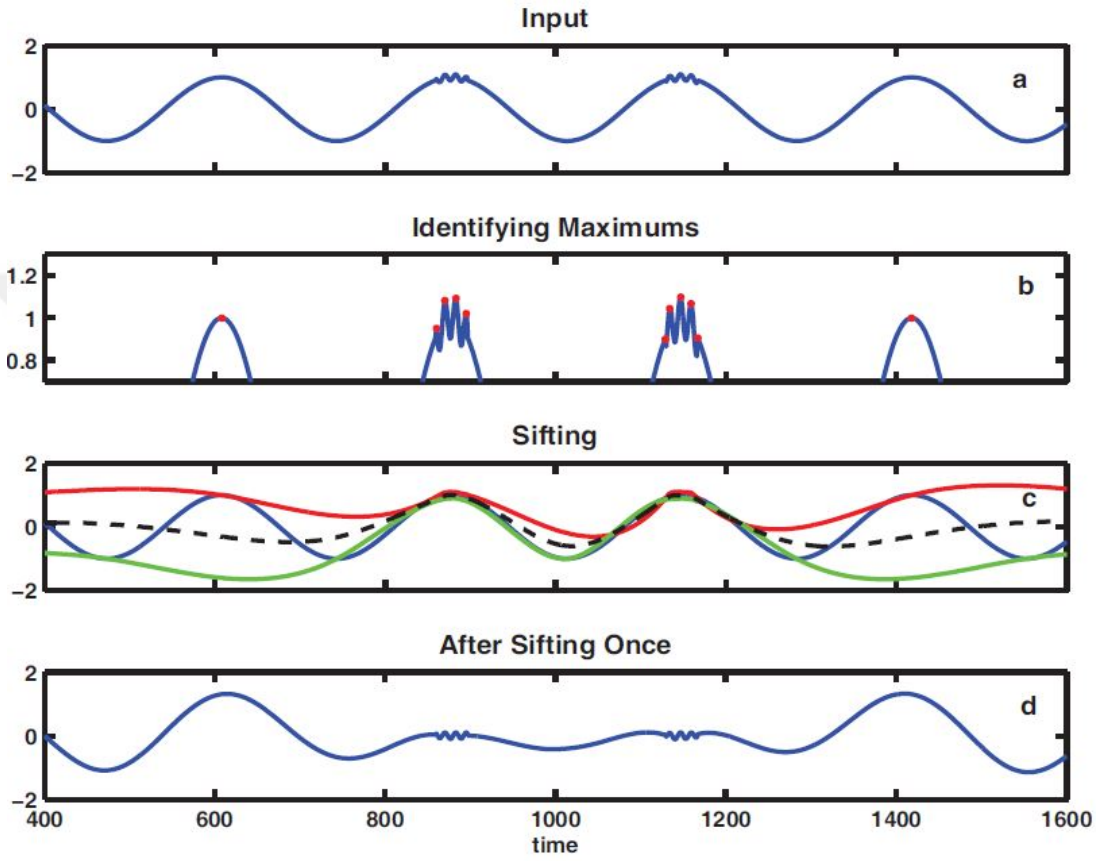


Figure 2.5 A synthetic signal that illustrates the mode mixing problem (Huang, 2014)

The lack of robustness in the decomposition obtained using EMD is an annoying consequence of mode mixing caused by intermittent waves. In real-world data, there is almost always some degree of random noise and discontinuities that users are not aware of beforehand, so it is important to consider whether the decomposition is sensitive to noise in the data. Unfortunately, EMDs generally do not meet this requirement, as they are perfectly adaptable to extreme distributions and the cause of this problem is mode mixing. To improve the reliability of the decomposition, Wu & Huang (2009) proposed EEMD, which aims to alleviate or eliminate the mode mixing problem.

EEMD is developed as follows:

1. Calculate $x_t^i = x_t + w_t^i$ where w_t^i with $i = 1, 2, \ldots, N$ represents different realizations of white noise and N denotes the ensemble number.

2. Decompose each signal x_t^i into the corresponding modes IMF_j^i by EMD.

3. Obtain the modes of EEMD by taking the ensemble means of IMF_j^i as the following:

$$\overline{IMF}_j = \frac{1}{N}\sum_{i=1}^{N} IMF_j^i \qquad j = 1, \cdots, k \tag{2.9}$$

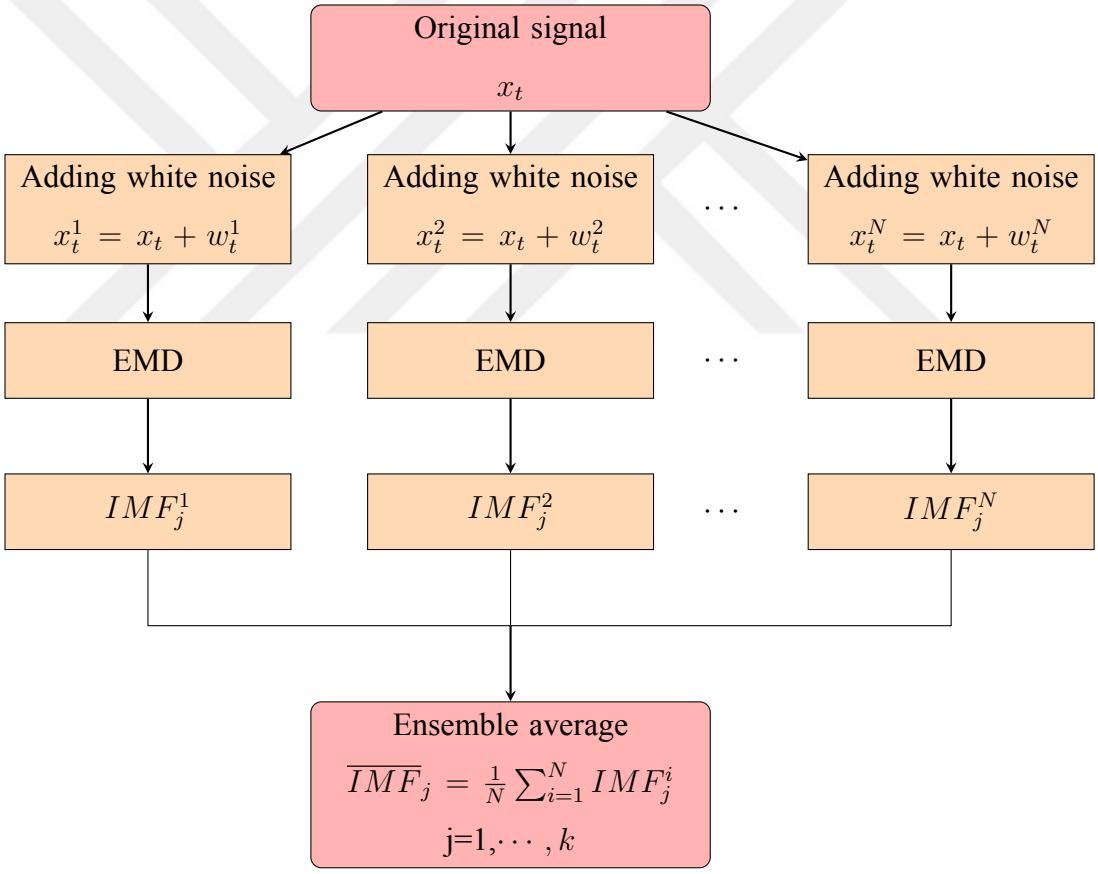


Figure 2.6 Flowchart of EEMD algorithm

To address the mode mixing problem, EEMD adds white noise of finite amplitude to the signal. This populates the entire time-frequency space relatively evenly, and when the signal is added to this uniform background, different scale signals are automatically

projected onto the white noise's proper scales of reference. Although each individual trial may produce very noisy results, the noise in each trial is expected to be canceled out in the ensemble mean of enough trials (Huang, 2014). By increasing the number of samples in the ensemble, the effect of the added white noise can be reduced to a negligible level. In general, an ensemble of a few hundred produces excellent results (Barbosh et al., 2020).

2.3 Vector Autoregressive Model

The VAR model was developed by Sims (1980). It is a widely used multivariate time series model in applied econometrics and finance studies. It is a simple model in which all the variables participating in the model are defined based on their own lagged values and the lagged values of the other variables in the model. VAR models can capture many complicated dynamics observed in stationary multivariate time series. However, a critical point for the VAR model is that the time series in question are stationary. In non-stationary series, it is necessary to achieve stationarity by applying the difference process.

The general form of the VAR model with p lag, denoted VAR(p), can be defined in Equation 2.10.

$$\boldsymbol{Y_t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 Y_{t-1}} + \boldsymbol{\beta_2 Y_{t-2}} + \cdots + \boldsymbol{\beta_p Y_{t-p}} + \boldsymbol{\varepsilon_t} \tag{2.10}$$

where $\boldsymbol{Y_t}$ is a $((k+1)\times 1)$ vector of endogenous variables; $\boldsymbol{\beta_0}$ is a $((k+1)\times 1)$ vector of intercept of the models, $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \cdots, \boldsymbol{\beta_p}$ are $(k+1)\times(k+1)$ matrices of lag coefficients and ε_t is a $((k+1)\times 1)$ vector of white noise residual process. Each variable of $\boldsymbol{Y_t}$ is a linear function of the lag p values for all variables of $\boldsymbol{Y_t}$. Equation 2.10 can be expressed using the delay operator D in Equation 2.11 and the closed form in Equation 2.12.

$$(1 - \boldsymbol{\beta_1} D - \boldsymbol{\beta_2} D^2 - \cdots - \boldsymbol{\beta_p} D^p)\boldsymbol{Y_t} = \boldsymbol{\beta_0} + \boldsymbol{\varepsilon_t} \tag{2.11}$$

$$\boldsymbol{\beta}(D)\boldsymbol{Y_t} = \boldsymbol{\beta_0} + \boldsymbol{\varepsilon_t} \tag{2.12}$$

VAR(1) can be written in matrix form as,

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \end{bmatrix} \tag{2.13}$$

or, equivalently, as in the following system of two equations

$$\begin{aligned} Y_{1t} &= \beta_{01} + \beta_{11} Y_{1t-1} + \beta_{12} Y_{1t-2} + \varepsilon_{t1} \\ Y_{2t} &= \beta_{02} + \beta_{21} Y_{2t-1} + \beta_{22} Y_{2t-2} + \varepsilon_{t2} \end{aligned} \tag{2.14}$$

Each variable in the model has one equation. The current (time t) value of each variable is determined by its own lagged values as well as the lagged values of the other variables in the model.

2.4 Autoregressive Model and Hjorth Parameters

In the literature, feature extraction has been done by using AR models and Hjorth parameters. Lawhern et al. (2012) used the parameters of AR model as features in a support vector machine (SVM) classifier to differentiate between artifact conditions. Zhang et al. (2018) proposed a feature extraction approach for EEG data related to emotions, focusing on the EMD and AR models, and developed a technique for identifying emotions using EEG data by categorizing different emotional states. The algorithm with multi-domain feature complimentary fusion proposed by Zhang et al. (2021) has been successfully applied to patients with stroke. The AR model was used to analyze the time-domain feature of the EEG signals. In another work (Aydemir, 2020), the olfactory responses of the brain were examined, and two identifications

were made: odor and subject. They extracted several features, such as Hjorth parameters and AR model parameters, and they were successful in identifying subjects and odors with high average classification accuracy rates. Additionally, the Hjorth parameters were used as features in Alzheimer's Disease (AD) detection for the three-class problem and they have achieved high accuracy in the detection of AD with the KNN classifier Safi & Safi (2021). AR model and Hjorth parameters are briefly mentioned below as the results will be compared with the proposed method detailed in Section 3.3.

2.4.1 AR Model

The concept of AR modeling was first introduced by the economist (Yule, 1927) and further developed by the statistician Box and Jenkins in the 1950s and 1960s (Box et al., 2015). It is a type of linear model that is used to investigate the dynamic relationship between a single time series variable and its own past values.

AR models were initially developed as a way to analyze economic time series data, such as inflation and interest rates. However, they have since been applied to a wide variety of other fields, including engineering, meteorology, and neuroscience.

An AR model is a statistical model used to understand the relationship between a time series and its past values. It is a type of linear model, which means that it uses a linear equation to describe the relationship between the current value of the time series and its past values. This allows the model to identify patterns and trends in the data, and make predictions about future values of the time series.

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_p x_{t-p} + \varepsilon_t \tag{2.15}$$

where p is the order of the model, $x_{t-1}, x_{t-2}, \cdots, x_{t-p}$ are the lagged values of the series, $\beta_0, \beta_1, \cdots, \beta_p$ are the coefficients of the model and ε_t is a white noise.

Recently, the development of more advanced statistical techniques, such as VAR models and NNs, has expanded the range of problems that can be addressed using AR modeling. Despite this, AR models remain a popular and widely used tool for analyzing time series data.

2.4.2 Hjorth Parameters

Bo Hjorth (1970) introduced Hjorth parameters, which are statistical measurements that are used to analyze signals in the time domain. The parameters are Activity, Mobility, and Complexity, and calculated as follows:

$$Activity = var(x_t) \tag{2.16}$$

$$Mobility = \sqrt{\frac{var(x_t^{'})}{Activity(x_t)}} \tag{2.17}$$

$$Complexity = \frac{Mobility(x_t^{'})}{Mobility(x_t)} \tag{2.18}$$

where $x_t^{'}$ is the first derivative of the signal.

As shown in Equation 2.16, activity is a measure of the signal's variance x_t. It represents the signal's power spectrum in the frequency domain. If the signal contains a lot of high-frequency components, the activity value will be high; otherwise, it is low. This can be referred to as the average power.

The mobility parameter is a measure of how quickly the signal's value is changing over time. It is calculated by taking the square root of the ratio of the variance of the first derivative of the signal (which represents the rate of change of the signal) to the variance of the signal itself, as seen in Equation 2.17. Mobility represents the average

frequency or the standard deviation ratio of the power spectrum. This can be referred to as the average power of the normalized derivative.

The complexity parameter is the ratio of the signal derivative's mobility to the signal's mobility, as in Equation 2.18. Since complexity includes the mobility parameter, it uses the signal's second derivative and is related to changes in signal frequency. The complexity of a signal is a measure of how closely its shape resembles a sine wave. It can be used to estimate the bandwidth of the signal, which is a measure of how much of the signal's power is present at different frequencies.

2.5 Feature Reduction

Feature reduction, or dimensionality reduction, is a process used in machine learning to simplify data by reducing the number of features or dimensions it contains. This is often done to improve the performance of a model, reduce the computational cost of training, and make the data easier to visualize.

There are several ways to reduce the number of features in a data set, including removing irrelevant or redundant features, combining similar features, and selecting a subset of the most important features. The specific approach used will depend on the characteristics of the data set and the goals of the feature reduction.

One common method of feature reduction also used in this thesis is **principal component analysis (PCA)**, which uses linear algebra to identify the underlying structure of the data and reduce the dimensions of the data set accordingly. It is a mathematical technique that was originally developed by Pearson (1901). Pearson's goal was to find a way to represent a data set with many variables in a form that could be easily visualized and analyzed. Over the years, the technique has been refined and extended, and it is now widely used in a various fields, including machine learning, data mining, and pattern recognition.

PCA is a statistical technique that is used to analyze variations in data. It does

this by transforming the data into a new set of variables, called principal components, which are uncorrelated with each other and are ranked based on how much variation they explain in the original data. This transformation makes it easier to visualize and analyze the data and has been shown to be effective when used in conjunction with signal decomposition methods in many studies (Sadiq et al., 2021).

The steps in performing PCA are as follows:

1. Begin with a dataset that has many variables.
2. Center and scale the data. This means that, for each variable in the data set, the mean value is subtracted from it and then the result is divided by the standard deviation. This step is important because it ensures that all variables are on the same scale and have a mean of zero.
3. Compute the covariance matrix of the centered and scaled data. This is a square matrix that contains the pairwise covariances between all the variables in the data set.
4. Compute the eigenvectors and eigenvalues of the covariance matrix. An eigenvector of a matrix is a vector that, when multiplied by the matrix, results in a scaled version of the original vector. The eigenvalues are the scalars that are used to scale the eigenvectors.
5. Select the principal components. The principal components are the eigenvectors that correspond to the largest eigenvalues. These components capture the most variation in the original data.
6. Transform the original data into the new space defined by the principal components. This is done by computing the dot product of the centered and scaled data with the matrix of principal components. The resulting transformed data will have the same number of rows as the original data, but a smaller number of columns (corresponding to the number of principal components selected in step 5).

7. Use the transformed data for visualization and analysis. The transformed data can be plotted in a lower-dimensional space, which allows for easier visualization and interpretation of the patterns and trends in the data. The principal components can also be used to build predictive models or to cluster the data into groups.

Overall, feature reduction is an important technique in machine learning that can help improve the performance and efficiency of models. It can also make the data easier to understand and interpret, which can be valuable for both research and practical applications.

2.6 Classification

Machine learning algorithms are computer learning methods that recognize patterns in EEG signals and are the final component of BCI systems. They use the ability of computers to perceive complex signal patterns and make decisions based on the data to which these patterns belong. They are specialized decision-making methods that are closely related to computational statistics and generally employ rule-based algorithm approaches in computational processes. They can predict the classes to which they belong by observing the feature functions that each of them has for many input values. Machine learning algorithms are widely used for signal pattern recognition in applications such as disease diagnosis, medical applications, and brain-computer interfaces.

Machine learning methods are divided into two according to the structure of the data. The first of these is the supervised method, which is based on the learning method rule. They are trained using labeled data, which consists of input examples and their corresponding desired outputs. The other is unsupervised methods in which the groups in the data are tried to be determined in the absence of any approaches. All of the developed algorithms differ according to their approach to the problem. Therefore, each algorithm has different performances in different problems. In this thesis, the most frequently used supervised machine learning algorithms were selected. KNN,

SVM, decision tree (DT), bagging, Extreme Gradient Boosting (XGBoost), logistic regression (LR), and ANN methods are used to classify the properties of chemosensory EEG signals. The aforementioned methods are briefly explained as follows.

2.6.1 K- Nearest Neighbors

KNN is one of the most basic and simple supervised learning methods developed by Fix & Hodges (1951). Cover & Hart (1967) later extended this method, which is used for classification and regression. The KNN algorithm is based on logic that is probably close to each other. The aim of the algorithm is to classify a new example using the previously classified samples. For this classification process, the value of k is chosen, and for each test data point, the distance is found to all training data points. After that, k nearest data points are examined, determined the most frequently occurring classes, and assigned the new data to that class.

There are different methods for calculating the distance, and the most known method is the Euclidean distance shown in 2.19.

$$d(x_0, x_i) = \sqrt{\sum_{i=1}^{k} (x_0 - x_i)^2} \tag{2.19}$$

Here; $d(x_0, x_i)$ is the Euclidean distance between the nearest k neighbors x_i of a new sample x_0 to be classified.

The most important disadvantage of the KNN algorithm is that the class to be determined changes according to the number of k-nearest neighbors. According to Figure 2.7, a new example, shown in green, has been added to the previously classified triangle and square classes. When the nearest neighbor value is $k = 3$, the two closest neighbors of the new sample are triangles and one neighbor is a square. Based on this information, KNN will determine that the class of the newly added sample should be triangular. When the nearest neighbor value is $k = 5$, the closest 3

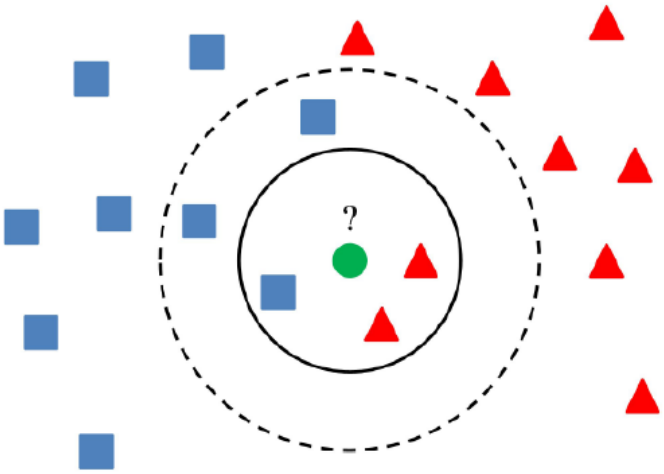


Figure 2.7 Visualization of K- nearest neighbors (Alaliyat, 2008)

neighbors of this new sample are square and 2 neighbors are triangles, and the class of the newly joined sample will be determined as square. In this thesis, various values of k ranging from 2 to 7 are considered and the highest accuracy is achieved when $k = 3$.

2.6.2 Support Vector Machines

SVM developed by Cortes & Vapnik (1995) is a type of supervised learning model that can be used for classification and regression tasks. SVM is one of the most robust prediction methods, and it is based on Vapnik's statistical learning framework (Vapnik, 1999). The purpose of an SVM model is to take a set of training data that has been labeled with two categories and use it to learn a function that can predict which category an example belongs to.

The goal of the SVM algorithm is to find the optimal boundary in an N-dimensional feature space that can be used to separate the data into different classes. The best decision boundary is called a **hyperplane**. The hyperplane is intended so that its distance from the nearest data point in each class is maximum. The dimensions of the hyperplane are determined by the number of features. In N-dimensional space, a hyperplane is a subspace of dimension $N - 1$. For instance,

if the number of input features is two, then the hyperplane is simply a line. If the number of input features is three, then the hyperplane becomes a two-dimensional plane.

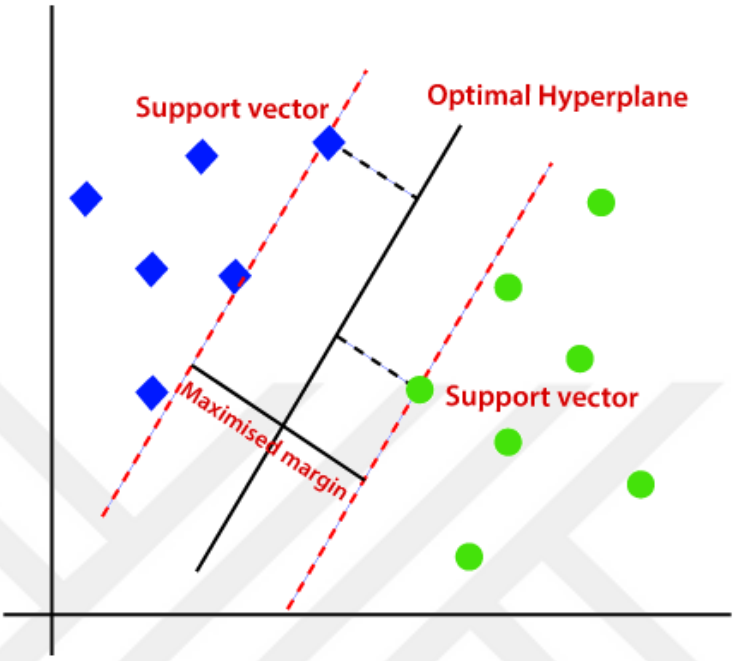


Figure 2.8 Basics of a support vector machine (Javatpoint, b)

In classification tasks, the goal is to predict which class a new data point belongs to based on its features. Given a set of training data with two classes, there may be many different hyperplanes that could be used to separate the data points into the two classes. The SVM algorithm tries to find the hyperplane that maximally separates the classes, which is known as the **maximum margin**. The maximum margin is the distance between the closest data points of the two classes, and it is important because it provides a clear boundary between the classes. By maximizing the margin, we can improve the confidence with which future data points can be classified (James et al., 2013).

The variance and bias must be well balanced to have a good fit model. If the margin decrease, the bias decrease, and the variance increase, and vice versa. It is worth noting that adding more instances outside of the margins has no effect on the decision boundary. The reason for this is that it is fully supported by the units located on the edge of the margin. These units are known as **support vectors**, shown in

Figure 2.8. They are separated by the hyperplane that provides the best fit given the adjusted constraints of the margin.

When all samples are kept off the margin, this is known as **hard margin** classification. This is only effective if the data is linearly separable and highly sensitive to outliers. To avoid these issues, a more flexible model with **soft margin** classifications is recommended. Hard margins may result in the over-fitting of the model. It is sometimes necessary to prefer a soft margin because it produces a more generalizable model when applied to the new data.

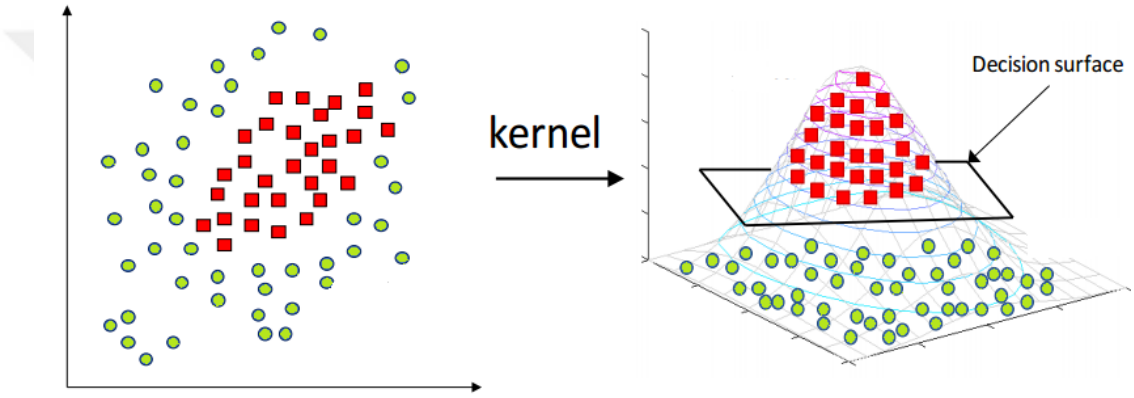


Figure 2.9 Kernel trick in SVM (Zhang, 2018)

Some data sets cannot be separated linearly. In Figure 2.9, we can see that we cannot draw a single line, or hyperplane, that correctly classifies the points. Adding more features is one approach to deal with nonlinear data. It cannot handle very complex data at a low polynomial degree, but at a high polynomial degree, it generates an enormous number of features, making the model too slow. As a result, the data is transformed rom a lower-dimensional space into a higher-dimensional space in order to find a decision boundary that can clearly separate the data points into different classes, and this is called the **kernel trick**. Kernels are the functions that assist us in doing this, and which kernel to use can be determined by hyper-parameter tuning. The most used kernels are polynomial and radial basis functions. In this thesis, the radial basis function kernel is used to evaluate the performance of the features extracted.

2.6.3 Decision Tree

The concept of DTs was first introduced by the computer scientist and statistician Hunt et al. (1966). However, it was not until the work of the computer scientist Breiman and the statisticians Friedman, Stone, and Olshen in the 1980s that DTs became a popular and widely used machine learning algorithm (Breiman et al., 2017).

In the early 1990s, the development of new algorithms, such as random forests and gradient boosting, expanded the range of problems that could be addressed using DTs. These algorithms combine multiple DTs to improve the performance and robustness of the model.

DTs are a popular type of supervised learning algorithm that can be used for classification and regression tasks. They use a tree-like model of decisions and their possible outcomes to make predictions about new data. A DT works by dividing the population or sample into smaller and smaller sub-groups based on the most significant splitter in the input variables (James et al., 2013).

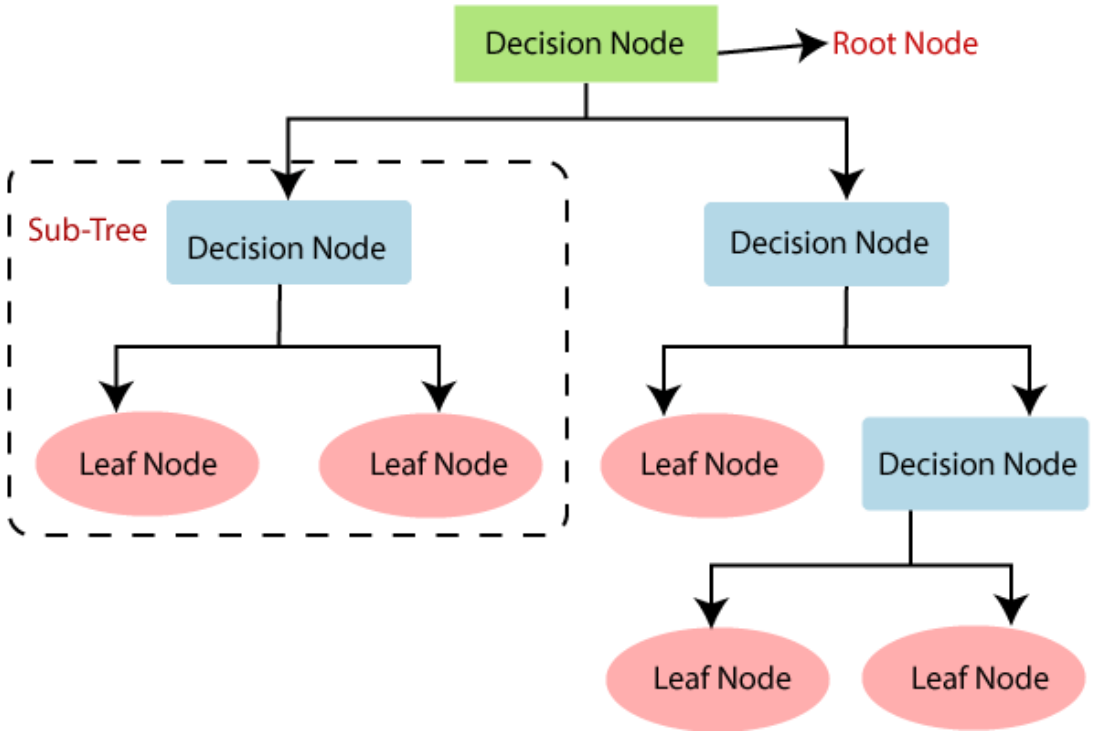


Figure 2.10 The components of a decision tree (Javatpoint, a)

As shown in Figure 2.10, a DT begins with a **root/decision node** symbolizing the entire population or sample, which is then divided into two or more homogeneous sets. **Splitting** is the process of separating the root node into sub-nodes based on the specified conditions. The process of removing sub-nodes from a decision node is known as **pruning**. The opposite of pruning is splitting. When a sub-node divides into further sub-nodes, it is referred to as an **internal/decision node**. The **leaf nodes** represent all of the potential outcomes in the data. After obtaining a leaf node, the tree cannot be divided any further.

If there are numerous characteristics that can be used to split the data, we would like to select the property that is most effective for classifying instances. While there are other approaches for selecting the best feature at each node, Information Gain and Gini Index are two of the most common splitting criteria for decision trees. **Information gain** indicates how effectively a particular attribute distinguishes training instances based on the desired classification. At each stage, a decision tree selects the option that provides the greatest information gain. When the information gain is zero, the feature does not partition the working set. To properly describe information gain, we employ entropy, a well-used quantity in information theory. Entropy characterizes the (im)purity of a random sample collection. **Gini Impurity** represents the probability of misclassification of new data if it was given a random class label based on the class distribution in the data. The details can be found in James et al. (2013).

2.6.4 Bagging

Bagging, or bootstrap aggregating, is a machine learning ensemble technique that is used to improve the stability and accuracy of statistical classification and regression models. Breiman (1996) proposed this ensemble technique, which is applicable to a wide variety of prediction models. This assists in reducing model variance and preventing over-fitting.

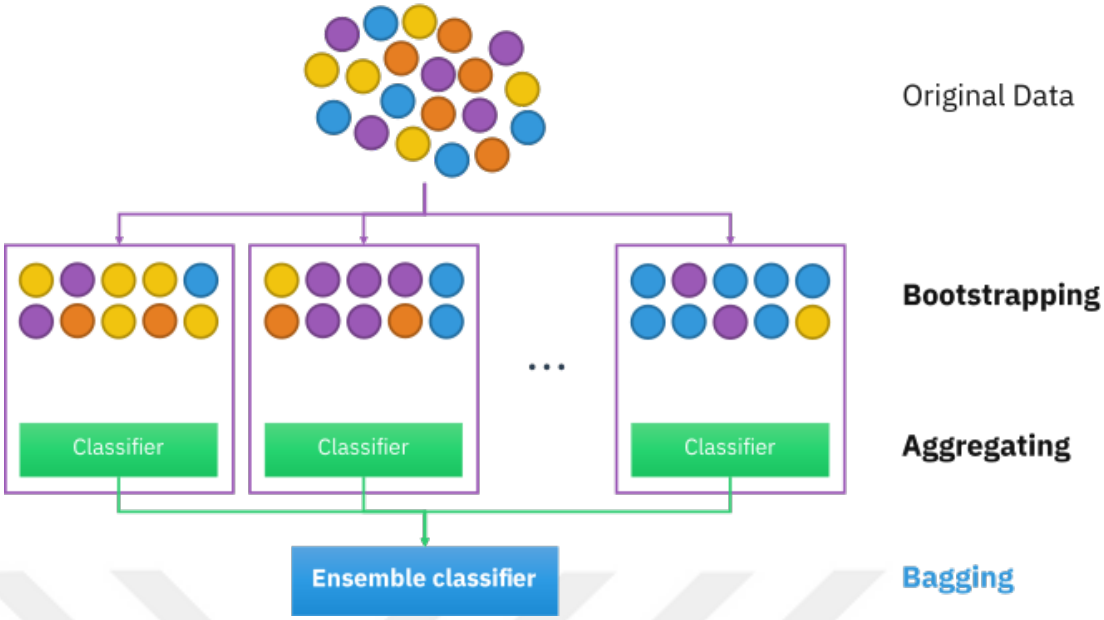


Figure 2.11 An illustration for the concept of bagging (Wikipedia, 2022b)

In bagging, a random sample of data is selected by replacement from a training set in bagging, as shown in Figure 2.11. This sample type is called a bootstrap sample. In this method, each sample is drawn from the original data set with replacement, which means that the same data point can be included in the sample multiple times. This guarantees that each bootstrap sample is independent of the others since it does not depend on the samples that were selected previously. Sampling with replacement is an important aspect of bagging because it allows the model to learn from different subsets of the data, which can improve the stability and accuracy of the final model.

2.6.5 Extreme Gradient Boosting

In Boosting, a subset of the data is extracted, similar to bagging. Again, a model is constructed using this subset. However, from now on, subsequent models will not be independent of the first; rather, they will contribute to its development. While each model has an equal impact on the outcome when bagging is used, boosting tries to fix these errors in the subsequent model by providing more weight to the incorrectly estimated data in the first place. This process is repeated by adding models until either the entire training dataset can be accurately predicted or the maximum number

of models has been reached.

Gradient boosting is a technique in which new models are created by calculating the error in the previous model, and then the remaining error is considered in the prediction to produce the final result. Using a gradient descent approach, gradient boosting minimizes the loss function. The loss function quantifies how well the coefficients of the model fit the underlying data.

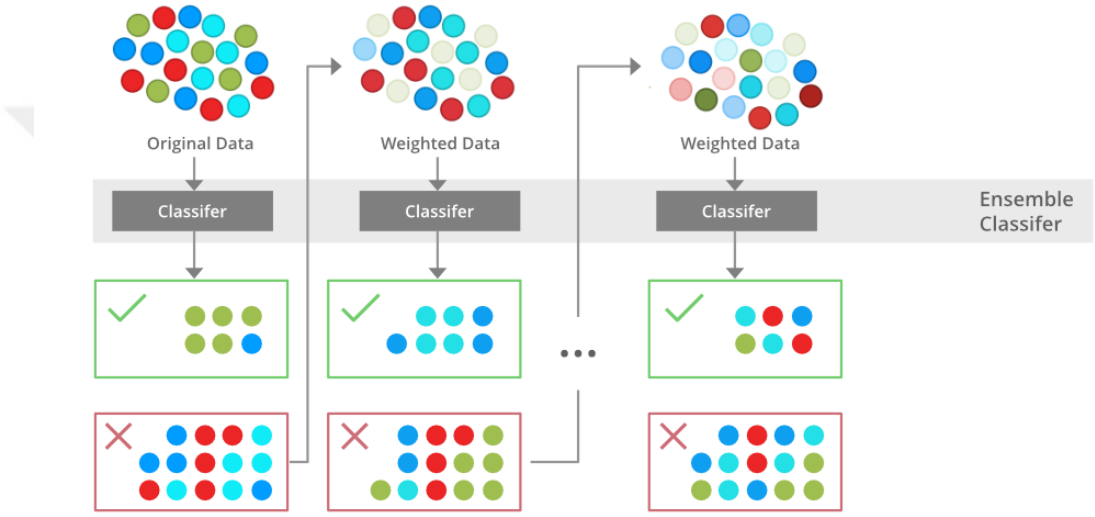


Figure 2.12 An illustration of the concept of boosting (Wikipedia, 2022a)

XGBoost started as a research project by Chen & Guestrin (2016). XGBoost is a machine-learning algorithm that combines multiple decision trees to create an ensemble model. It is based on the gradient boosting framework and is specifically designed to prevent overfitting. One of the key features of XGBoost is its ability to handle missing data and make predictions quickly. In XGBoost, decision trees are created in a sequential fashion, and weights are assigned to each independent variable. If a decision tree makes incorrect predictions, the weight of the variables that were misclassified is increased, and these variables are then passed to a second decision tree. By combining the predictions of multiple decision trees, XGBoost is able to create a more accurate and robust model.

2.6.6 Logistic Regression

The concept of LR was developed by a statistician Cox (1958). Since its development, LR has become a widely used tool in statistics and machine learning. LR has been extended over the years, and it is now a cornerstone of many machine learning and data analysis techniques. It continues to be an active area of research and development, with new algorithms and variations being developed to improve its performance and applicability to a wider range of problems.

LR is a classification method that is used to predict the probability of a particular class for a given set of independent variables. In classical regression methods, it is not always possible to obtain values between 0 and 1 for the dependent variable when it is binary.

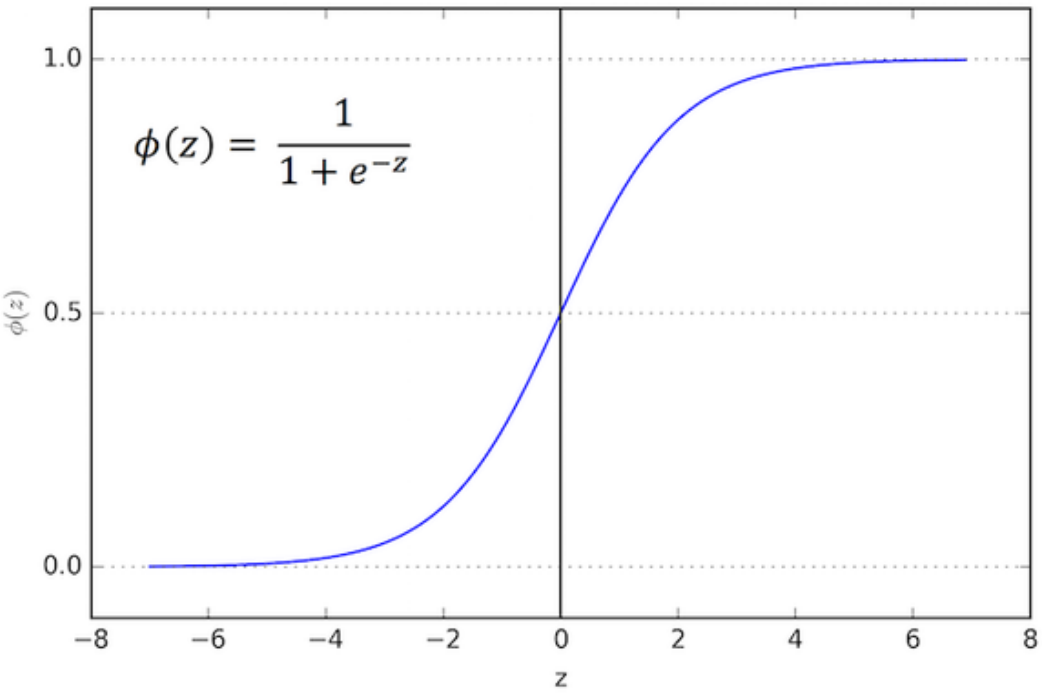


Figure 2.13 Sigmoid function (Sharma, 2017)

Logistic regression uses the sigmoid function to predict the probability that a given data point belongs to a certain class. The sigmoid function, shown in Figure 2.13, is a mathematical function that maps real-valued input to a probability value between 0 and 1. This function takes the predicted values produced by the logistic regression

model and maps them to a probability value, indicating the likelihood that the data point belongs to a specific class. A threshold value is then determined, and values above this threshold are considered 1, and values below it are considered 0, as seen in Figure 2.14. This allows the LR model to make predictions about which class an observation belongs to (James et al., 2013).

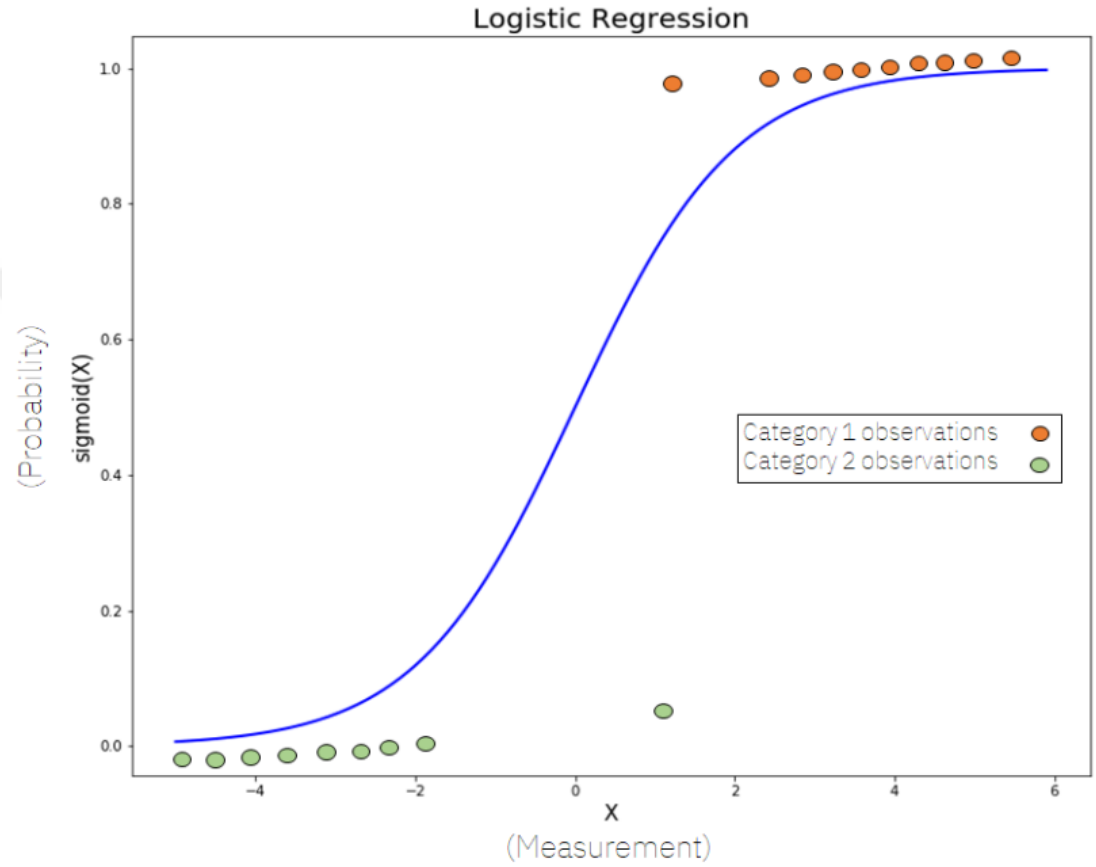


Figure 2.14 Logistic regression for classification (Gupta S., 2022)

2.6.7 Artificial Neural Networks

The story of ANN began with McCulloch & Pitts (1943), who developed a computational model for NN based on algorithms called threshold logic. The idea of ANN has its roots in the work of researchers in the 1950s and 1960s who were exploring the concept of ”cybernetics” or the study of systems that can receive, process, and transmit information. One of the key pioneers in this field was Rosenblatt (1958), who developed the first prototype of a neural network machine,

known as the ”Perceptron”. However, the early versions of NNs were limited in their capabilities and were unable to solve many real-world problems. It wasn’t until the 1980s and 1990s, when the development of more powerful computers and new learning algorithms made it possible to train larger and more complex neural networks, that the true potential of these systems began to be realized.

Today, ANN is a key part of the field of machine learning and has been applied to various tasks, from image and speech recognition to natural language processing and even playing games like chess. NNs continues to be an active area of research and development, with the goal of creating machines that can simulate the complex workings of the human brain.

ANNs are computational models that are inspired by the structure and functions of the human brain. They are composed of many interconnected processing nodes, called ”neurons”, which are organized into layers. The input data is fed through the neural network, one layer at a time, and transformed by the connections between the neurons until it reaches the output layer, where a prediction or classification is produced.

The key elements of an artificial neural network include the following:

- **Inputs:** It refers to the raw data that is fed into the network. This data is typically a set of examples representing the different classes or categories that the network is being trained to recognize. The input is passed through the network, one example at a time, and it is transformed by the connections between the neurons in the network.
- **Nodes/ Neurons:** These are the computational units of a neural network, and they are organized into layers. Each neuron receives input from other neurons in the previous layer and applies a mathematical operation to those inputs to produce an output.
- **Layers:** Neural networks are composed of multiple layers of neurons, which are organized into an input layer, one or more hidden layers, and an output layer.

The input layer receives the raw input data, the hidden layers apply non-linear transformations to the data, and the output layer produces the final prediction or classification.

- **Connections:** The connections between neurons carry the data that flows through a neural network, and they are associated with weights that determine the strength of the connection. The weights of the connections are adjusted during the training process to improve the network's performance.
- **Activation function:** This is a mathematical function that is applied by each neuron to the input it receives from the previous layer. The activation function determines the output of the neuron and is a key factor in the non-linear processing capabilities of a neural network.
- **Output:** It refers to the final result of the computation performed by the network, and it represents the network's prediction or decision-based on the input data.

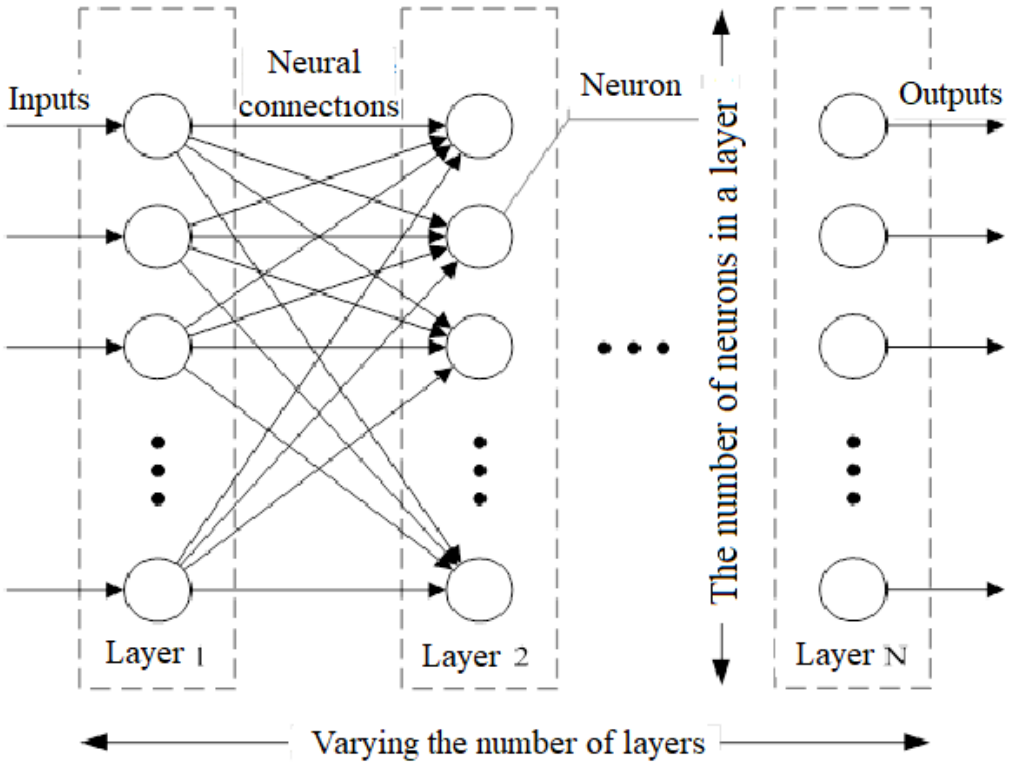


Figure 2.15 Block diagram of an artificial neural network (Vasiliev et al., 2019)

ANNs, illustrated in Figure 2.15, are composed of layers of interconnected nodes/neurons, which are inspired by the structure of neurons in the human brain. A node is a computational unit that receives input from the data and applies a set of coefficients, called weights, to amplify or dampen that input. The input data is multiplied by a set of weights, and the resulting products are summed and processed by an activation function. The activation function determines whether and to what extent the processed data should be passed on to the next layer of the neural network. This process is repeated for each layer of the network until a final output is produced. The weights of the connections between the nodes are adjusted during the training process to improve the network's performance. ANN can learn from the data and improve their performance over time. To train a NN, a set of labeled training data is fed into the network, and the network's weights and biases are adjusted to minimize a loss function that measures the difference between the network's predictions and the true labels in the data.

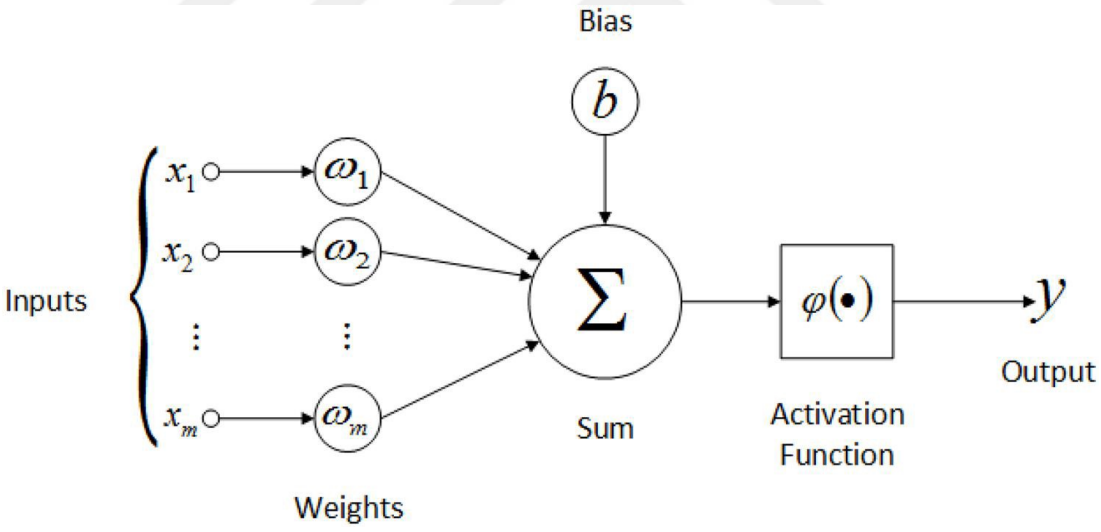


Figure 2.16 A single layer neural network (Raut, 2020)

In this thesis, a single hidden layer NN is used and is a type of ANN that is composed of three layers: an input layer, a hidden layer, and an output layer. The input layer receives the raw input data, the hidden layer applies a non-linear transformation to the data, and the output layer produces the network's final prediction or classification. Single hidden layer NN, shown in Figure 2.16, is considered a relatively simple architecture, compared to more complex NNs with

multiple hidden layers. However, they are still capable of solving many real-world problems and can be trained to achieve high levels of performance. In a single hidden layer NN, the neurons in the hidden layer are the key computational units. Each neuron receives input from the input layer and applies a mathematical operation to those inputs to produce an output. This output is then passed on to the output layer, where it is used to make the final prediction or classification.

Overall, ANNs are powerful tools for generating predictions and classifications based on complex input data. They have been applied to various applications, including image and speech recognition, natural language processing, and even medical diagnosis.

2.7 Performance Evaluation

The key concept for accurately measuring a classifier's performance is a generalization. For this purpose, we use leave-one-out cross-validation (LOOCV) which is a model validation technique that is used to evaluate the performance of a machine learning model in this thesis. It involves training the model on all but one of the available data points and then testing the model on the remaining data point shown in Figure 2.17. This process repeats n times (where n is the total number of observations in the data), excluding a different observation from the training set each time (James et al., 2013). The final performance of the model is then calculated as the average performance across all the test sets. LOOCV is a type of k-fold cross-validation, where k is equal to the number of data points in the data set. It is commonly used when the data set is small and it is not feasible to split the data into many folds. One advantage of LOOCV is that it provides a very thorough evaluation of the model's performance since each data point is used as a test set exactly once.

Assessing the performance of a classification algorithm is the process of evaluating how well the algorithm can predict the correct class labels for a given set of data (Hossin & Sulaiman, 2015). There are several metrics that can be used to assess the

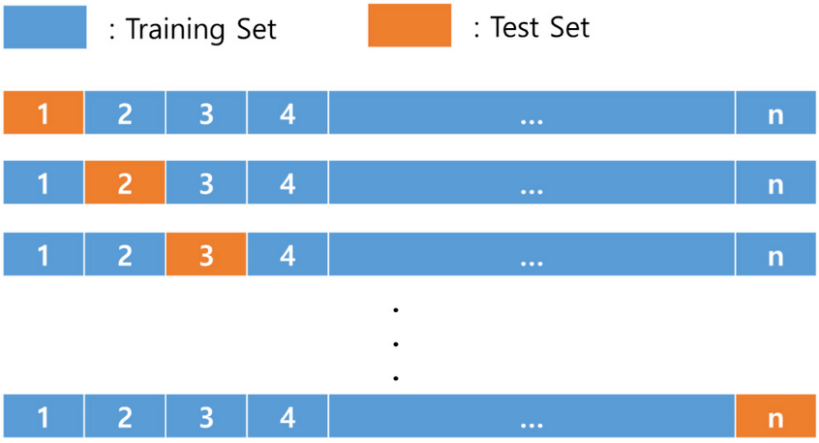


Figure 2.17 Schematic representation of the leave-one-out cross-validation (Cha et al., 2020)

performance of a classification algorithm. These performance metrics were calculated from the confusion matrix in Table 2.1. A confusion matrix is a tool for evaluating the performance of a classification algorithm. It is a table that is used to visualize the performance of the algorithm, with the columns representing the predicted class labels and the rows representing the true class labels. It is a useful tool for understanding the performance of a classification algorithm and for identifying areas where the algorithm may be making mistakes.

Table 2.1 Structure of a confusion matrix

	Actual	
Predicted	**Positive**	**Negative**
Positive	TP	FP
Negative	FN	TN

A **true positive (TP)** is a prediction where the algorithm correctly predicts the positive class. For example, in the case of a disease diagnosis, a TP would be a prediction that a subject has the disease and it actually has the disease.

A **true negative (TN)** is a prediction where the algorithm correctly predicts the negative class. Continuing with the disease diagnosis example, a TN would be a prediction that a subject has no disease and it actually has not the disease.

A **false positive (FP)** is a prediction where the algorithm incorrectly predicts the positive class. In the disease diagnosis example, a false positive would be a prediction

that a subject has the disease but actually has not the disease.

A **false negative (FN)** is a prediction where the algorithm incorrectly predicts the negative class. In the disease diagnosis example, an FN would be a prediction that a subject has no disease but it actually has the disease.

The performance of the classifiers is evaluated based on using accuracy (ACC), sensitivity (SEN), specificity (SPF) and area under the curve (AUC) in this thesis. The mathematical expressions of the performance measure parameters are given in Equations 2.20-2.22 and briefly explained below.

Accuracy (ACC) is a metric used to evaluate the performance of a classification algorithm. It is defined as the proportion of the correct predictions made by the algorithm. In other words, it is the number of times the algorithm predicted the correct class label, divided by the total number of predictions made. Accuracy is a useful metric for evaluating the overall performance of a classification algorithm, but it can be misleading if the data set is unbalanced (i.e. if one class is much more prevalent than the other). In this case, the accuracy of the algorithm may be high even if it is not making very good predictions for the less prevalent class. Therefore, it is often useful to consider other metrics in addition to accuracy when evaluating the performance of a classification algorithm.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{2.20}$$

Sensitivity (SEN), also known as the true positive rate or recall, is a metric used to evaluate the performance of a binary classification algorithm. It is defined as the proportion of actual positive instances that were correctly predicted by the algorithm. In other words, it is the number of true positives (i.e. the number of times the algorithm predicted the positive class and was correct) divided by the total number of actual positive instances in the data. Sensitivity is a useful metric for evaluating the ability of the algorithm to correctly identify positive instances. It is often used along

with other metrics, such as precision, to get a more complete picture of the performance of the algorithm.

$$SEN = \frac{TP}{TP + FN} \times 100 \tag{2.21}$$

Specificity (SPF), also known as the true negative rate, is a metric used to evaluate the performance of a binary classification algorithm. It is defined as the proportion of actual negative instances that were correctly predicted by the algorithm. In other words, it is the number of true negatives (i.e. the number of times the algorithm predicted the negative class and was correct) divided by the total number of actual negative instances in the data. Specificity is a useful metric to evaluate the ability of an algorithm to correctly identify negative instances. It is often used along with other metrics, such as sensitivity, to get a more complete picture of the performance of the algorithm.

$$SPF = \frac{TN}{TN + FP} \times 100 \tag{2.22}$$

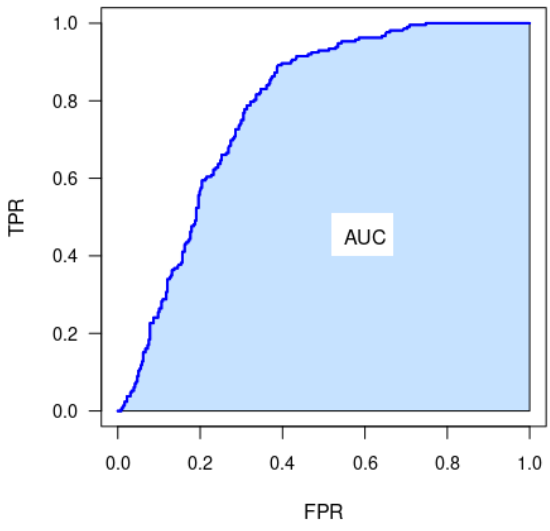


Figure 2.18 The area under the curve (Wikipedia, 2022c)

The area under the curve (AUC) is a metric used to evaluate the performance of a classification algorithm. It is calculated by finding the area under the receiver

operating characteristic (ROC) curve, as seen in Figure 2.18. It is a plot of the true positive rate (TPR), calculated as TP/TP+FN, versus the false positive rate(FPR), calculated as FP/FP+TN, for different classification thresholds. The AUC provides a single number that summarizes the performance of the algorithm across all possible thresholds. A perfect classifier would have an AUC of 1, while a completely random classifier would have an AUC of 0.5. The higher the AUC, the better the performance of the classification algorithm (Ling et al., 2003).

CHAPTER THREE
A NOVEL HYBRID FEATURE EXTRACTION METHOD

In this chapter, we propose a novel approach for extracting features from EEG signals. Our proposed procedure is illustrated in Figure 3.1. We start the process with a pre-processing stage using the raw EEG signals. IMFs are obtained using EEMD and are used as independent variables in the VAR model. The coefficients of the VAR model are employed as features in the proposed method. PCA is employed for feature reduction. The selected features are fed to the classifier to distinguish subjects with PD from healthy subjects. In the final stage, the performance of the classifiers is evaluated. The data used in this thesis and every step of the method are explained in the following subsections.

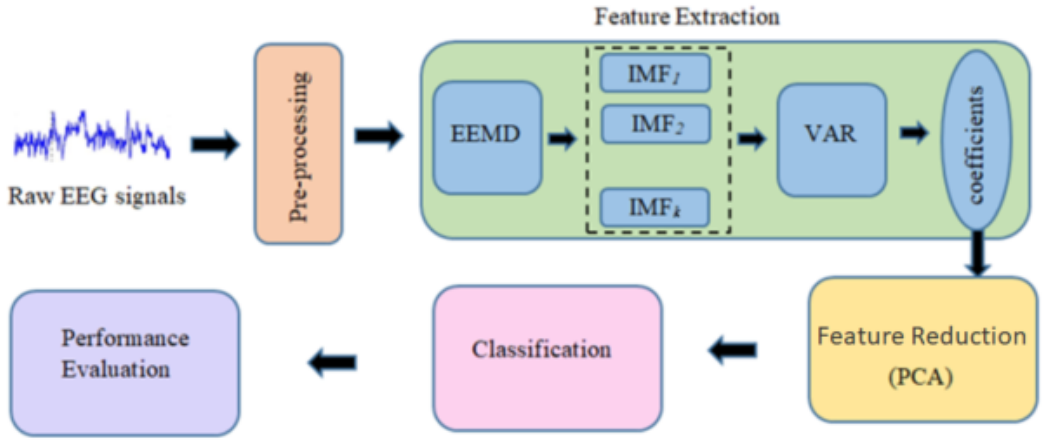


Figure 3.1 The general process of the proposed procedure

3.1 Data Acquisition

The data used in this thesis were recorded at the Department of Biophysics of Dokuz Eylul University (DEU) and approved by the Clinical Ethical Committee of DEU. The data set was used previously by Guducu (2015) and Guducu et al. (2015). Twenty newly diagnosed patients with PD (7 males, 13 females; mean age: 53.92 $\pm$ 9.04 years) and twelve healthy controls (7 males, 5 females; mean age: 52.42 $\pm$ 7.69 years) were included in the study. The subjects with PD were newly-diagnosed

unmedicated PD patients. The participants in the control group had no previous diagnosis of neurological or psychiatric diseases. Before the recording, the subject's entire medical history was collected and examined by the same otolaryngologist to rule out any sinus disease that may have a negative impact on olfactory function. Informed consent was obtained before participation.

Chemosensory stimuli were presented using an air-diluted olfactometer (Om2b, Burghart Messtechnik, Germany). The olfactory stimuli in the olfactometer were 2-phenyl-ethyl-alcohol (PEA, 60% v/v, Merck, Germany). The stimulus duration was set as 200 milliseconds (ms), and the stimuli were randomly delivered every 15-17 seconds. A 64-channel recording device was used to record all EEG signals at a sampling frequency of 1 kHz (Synamps 64, Neuroscan Inc., USA).

3.2 Pre-Processing

Pre-processing is a common step in the data analysis process that involves cleaning and transforming the raw data before it is used for further analysis. This typically involves a number of steps, such as removing missing data, scaling the data to ensure that all variables are on the same scale, and transforming the data to make it more suitable for the intended analysis. Pre-processing is an important step because it ensures that the data is in a form that can be easily and accurately analyzed, and it can also help improve the performance of any subsequent analysis or modeling.

The pre-processing and ERP analysis procedure was as follows:

- The raw EEG data were filtered with a 50-Hz notch filter to reduce the power line effects in the recordings. Power line effects are a common source of noise and interference in electrical signals, such as those recorded from the human body using electrodes. These effects are caused by the presence of alternating current in the power grid, which can produce voltage fluctuations that can interfere with the signal of interest. Power line effects can be particularly

problematic for signals with a low signal-to-noise ratio, such as EEG signals, where they can make it difficult to accurately interpret the data. To reduce the effects of power line interference, signal processing techniques, such as filtering, can be used to remove unwanted frequencies from the signal.

- An offline band-pass filter with 0.5-48 Hz cut-off frequencies was applied. An offline band-pass filter is a digital filter that is applied to a signal after it has been recorded, or ”offline”. This type of filter is used to remove frequencies that are outside a specified range, known as the ”passband” while allowing frequencies within the passband to pass through. The result is a ”filtered” version of the original signal that has reduced noise and artifacts and is more suited for further analysis.

- Trials with the ocular activity, which is the movements of the eyes, including both voluntary and involuntary movements exceeding 50 μV were discarded.

- The EEG signals were divided into epochs, a fixed-length segment of the time series, of 3,000ms based on the event (1000ms before target onset to 2000ms after target onset). Epoching the data allows for a more detailed and fine-grained analysis of the time series data and can help to identify subtle differences or trends that might be difficult to detect in the raw data.

- In signal analysis, baseline correction is a method used to remove any offsets or variations in the baseline, or the average background level, of a signal. The baseline correction can be performed using various mathematical algorithms, such as polynomial fitting or moving average smoothing. The goal is to accurately determine the baseline level so that any peaks in the signal can be accurately measured and identified. The sum of ERPs was constructed from trials that consisted of 3,000 milliseconds of epochs. The first 1,000 milliseconds of this period, referred to as the pre-stimulus period, was used to adjust the baseline for the ERP epochs, which lasted for 2,000 milliseconds. The same number of stimuli was delivered to all subjects. However, after filtering and artifact rejection, the number of trials used in the analysis differed.

On average, 24.9 ± 5.1 trials were used in the data analysis.

- In order to provide the stationarity assumption of the VAR model, the first difference of the time series obtained in the previous stage was taken.

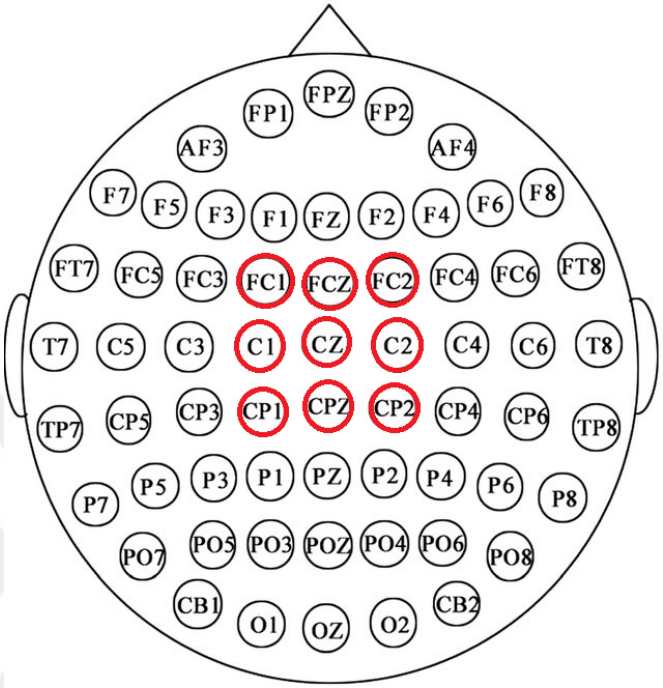


Figure 3.2 The EEG cap layout for 62 electrodes

- In an EEG system, multiple electrodes are typically placed on the scalp to record the electrical activity of the brain. However, not all of these electrodes may be suitable for a particular analysis, and choosing the correct electrodes can improve the accuracy and reliability of the results. Channel selection in EEG signal analysis may be based on factors such as the location of the electrodes on the scalp, the signal-to-noise ratio of the recorded signals, and the specific brain activity studied. In this thesis, CZ, FC1, FC2, FCZ, C1, C2, CP1, CPZ, and CP2 electrodes were used and labeled in red circles in Figure 3.2 which shows the layout of EEG electrodes on the cap. Boesveldt et al. (2007) and Whitcroft et al. (2017) proposed the use of 5 electrodes including the FZ, CZ, PZ, C3, and C4. This selection was adapted according to our cap positions and within the central line of electrodes since this region of interest (ROI) includes any activity related to olfactory processing. Besides, the lateral

electrodes were included to increase the information related to olfactory processing.

- The 400-900ms of the post-stimulus period were considered for each electrode of each subject since the subjects are predicted to respond to the stimulus approximately within this range (Guducu et al., 2019, Huart et al., 2013).

3.3 Proposed Method

In this section, the details of the EEMD_VAR, the proposed feature extraction procedure, are given. As can be seen in Figure 3.3, the algorithm is divided into two phases. The EEMD is applied to the stationary data in the feature extraction phase, which is the first phase. The signal is decomposed into IMF components via the EEMD method. Then, the IMFs are used as inputs in the VAR model. The acquired features are selected by the PCA. In the second phase, the selected features feed the classification algorithms and the performance metrics are calculated.

To analyze the statistical relationship between the EEG signal and its IMFs, a VAR model approach incorporating different numbers of IMFs was used in this thesis. IMFs obtained from Equation 2.9 and original signal x_t are used as variables in Equation 2.10. In the EEMD_VAR, we define $\boldsymbol{Y_t}$ as a $((k+1)\times 1)$ vector as follow:

$$\boldsymbol{Y_t} = \begin{bmatrix} x_t \\ \overline{IMF}_1 \\ \vdots \\ \overline{IMF}_k \end{bmatrix} \tag{3.1}$$

For the sake of ease, we give an example of the EEMD_VAR model with two IMFs and two lags, the components of the VAR model are defined in Equations 3.2-3.3.

$$\boldsymbol{Y_t} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 Y_{t-1}} + \boldsymbol{\beta_2 Y_{t-2}} + \boldsymbol{\varepsilon_t} \tag{3.2}$$

$$\begin{aligned}\begin{bmatrix} x_t \\ \overline{IMF}_1 \\ \overline{IMF}_2 \end{bmatrix} &= \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \end{bmatrix} + \begin{bmatrix} \beta_{11}^1 & \beta_{12}^1 & \beta_{13}^1 \\ \beta_{21}^1 & \beta_{22}^1 & \beta_{23}^1 \\ \beta_{31}^1 & \beta_{32}^1 & \beta_{33}^1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \overline{IMF}_{1t-1} \\ \overline{IMF}_{2t-1} \end{bmatrix} \\ &+ \begin{bmatrix} \beta_{11}^2 & \beta_{12}^2 & \beta_{13}^2 \\ \beta_{21}^2 & \beta_{22}^2 & \beta_{23}^2 \\ \beta_{31}^2 & \beta_{32}^2 & \beta_{33}^2 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ \overline{IMF}_{1t-2} \\ \overline{IMF}_{2t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \varepsilon_{t3} \end{bmatrix}\end{aligned} \tag{3.3}$$

Here, we only consider the coefficients of the first model in Equation 3.3 for classification since the procedure explains the effects of IMFs on the EEG signal. The model, whose coefficients are concerned, is given in Equation 3.4. The input signal, x_t is a linear function of the lagged values of itself and the lagged values of its IMFs.

$$\begin{aligned} x_t = \beta_{01} + \beta_{11}^1 x_{t-1} + \beta_{12}^1 \overline{IMF}_{1t-1} + \beta_{13}^1 \overline{IMF}_{2t-1} \\ + \beta_{11}^2 x_{t-2} + \beta_{12}^2 \overline{IMF}_{1t-2} + \beta_{13}^2 \overline{IMF}_{2t-2} + \varepsilon_{t1} \end{aligned} \tag{3.4}$$

According to Equation 3.4, seven coefficients are obtained for two IMFs and two lags. It is required to reduce the dimension since too many coefficients are obtained for large values of k and p. Therefore, PCA is applied to the obtained autoregressive coefficients of the VAR model. In this study, two components are used as features according to the cumulative explained variance ratio which is more than 90% of the total variability.

Feature extraction phase

Input: x_t,$\bar{x}_t$,$\bar{x}_t^{\varphi_{tr}}$,φ_{tr},p,n_{f_t},$f_t{}^{\varphi_{tr}}$,$f_{t_{n_{f_t}}}^{\varphi_{tr}}$,$\delta_{VAR}{}^{\varphi_{tr}}$

Given: x_t, $\bar{x}_t$, $\bar{x}_t^{\varphi_{tr}}$

for each x_t **do**

  Compute $\bar{x}_t = x_t - x_{t-1}$

end for

for each $\bar{x}_t^{\varphi_{tr}}$ **do**

  for each n_{f_t} **do**

    for each p **do**

      Run EEMD model $\bar{x}_{t_p}^{\varphi_{tr}}$

      Save $n_{f_t}^{th}$ $f_t{}^{\varphi_{tr}}$ intrinsic mode function

    end for

    Run VAR model $\bar{x}_t^{\varphi_{tr}}$ $f_{t_{n_{f_t}}}^{\varphi_{tr}}$

    Save $\delta_{VAR}{}^{\varphi_{tr}}$

  end for

  Run PCA $\delta_{VAR}{}^{\varphi_{tr}}$

end for

Output: Component ϑ_C to feed classification algorithms

Classification phase

Input: ϑ_C

Given: y

for $i = 1$ to $length(\vartheta_C)$ **do**

  Run C classifier for $\vartheta_C[i]$

  Calculate $y^{'}$ = test$(C, \vartheta_C[i])$ ▷ test C on $\vartheta_C[i]$

  Update confusion_matrix

end for

Calculate performance_metrics

Output: ACC, SEN, SPF, AUC

Figure 3.3 The EEMD_VAR Algorithm

CHAPTER FOUR
RESULTS AND DISCUSSIONS

Twenty newly diagnosed patients with PD and twelve healthy controls were included in the study. The signals and the IMFs of randomly selected healthy and PD subjects in the range of 400-900 ms are shared as examples in Figure 4.1 and Figure 4.2, respectively. The number of IMFs (k) and lag (p) is selected between 3-10 (eight different cases) and 1-20 (twenty different cases), respectively. The scenarios are created by using all p values for each p and the procedure given in Figure 3.1 is employed for all scenarios. and accuracy, sensitivity, specificity, and area under the curve performance metrics are calculated for each of the selected nine electrodes and used classification methods. All analyzes were performed in R 4.1.2. A heat map shows all the accuracy results to see the variation in results for 160 scenarios using LR in C2 as an example in the Figure 4.3.

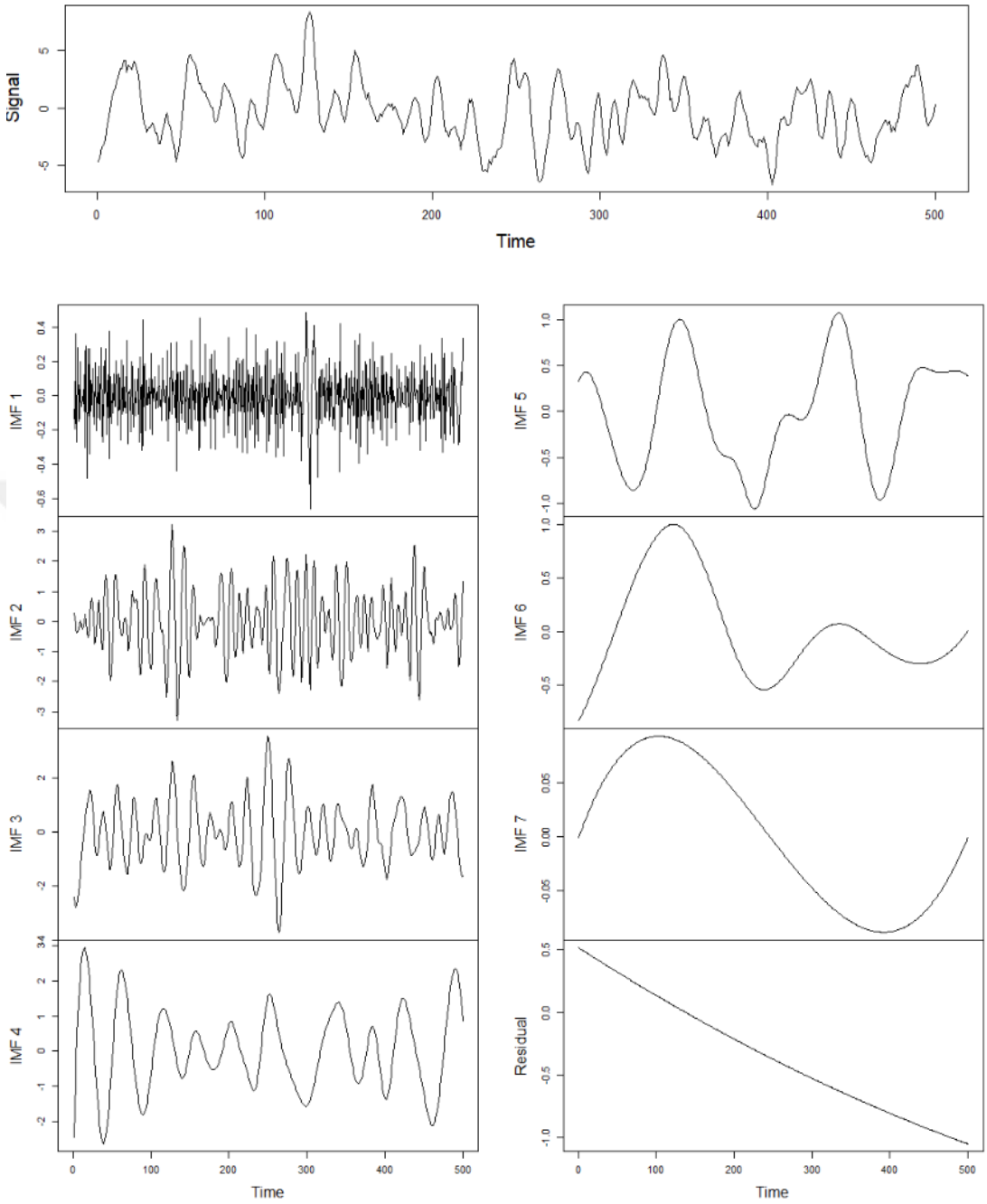


Figure 4.1 The EEG signal for a control subject and its IMF components from data through the EEMD method

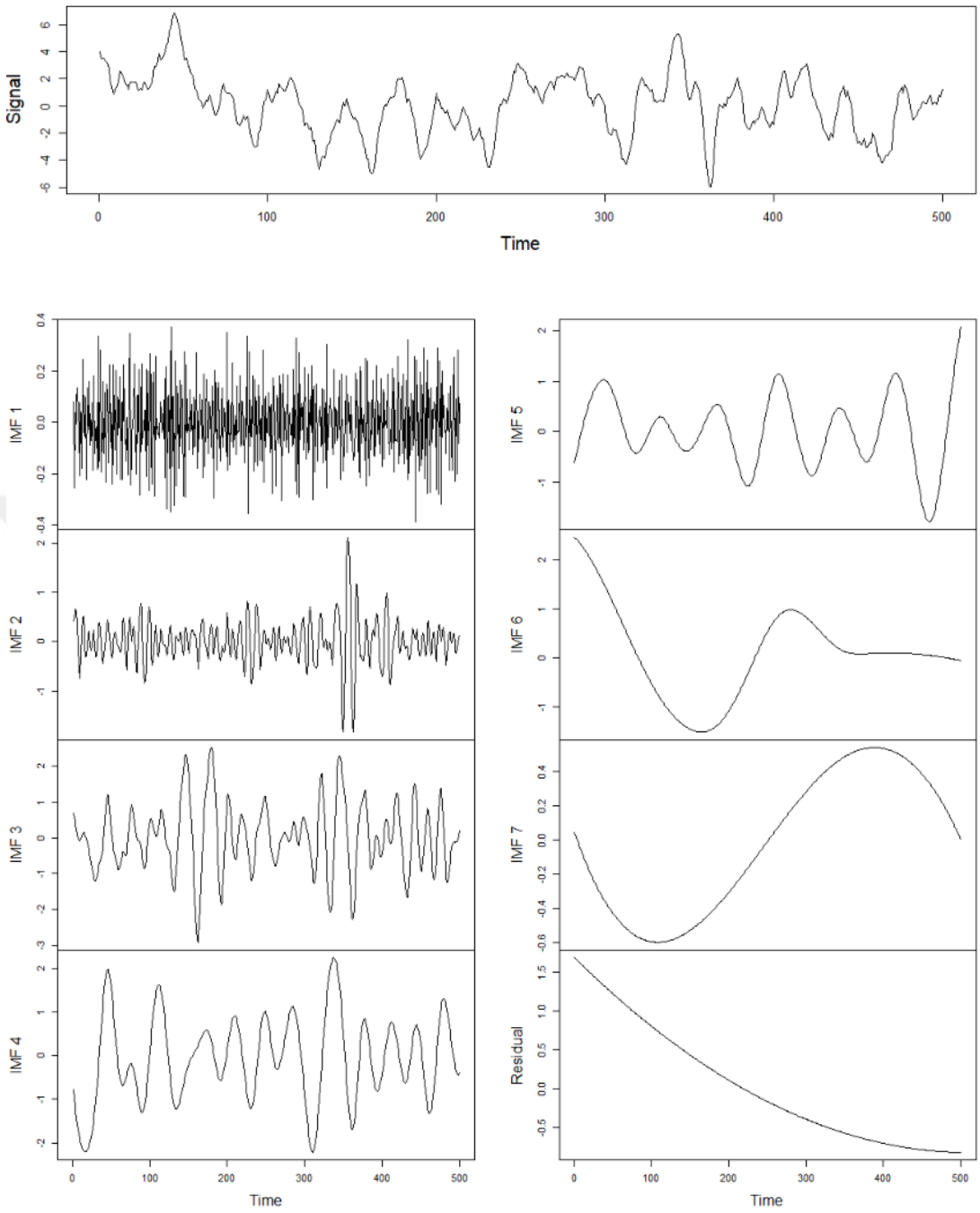


Figure 4.2 The EEG signal for a subject with PD and its IMF components from data through the EEMD method

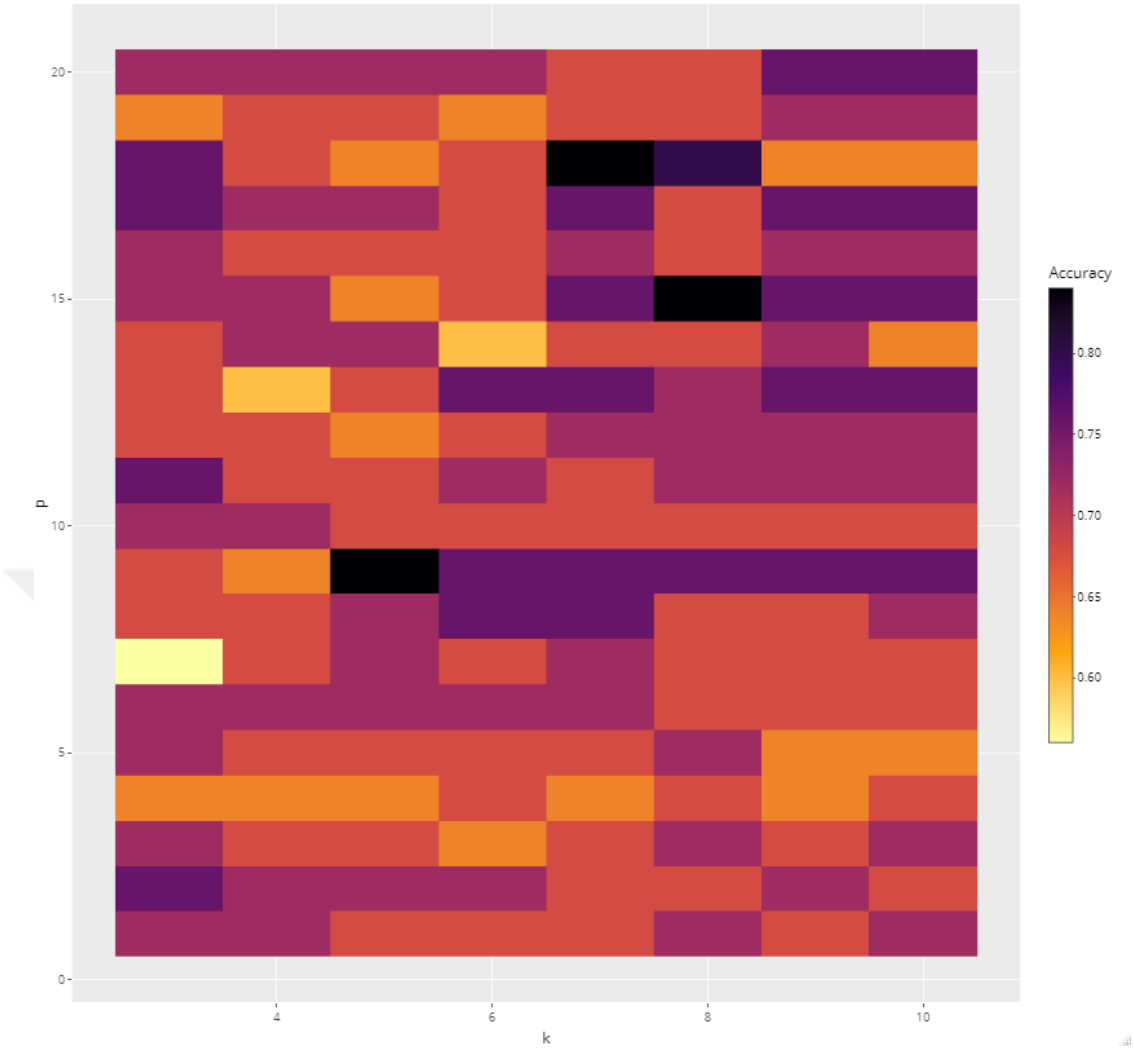


Figure 4.3 Accuracy results for 160 scenarios using LR in C2

4.1 Findings

Only results that have maximum accuracy have been presented due to the lack of space and ease of understanding in Table 4.1. Bold values indicate the highest maximum accuracy obtained from the related classifier. The asterisks (*) indicate that the marked maximum accuracy value has reached more than once in 160 different scenarios. The values in parentheses show k and p reached that maximum accuracy. In other words, they indicate in which of the 160 scenarios, maximum accuracy is achieved. The purpose of sharing the values in parentheses is not to show the

optimum k and p values that should be used, but to give an idea to the researcher trying to determine these values. We should first highlight that almost all accuracy values in achieved more than 0.75 for each electrode and each classification method. The highest accuracy of each electrode was as follows: FC1, 0.8125 (DT and ANN); C1, 0.8966 (DT and ANN); CP1, 0.8276 (ANN); FCZ, 0.8125 (KNN, SVM, DT, and LR); CZ, 0.875 (Bagging); CPZ, 0.8438 (DT); FC2, 0.8966 (LR); C2, 1.000 (ANN); and CP2, 0.9286 (LR). According to the table, the highest accuracy in distinguishing between the subjects with PD and the controls exhibited 100% accuracy using ANN in C2. Moreover, 0.9286 and 0.8966 were obtained from the LR. Overall, higher accuracies were achieved from right-side electrodes, which may be a hint that a clearer picture of olfactory dysfunction can be obtained from the right side of the brain.

Table 4.1 The maximum accuracy values obtained using the 160 scenarios for the EEMD_VAR method

	KNN	SVM	DT	Bagging	XGBoost	LR	ANN
FC1	0.7813 (8,4)	0.7813 (7,18)	0.8125 (6,10)	0.7813 (5,20)	0.7500 (5,18)	0.7813 (4,15)	0.8125 (8,10)
C1	**0.8621** (5,5)	0.7586* (5,5)	**0.8966** (10,20)	0.7931* (6,5)	0.7586 (7,3)	0.7931* (5,5)	0.8966 (5,5)
CP1	0.7931 (7,5)	0.7586* (6,5)	0.7931* (9,10)	0.7586 (7,5)	0.7241* (9,10)	0.7931 (4,6)	0.8276 (3,5)
FCZ	0.8125 (5,18)	0.8125 (9,19)	0.8125 (4,9)	0.7813 (8,10)	0.7813 (7,2)	0.8125* (3,3)	0.7813* (8,2)
CZ	0.8125 (6,2)	0.8125 (7,13)	0.8438 (4,4)	**0.8750** (8,5)	0.8438 (7,1)	0.8125* (5,8)	0.8438 (5,4)
CPZ	0.7500 (3,11)	0.7813 (6,5)	0.8438 (8,5)	0.7500 (10,16)	0.7188 (3,13)	0.7813 (9,15)	0.7500* (9,20)
FC2	**0.8621*** (5,18)	0.8621 (10,1)	0.8621 (5,18)	0.8621 (6,7)	**0.8621*** (5,11)	**0.8966** (6,7)	0.8621* (10,10)
C2	0.8400* (10,10)	**0.8800** (3,19)	0.8800* (8,11)	0.8400* (8,15)	0.8400* (7,18)	0.8400* (4,19)	**1.000** (6,19)
CP2	0.7857 (8,19)	0.8214 (3,2)	0.8571 (6,19)	0.7857* (10,5)	0.8214 (6,2)	**0.9286** (5,2)	0.7500* (6,2)

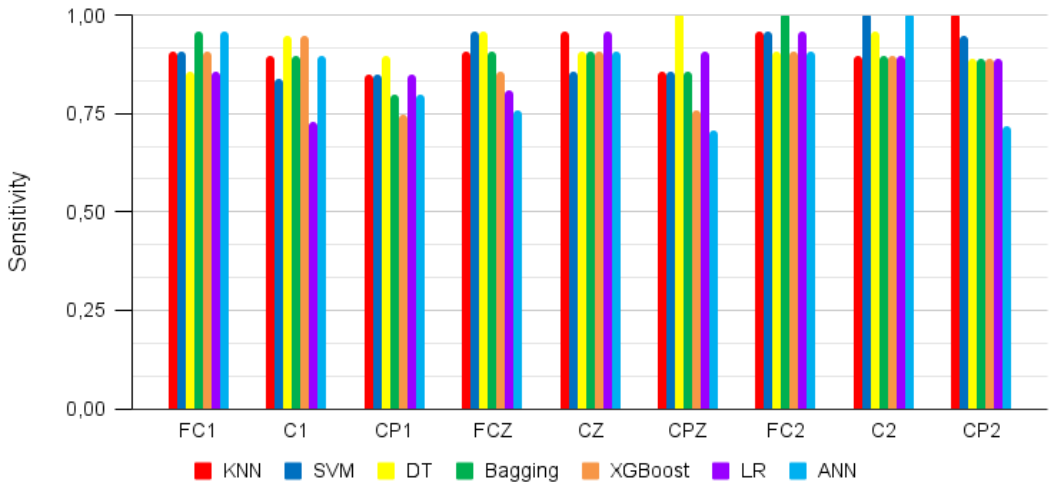


Figure 4.4 The sensitivity of the results with the highest accuracy for the EEMD_VAR method

According to Figure 4.4, the sensitivity of the results with the highest accuracy was measured as greater than 0.85 in most cases. The highest calculated sensitivity of each electrode reflects their performance for best correctly classifying subjects with PD identified as the subjects with PD. The highest sensitivity of each electrode was as follows: FC1 0.95 (Bagging and ANN); C1, 0.9444 (DT and XGBoost); CP1, 0.8947 (DT); FCZ, 0.95 (SVM and DT); CZ, 0.95 (KNN and LR); CPZ, 1.000 (DT); FC2, 1.000 (Bagging); C2, 1.000 (SVM and ANN); and CP2, 1.000 (KNN). It seems that DT was more likely to give more sensitive results. It is clear that the sensitivity obtained from the electrodes on the right side was generally higher.

Figure 4.5 shows the specificity of the results with the highest accuracy. The highest calculated specificity of each electrode reflects their performance for best correctly classifying subjects with the controls identified as the control subjects. The highest specificity of each electrode was as follows: FC1 075 (DT); C1, 0.9091 (LR and ANN); CP1, 0.90 (ANN); FCZ, 0.8333 (LR and ANN); CZ, 0.8333 (Bagging); CPZ, 0.8333 (ANN); FC2, 0.7778 (DT, XGBoost, LR and ANN); C2, 1.000 (ANN); and CP2, 1.000 (LR). The ANN often gives a better result in terms of specificity. Based on Figure 4.5, the highest specificity, 1.000, was obtained from the two right side electrodes, C2 and CP2.

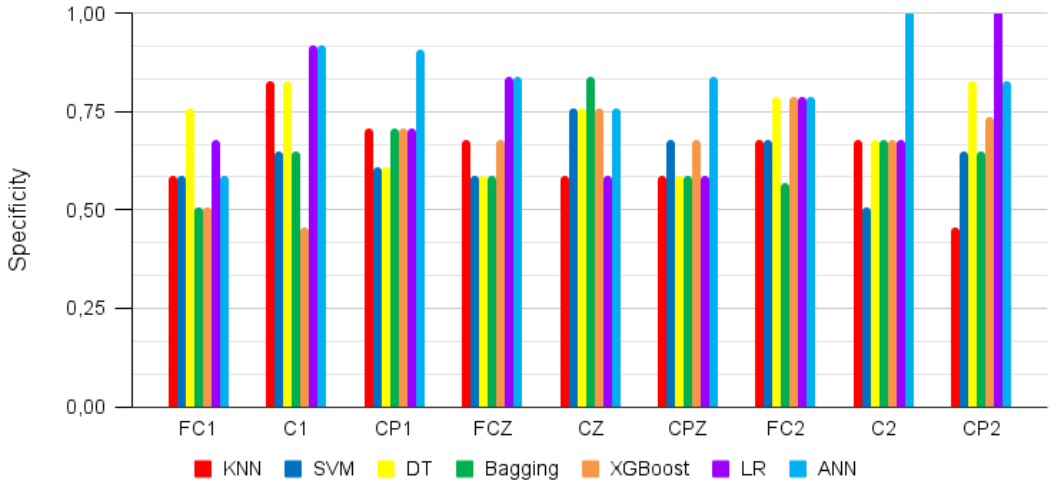


Figure 4.5 The specificity of the results with the highest accuracy for the EEMD_VAR method

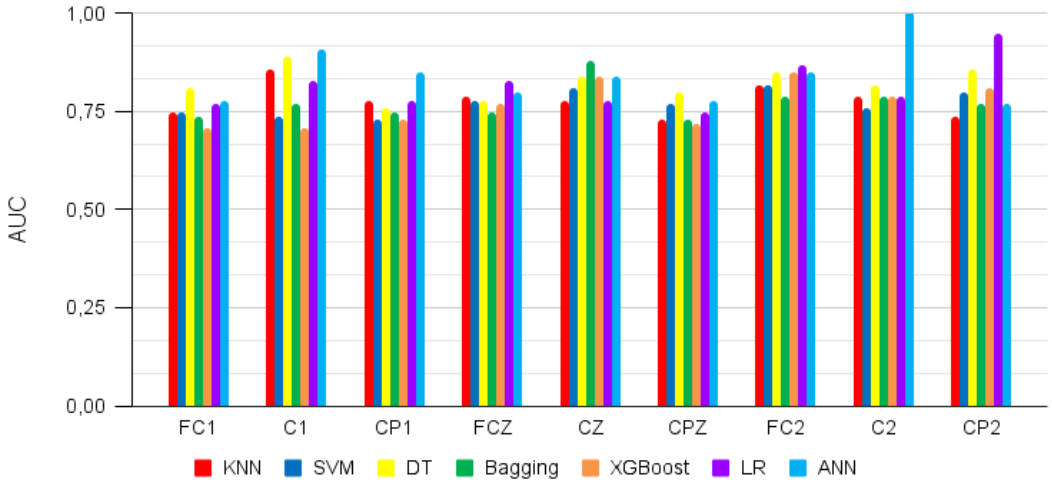


Figure 4.6 The area under the curve of the results with the highest accuracy for the EEMD_VAR method

Figure 4.6 gives the AUC of the results with the highest accuracy. The highest sensitivity of each electrode was as follows: FC1 0.80 (DT); C1, 0.90 (ANN); CP1, 0.84 (ANN); FCZ, 0.82 (LR); CZ, 0.87 (Bagging); CPZ, 0.79 (DT); FC2, 0.86 (LR); C2, 1.000 ANN); and CP2, 0.94 (LR). It seems that higher AUC values were more likely to be obtained using ANN and LR. The electrodes on the right tended to give more accurately classified subjects with an AUC.

4.2 Comparative Analysis and Discussions

We also evaluated the performance of the proposed method by comparing it to some traditional methods. The proposed method involves a process for extracting features from an EEG signal, which combines the use of EEMD and the VAR models to compute autoregressive coefficients. These coefficients are then used as the features that are extracted from the EEG signal. Additionally, feature extraction was performed using the AR model and Hjorth parameters, which are frequently used in the literature. The extracted features from each method were used as the inputs for classification algorithms. The performance metrics obtained using the AR model and the Hjorth parameters were compared with the performance metrics of the EEMD_VAR method.

Table 4.2 The accuracy values obtained using the Hjorth parameters

	KNN	SVM	DT	Bagging	XGBoost	LR	ANN
FC1	0.4688	0.3750	0.4375	0.5938	0.4063	0.5000	0.5625
C1	0.4483	0.4138	0.5172	0.3448	0.5517	0.6207	0.6207
CP1	0.5172	0.4483	0.6552	0.3793	0.6897	0.5172	0.4828
FCZ	0.5000	0.4375	0.5000	0.3750	0.4688	0.5625	0.6250
CZ	0.6563	0.6563	0.5938	0.5625	0.5625	0.6875	0.5625
CPZ	0.6250	0.5938	0.6250	0.5313	0.5313	0.6563	0.5938
FC2	0.5862	0.5172	0.5172	0.5862	0.4828	0.6207	0.6552
C2	0.6800	0.5600	0.5200	0.4800	0.6800	0.6400	**0.7200**
CP2	0.4286	0.6786	0.6071	0.4286	0.6429	0.6786	0.6071

While almost all accuracy values achieved using the EEMD_VAR method were more than 0.75 for each electrode and each classification method, the accuracy was obtained maximum 0.72 and 0.7241 by using Hjorth parameters and AR model,

respectively (see Table 4.2 and Table 4.3). The highest accuracy in distinguishing between the subjects with PD and the controls was exhibited using the ANN in C2 for Hjorth parameters and LR in FC2 for the AR model. A similar conclusion was obtained for our proposed method.

Table 4.3 The accuracy values obtained using the AR model

	KNN	SVM	DT	Bagging	XGBoost	LR	ANN
FC1	0.5313	0.5625	0.3750	0.4063	0.3750	0.5625	0.5000
C1	0.5172	0.6207	0.3793	0.5862	0.5517	0.4138	0.6897
CP1	0.5862	0.5172	0.6207	0.5172	0.6552	0.4828	0.5172
FCZ	0.6563	0.5625	0.5313	0.2813	0.5000	0.5313	0.5625
CZ	0.4375	0.5000	0.3750	0.3438	0.5000	0.5625	0.6250
CPZ	0.4063	0.5000	0.4375	0.6563	0.4375	0.5000	0.4063
FC2	0.6897	0.6897	0.5862	0.4138	0.6207	**0.7241**	0.6897
C2	0.6800	0.6400	0.6800	0.6800	0.7200	0.6800	0.6400
CP2	0.5357	0.3929	0.6071	0.5714	0.5000	0.7143	0.5000

When the sensitivity of the results with the highest accuracy for the EEMD_VAR method (Figure 4.4) is compared with the sensitivity of the methods using the Hjorth parameters (Figure 4.7-a) and the AR model (Figure 4.7-b), it is clear that the proposed method is more successful in classifying the control subjects as the controls. The sensitivity of the methods using the Hjorth parameters and the AR model supported the result that the sensitivity was higher in the right side electrodes obtained using the EEMD_VAR method.

In contrast to the specificity of the results with the highest accuracy for the EEMD_VAR method (Figure 4.5), the specificity obtained using the Hjorth parameters (Figure 4.8-a) and the AR model (Figure 4.8-b) are generally quite low.

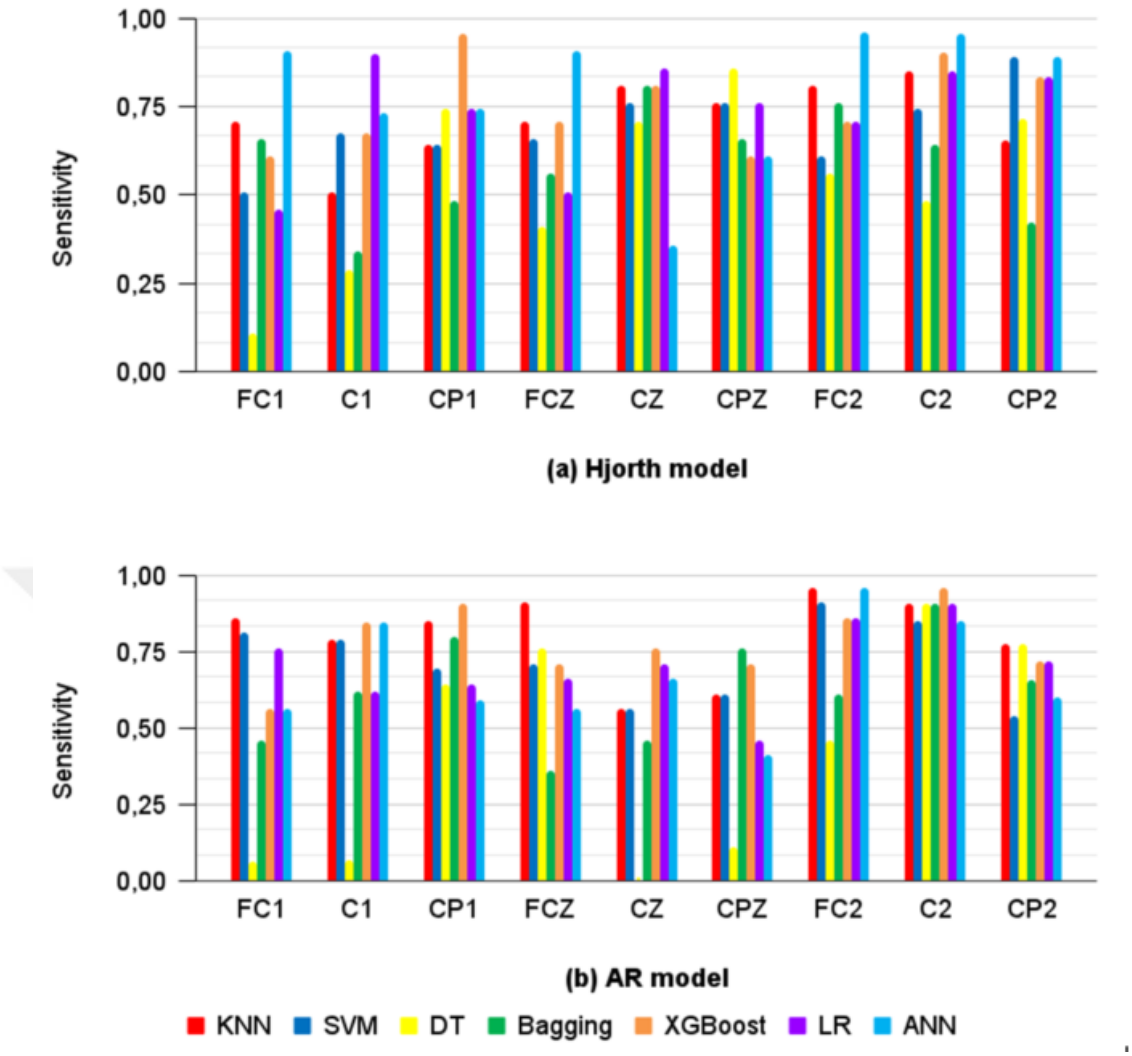


Figure 4.7 The sensitivities obtained using the Hjorth parameters and the AR model

Based on Figure 4.5, in general, the highest specificity values are given by DT stand out. Contrary to the EEMD_VAR method, the extremely low results obtained from the C2 electrode are also striking.

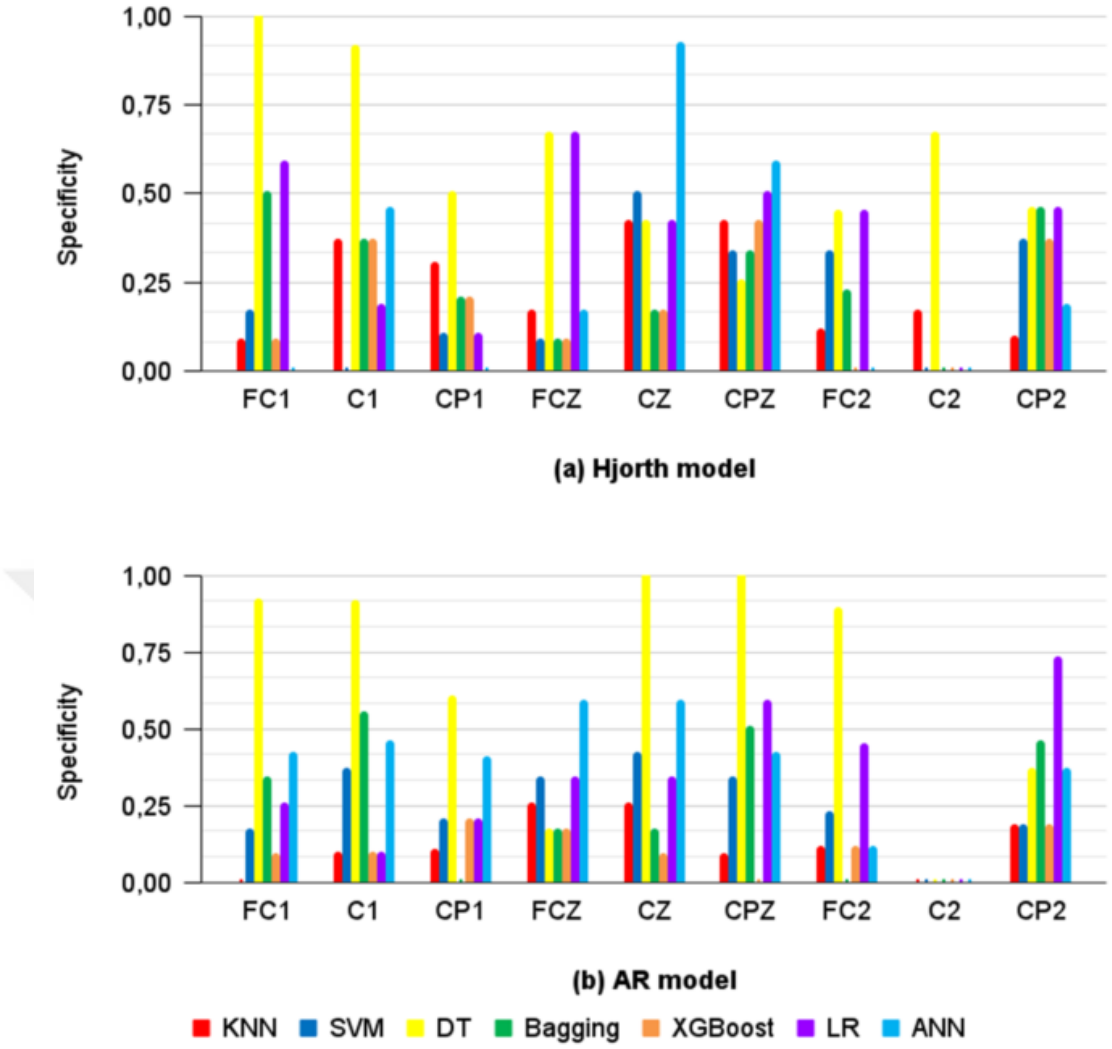


Figure 4.8 The specificity obtained using the Hjorth parameters and the AR model

For the EEMD_VAR method, the AUC values of the results with the highest accuracy are mostly above 0.75 (Figure 4.6), while maximum AUCs are 0.64 and 0.72 using the Hjorth parameters (Figure 4.9-a) and the AR model (Figure 4.9-b), respectively. This means that the EEMD_VAR method has more chance of distinguishing between the subjects with PD and the control subjects.

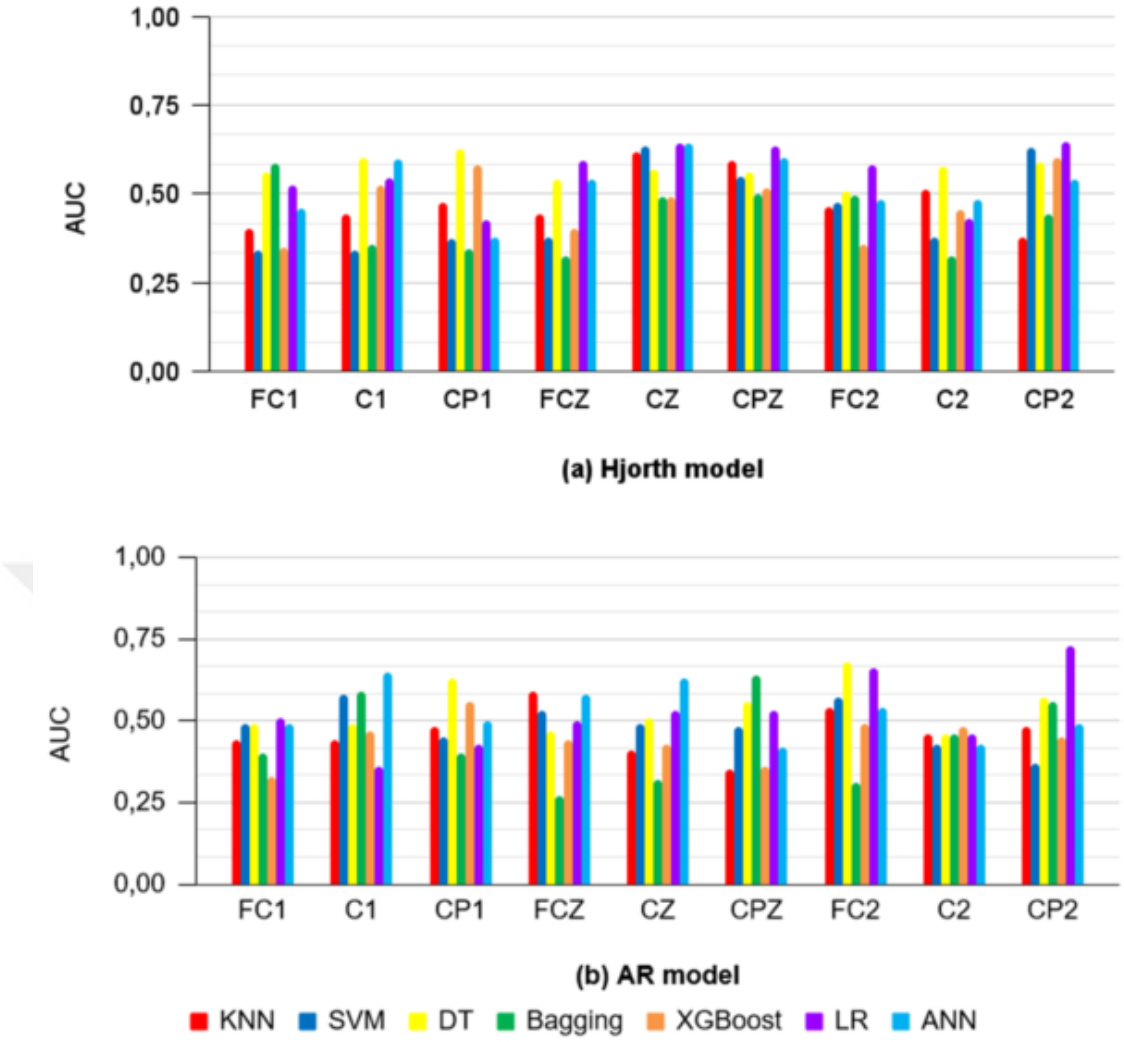


Figure 4.9 The AUC values obtained using the Hjorth parameters and the AR model

Several studies have provided statistically significant results based on the differences in the olfactory response in patients with PD and control groups. Guducu et al. (2019) used CZ and PZ electrodes of the same dataset and proposed an automated diagnostic tool based on entropy to detect chemosensory responses for identifying anosmia. Anosmic patients were correctly above 75% diagnosed using this tool. Iannilli et al. (2017),Versace et al. (2017) and Cozac et al. (2017) showed a reduction in olfactory response in patients with PD by using parametric and nonparametric statistical tests. When comparing our results to those of former studies, it must be pointed out that this study goes beyond previous studies and presents a novel approach to extracting features from olfactory EEG signals to train machine

learning algorithms for the automatic detection of PD. Another significance of our work lies in the fact that our proposed method has achieved high sensitivity, specificity, and accuracy.

Additionally, the right side superiority has been revealed for the PD classification in our study. The findings of lateralization of the human olfactory cortex were supported by Zatorre et al. (1992). In the previous studies, the right hemisphere has been associated with familiarity with odor (Royet & Plailly, 2004) and this familiarity has already been diminished in PD (Hudry et al., 2003, Tremblay et al., 2018).

CHAPTER FIVE
CONCLUSIONS AND FUTURE WORKS

In this thesis, we propose a hybrid feature extraction method for EEG signals. The KNN, SVM, DT, Bagging, XGBoost, LR and ANN are selected as classifiers to classify these features. The proposed feature extraction approach, called EEMD_VAR combines the EEMD and the VAR methods. An empirical comparison is made between the EEMD_VAR method and the Hjorth parameters and the AR model. The proposed algorithm achieves a maximum classification accuracy of 100%, but the AR method and the Hjorth parameters only achieve a maximum of 72%, demonstrating that our proposed method for feature extraction may achieve good classification performance. The results outperform overall diagnostic accuracy for the subjects with PD also according to the sensitivity, specificity, and AUC value. ANN comes into prominence among the classification algorithms according to all the performance evaluation metrics that we consider in this thesis. ANN is capable of tolerating the difficulty in the data (Pahuja et al., 2022). Hence, it is no surprise that ANN works well. Furthermore, the superiority of right-side electrodes regarding classification performance might be a reflection of impaired odor processing in PD. This idea needs to be supported in further studies with larger and different groups who have olfactory dysfunction.

The main benefits of the study are as follows:

- The proposed hybrid feature extraction method is effective at detecting PD with a high classification performance.
- The methods used for the hybrid approach are well-known and basic methods. By using this, an effective learning method is proposed for classification.
- The high-dimensional extracted features are mapped onto a new reduced feature space by PCA to make easier the classification stage. Therefore, high classification accuracies were obtained without increasing the computational complexity.

- The number of IMF (k) and lag (p) are selected between 3-10 and 1-20 to cover it from a broad perspective. 160 scenarios were examined separately and seven classifiers and nine electrodes were used to evaluate the performance of the proposed hybrid feature extraction method. his provide a comprehensive benchmark.

- The proposed hybrid feature extraction method demonstrated a higher classification accuracy than the other methods when tested using seven different classifiers. This indicates that the proposed method is highly effective at accurately classifying the data.

- The proposed approach for detecting Parkinson's disease (PD) achieved excellent classification results (100.0%) when tested on newly diagnosed patients with PD.

- The higher results from right-side electrodes have been revealed in our study.

The limitations of the proposed method are summarized as follows:

- The model was developed using a small dataset consisting of 20 PD patients and 12 healthy subjects. In the future, we plan to further test the validity of the method by applying it to a larger and more diverse dataset that includes patients in the early stages of PD. This will help to confirm the effectiveness of the method in a wider range of contexts.

- Another limitation of the EEMD_VAR method lies in the lack of a criterion for choosing the best combination of the number of IMFs (k) and lag (p) for each subject. The aim of the proposed work is to use the coefficients of the model as features. Any parameter tuning technique could not be used in this study since it proposes models with a different number of model coefficients using a number of IMF and time delays for each subject. The disparity of the number of coefficients (features) could not be involved in classifiers as it means a different number of inputs for each object.

Future research may consider the potential effects of stimulus properties (i.e. stimulus duration and intensity) on the classification performances. As a potential extension of this study, it would be interesting to test the best-performing model on a population of patients with a variety of different medical histories and health conditions. This would help to determine the generalizability of the model and its effectiveness in a more diverse range of patients. In additional work, other classification techniques, as well as other EEG signals, and other metrics obtained from the configuration of these signals can be tried. Additionally, it is possible to see many brain-computer interface applications in the literature (Masud et al., 2022, Xia et al., 2018, Mullen et al., 2015). Our high-performance method can be used in an application to allow patients with neurological disease to assess their symptoms or remotely track disease progression. Hence, identifying neurological disease in its earliest stages and assisting patients in receiving prompt, effective care can be possible.

REFERENCES

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices*, *2014*.

Alaliyat, S. (2008). Video-based fall detection in elderly's houses. Master's thesis, Gjøvik University College.

Albaqami, H., Hassan, G. M., Subasi, A., & Datta, A. (2021). Automatic detection of abnormal eeg signals using wavelet feature extraction and gradient boosting decision tree. *Biomedical Signal Processing and Control*, *70*, 102957.

Aljalal, M., Aldosari, S. A., AlSharabi, K., Abdurraqeeb, A. M., & Alturki, F. A. (2022). Parkinson's disease detection from resting-state eeg signals using common spatial pattern, entropy, and machine learning techniques. *Diagnostics*, *12*(5), 1033.

Aydemir, O. (2017). Olfactory recognition based on eeg gamma-band activity. *Neural computation*, *29*(6), 1667–1680.

Aydemir, O. (2020). Odor and subject identification using electroencephalography reaction to olfactory. *Traitement du Signal*, *37*(5), 799–805.

Barbosh, M., Singh, P., & Sadhu, A. (2020). Empirical mode decomposition and its variants: A review with applications in structural health monitoring. *Smart Materials and Structures*, *29*(9), 093001.

Beres, A. M. (2017). Time is of the essence: A review of electroencephalography (eeg) and event-related brain potentials (erps) in language research. *Applied psychophysiology and biofeedback*, *42*(4), 247–255.

Boesveldt, S., Haehner, A., Berendse, H. W., & Hummel, T. (2007). Signal-to-noise ratio of chemosensory event-related potentials. *Clinical neurophysiology*, *118*(3), 690–695.

Boonyakitanont, P., Lek-Uthai, A., Chomtho, K., & Songsiri, J. (2020). A review of feature extraction and performance evaluation in epileptic seizure detection using eeg. *Biomedical Signal Processing and Control*, *57*, 101702.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Cha, G.-W., Moon, H. J., Kim, Y.-M., Hong, W.-H., Hwang, J.-H., Park, W.-J., & Kim, Y.-C. (2020). Development of a prediction model for demolition waste generation using a random forest algorithm based on small datasets. *International Journal of Environmental Research and Public Health*, *17*(19), 6997.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, (785–794).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21-27.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, *20*(2), 215–232.

Cozac, V. V., Auschra, B., Chaturvedi, M., Gschwandtner, U., Hatz, F., Meyer, A., Welge-Lüssen, A., & Fuhr, P. (2017). Among early appearing non-motor signs of parkinson's disease, alteration of olfaction but not electroencephalographic spectrum correlates with motor function. *Frontiers in Neurology*, *8*, 545.

De Oliveira, A. P. S., De Santana, M. A., Andrade, M. K. S., Gomes, J. C., Rodrigues, M. C., & dos Santos, W. P. (2020). Early diagnosis of parkinson's disease using

eeg, machine learning and partial directed coherence. *Research on Biomedical Engineering*, *36*(3), 311–331.

Doty, R. L. (2012). Olfactory dysfunction in parkinson disease. *Nature Reviews Neurology*, *8*(6), 329–339.

Doty, R. L., Deems, D. A., & Stellar, S. (1988). Olfactory dysfunction in parkinsonism: a general deficit unrelated to neurologic signs, disease stage, or disease duration. *Neurology*, *38*(8), 1237–1237.

Fix, E., & Hodges, J. L. (1951). Discriminatory analysis, nonparametric estimation: consistency properties. *Report 4, Project nº 21-49*, *4*.

Gomathi, K., Leela, D., & Prasad, S. (2013). Smart system to recognize eeg signal for finding brain diseases using k means clustering. *International Journal of Advanced Computer Research*, *3*(4), 325.

Guducu, C. (2015). *The brain responsiveness to chemosensory stimuli in Parkinson's disease*. PhD thesis.

Guducu, C., Olcay, B. O., Schäfer, L., Aziz, M., Schriever, V., Özgören, M., & Hummel, T. (2019). Separating normosmic and anosmic patients based on entropy evaluation of olfactory event-related potentials. *Brain Research*, *1708*, 78–83.

Guducu, C., Taslica, S., Cakmur, R., Ozgoren, M., Ikiz, A. O., & Oniz, A. (2015). Assessing olfactory function in parkinson's disease via entropy analysis of chemosensory event related potentials. *The Tohoku journal of experimental medicine*, *237*(2), 111–116.

Gupta S. (2022). Why is logistic regression a classification algorithm? https://builtin.com/machine-learning/logistic-regression-classification-algorithm. Online; accessed 13-December-2022.

Hjorth, B. (1970). Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, *29*(3), 306–310.

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.

Huang, N. E. (2014). *Hilbert-Huang transform and its applications*, volume 16. World Scientific.

Huang, N. E., Shen, Z., & Long, S. R. (1999). A new view of nonlinear water waves: the hilbert spectrum. *Annual review of fluid mechanics*, *31*, 417–457.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, *454*(1971), 903–995.

Huang, N. E., Wu, M.-L. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P., & Fan, K. L. (2003). A confidence limit for the empirical mode decomposition and hilbert spectral analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *459*(2037), 2317–2345.

Huart, C., Rombaux, P., & Hummel, T. (2013). Plasticity of the human olfactory system: the olfactory bulb. *Molecules*, *18*(9), 11586–11600.

Hudry, J., Thobois, S., Broussolle, E., Adeleine, P., & Royet, J.-P. (2003). Evidence for deficiencies in perceptual and semantic olfactory processes in parkinson's disease. *Chemical senses*, *28*(6), 537–543.

Hunt, E. B., Marin, J., & Stone, P. J. (1966). Experiments in induction.

Iannilli, E., Stephan, L., Hummel, T., Reichmann, H., & Haehner, A. (2017). Olfactory impairment in parkinson's disease is a consequence of central nervous system decline. *Journal of neurology*, *264*(6), 1236–1246.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Javatpoint. Machine learning decision tree classification algorithm. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm. Online; accessed 07-December-2022.

Javatpoint. Support vector machine (svm) algorithm. https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm. Online; accessed 13-December-2022.

Khatwani, P., & Tiwari, A. (2013). A survey on different noise removal techniques of eeg signals. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(2), 1091–1095.

Lawhern, V., Hairston, W. D., McDowell, K., Westerfield, M., & Robbins, K. (2012). Detection and classification of subject-generated artifacts in eeg signals using autoregressive models. *Journal of neuroscience methods*, *208*(2), 181–189.

Leoni, J., Strada, S. C., Tanelli, M., Jiang, K., Brusa, A., & Proverbio, A. M. (2021). Automatic stimuli classification from erp data for augmented communication via brain–computer interfaces. *Expert Systems with Applications*, *184*, 115572.

Ling, C. X., Huang, J., & Zhang, H. (2003). Auc: a better measure than accuracy in comparing learning algorithms. In *Conference of the canadian society for computational studies of intelligence*, (329–341). Springer.

Maria, A. R., Bogdan, H., & Sever, P. (2018). Wavelet transform for seizures detection in eeg records. In *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, (1–6). IEEE.

Masud, U., Saeed, T., Malaikah, H. M., Islam, F. U., & Abbas, G. (2022). Smart assistive system for visually impaired people obstruction avoidance through object detection and classification. *IEEE Access*, *10*, 13428–13441.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133.

Mullen, T. R., Kothe, C. A., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., & Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using

wearable dry eeg. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2553–2567.

Nagel, S. (2019). *Towards a home-use BCI: fast asynchronous control and robust non-control state detection*. PhD thesis, Universität Tübingen.

Pahuja, S., Veer, K. et al. (2022). Recent approaches on classification and feature extraction of eeg signal: a review. *Robotica*, 1–25.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, *2*(11), 559–572.

Rahman, M. M., & Fattah, S. A. (2020). An efficient feature extraction scheme for classification of mental tasks based on inter-channel correlation in wavelet domain utilizing eeg signal. *Biomedical Signal Processing and Control*, *61*, 102033.

Raut, O. S. (2020). Single layer neural network for classification. https://www.kaggle.com/code/omkarsantoshraut/single-layer-neural-network-for-classification. Online; accessed 21-December-2022.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Royet, J.-P., & Plailly, J. (2004). Lateralization of olfactory processes. *Chemical senses*, *29*(8), 731–745.

Sadiq, M. T., Yu, X., & Yuan, Z. (2021). Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces. *Expert Systems with Applications*, *164*, 114031.

Sadiq, M. T., Yu, X., Yuan, Z., & Aziz, M. Z. (2020). Motor imagery bci classification based on novel two-dimensional modelling in empirical wavelet transform. *Electronics Letters*, *56*(25), 1367–1369.

Sadiq, M. T., Yu, X., Yuan, Z., Fan, Z., Rehman, A. U., Li, G., & Xiao, G. (2019). Motor imagery eeg signals classification based on mode amplitude and frequency components using empirical wavelet transform. *Ieee Access*, *7*, 127678–127692.

Sadiq, M. T., Yu, X., Yuan, Z., Zeming, F., Rehman, A. U., Ullah, I., Li, G., & Xiao, G. (2019). Motor imagery eeg signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces. *IEEE access*, *7*, 171431–171451.

Safi, M. S., & Safi, S. M. M. (2021). Early detection of alzheimer's disease from eeg signals using hjorth parameters. *Biomedical Signal Processing and Control*, *65*, 102338.

Saha, A., Konar, A., Chatterjee, A., Ralescu, A., & Nagar, A. K. (2014). Eeg analysis for olfactory perceptual-ability measurement using a recurrent neural classifier. *IEEE Transactions on Human-Machine Systems*, *44*(6), 717–730.

Sergeeva, M., Moiseeva, K., Kuznetsov, A., Shabakaeva, D., Smirnova, D., Khivintseva, E., Agapov, S., Bulanov, V., & Zakharov, A. (2022). Review of analytical instruments for eeg analysis. In *2022 Fourth International Conference Neurotechnologies and Neurointerfaces (CNN)*, (140–143). IEEE.

Sharma, S. (2017). Activation functions in neural networks. https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6. Online; accessed 21-December-2022.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Sörnmo, L., & Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications*, volume 8. Academic press.

Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial psychiatry journal*, *18*(1), 70.

Tremblay, C., Martel, P. D., & Frasnelli, J. (2017). Trigeminal system in parkinson's disease: A potential avenue to detect parkinson-specific olfactory dysfunction. *Parkinsonism & related disorders*, *44*, 85–90.

Tremblay, C., Martel, P. D., & Frasnelli, J. (2018). Chemosensory perception is specifically impaired in parkinson's disease. *Parkinsonism & Related Disorders*, *57*, 68–71.

Tuncer, T., Dogan, S., & Acharya, U. R. (2020). Automated detection of parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels. *Biocybernetics and Biomedical Engineering*, *40*(1), 211–220.

Tuncer, T., Dogan, S., & Subasi, A. (2021). Eeg-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection. *Biomedical Signal Processing and Control*, *68*, 102591.

Uktveris, T., & Jusas, V. (2017). Application of convolutional neural networks to four-class motor imagery classification problem. *Information Technology and Control*, *46*(2), 260–273.

Uyulan, C., & Erguzel, T. T. (2017). Analysis of time–frequency eeg feature extraction methods for mental task classification. *International Journal of Computational Intelligence Systems*, *10*(1), 1280.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Vasiliev, V., Velmakina, Y., & Mayborodin, A. B. (2019). Using artificial neural networks when integrating the requirements of standards for management systems in qms. *IOP Conference Series: Materials Science and Engineering*, *666*(1), 012058.

Versace, V., Langthaler, P. B., Sebastianelli, L., Höller, Y., Brigo, F., Orioli, A., Saltuari, L., & Nardone, R. (2017). Impaired cholinergic transmission in patients with parkinson's disease and olfactory dysfunction. *Journal of the Neurological Sciences*, *377*, 55–61.

Whitcroft, K., Aziz, M., Croy, I., Schriever, V., & Hummel, T. (2017). Short inter-stimulus intervals can be used for olfactory electroencephalography in patients of varying olfactory function. *Neuroscience*, *363*, 26–33.

Wikipedia (2022a). Boosting (machine learning). Online; accessed 09-December-2022.

Wikipedia (2022b). Bootstrap aggregating. Online; accessed 09-December-2022.

Wikipedia (2022c). Partial area under the roc curve. Online; accessed 13-December-2022.

Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, *1*(01), 1–41.

Xia, Y., Zhang, H., Xu, L., Gao, Z., Zhang, H., Liu, H., & Li, S. (2018). An automatic cardiac arrhythmia classification system with wearable electrocardiogram. *IEEE Access*, *6*, 16529–16538.

Yule, G. U. (1927). Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *226*(636-646), 267–298.

Zatorre, R. J., Jones-Gotman, M., Evans, A. C., & Meyer, E. (1992). Functional localization and lateralization of human olfactory cortex. *Nature*, *360*(6402), 339–340.

Zhang, G. (2018). What is the kernel trick? why is it important? https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d. Online; accessed 07-December-2022.

Zhang, Y., Chen, W., Lin, C.-L., Pei, Z., Chen, J., & Chen, Z. (2021). Boosting-lda algriothm with multi-domain feature fusion for motor imagery eeg decoding. *Biomedical Signal Processing and Control*, *70*, 102983.

Zhang, Y., Zhang, S., & Ji, X. (2018). Eeg-based classification of emotions using empirical mode decomposition and autoregressive model. *Multimedia Tools and Applications*, *77*(20), 26697–26710.