Early Detection of Neurodegenerative Diseases from Bio-Signals: A Machine Learning Approach

By

Shamaila Iram
BSc, MSc

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the Degree of Doctor of Philosophy

June 2014

Acknowledgements

At first, all praise is due to Allah (s.w.t) for granting me the ability, strength and determination to complete this PhD. I am profoundly grateful to my supervisor, Dr. Dhiya Al-jumeily for offering me a PhD scholarship to start my research career at Liverpool John Moores University (LJMU) and also for encouraging and assisting me throughout the PhD. I would like to avail this opportunity to appreciate his support and his advices that helped me to face the challenges more successfully. I think myself fortunate enough to be one of his students.

I am also deeply obliged to Dr Paul Fergus for his precious time that he dedicated to my research work and discussion on ideas that eventually became a part of this thesis. I am so grateful to him for polishing my writing skills that ended up publishing this research work to well-known conferences and journals. Also, very special thank goes to Dr. Abir Husain for reviewing my work and giving me valuable feedback. My gratitude also extends to The Director of School of Computing and Mathematical Sciences, Professor Majid Merabti for his support through some of the issues during PhD.

During PhD, a biggest challenge was the collection of EEG signals for Alzheimer's patients to complete this research work. Here, I would like to express my heartfelt gratitude to Professor Francois Vialatte for providing me a prestigious opportunity to work at ESPCI ParisTech SIGMA laboratory (France) to analyse EEG Signals. He deserves much more respect and gratitude for helping me; during French visa application process, sorting out accommodation problems and then finally discussing the new ideas and approaches that could be implemented to accomplish this research work. Also, those hours and hours of work and discussions in the lab for debugging the code to get better results are totally unforgettable. Thanks to Dr. Gérard Dreyfus, Antoine Gaume, Dr. Mannual, Parvaneh Adibpour for making my stay more fruitful and pleasurable.

Also, I would like to express my deepest gratitude to my parents and other family members for their prayers and support which enabled me to successfully complete my PhD. I can never say enough thanks to my brother, Muhammad Irfan Qamar who motivated me and supported me during the whole time span of PhD. I am so grateful to my brother, Muhammad Rizwan Iqbal, for his financial support.

I would also like to thanks Dr Muhammad Zahid Khan, Dr Nargis Bibi, Laura Beltran for their friendly advices that helped me to stay positive and more focused. Special thanks go to the staff and technicians of the School of Computing and Mathematical Sciences, Liverpool John Moores University, for their assistance and providing me with a pleasant work environment along with excellent computing facilities.

Abstract

Given the fact that people, especially in advanced countries, are living longer due to the advancements in medical sciences which resulted in the prevalence of age-related diseases like Alzheimer's and dementia. The occurrence of such diseases continues to increase and ultimately the cost of caring for these groups will become unsustainable. Addressing this issue has reached a critical point and failing to provide a strategic way forward will negatively affect patients, national health services and society as a whole.

Three distinctive development stages of neurodegenerative diseases (Retrogenesis, Cognitive Impairment and Gait Impairment) motivated me to divide this research work into two main parts. To fully achieve the purpose of early detection/diagnosis, I aimed at analysing the gait signals as well as EEG signals, separately, as both of these signals severely get affected by any neurological disease.

The first part of this research work focuses on the *discrimination analysis* of gait signals of different neurodegenerative diseases (Parkinson's, Huntington, and Amyotrophic Lateral Sclerosis) and also of control subjects. This involves relevant feature extraction, solving the issues of imbalanced datasets and missing entries and lastly classification of multiclass datasets. For the classification and discrimination of gait signals, eleven (11) classifiers are selected representing linear, non-linear and Bayes normal classification techniques. Results revealed that three classifiers have provided us with higher accuracy rate which are *UDC*, *LDC* and *PARZEN* with *65%, 62.5%* and *60%* accuracy, respectively. Further, I proposed and developed a new *classifier fusion strategy* that combined classification algorithms with combining rules (voting, product, mean, median, maximum and minimum). It generates better results and classifies subjects more accurately than *base-level* classifiers.

The last part of this research work is based on the rectification and computation of EEG signals of mild Alzheimer's disease patients and control subjects. To detect the perturbation in EEG signals of Alzheimer's patients, three neural synchrony measurement techniques; *phase synchrony*, *magnitude squared coherence* and *cross correlation* are applied on three different databases of mild Alzheimer's disease (MiAD) patients and healthy subjects. I have compared right and left temporal parts of brain with rest of the brain area (frontal, central and occipital), as temporal regions are relatively the first ones to be affected by Alzheimer's. Two novel methods are proposed to compute the neural synchronization of the brain; *Average synchrony measure* and *PCA based synchrony measure*. These techniques are evaluated for three different datasets of MiAD patients and control subjects using the *Wilcoxon ranksum test (Mann-Whitney U test)*. Results demonstrated that PCA based method helped us to find more significant features that can be used as biomarkers for the early diagnosis of Alzheimer's.

Publications

Journal Papers

Shamaila Iram, Paul Fergus, Dhiya Al-Jumeily, Abir Hussain, Martin Randles, *A Classifier Fusion Strategy to Improve the Early Detection of Neurodegenerative Diseases*. Accepted for publication in the International Journal of Artificial Intelligence and Soft Computing, 2013

William Hurst, Madjid Merabti, **Shamaila Iram**, Paul Fergus, "*Protecting Critical Infrastructures Through Behavioural Observation*", Accepted for publication in the International Journal of Critical Infrastructures 2013.

Paul Fergus, Pauline Cheung, Abir Hussain, Dhyia Al-Jumeily, **Shamaila Irma** *"Prediction of Preterm Deliveries from EHG Signals using Machine Learning" PLOS ONE, vol. 8, p. e77154, 2013.*

Paul Fergus, **Shamaila Iram**, Dhiya Al-Jumeily, Martin Randles, Andrew Attwood, *Home-based Health Monitoring and Measurement for Personalised Healthcare*, Journal of Medical Imaging and Health Informatics (JMIHI), vol. 2(1), pp. 35-43, March, 2012.

Shamaila Iram, Dhiya Al-Jumeily, Francois Vialatte, Paul Fergus "*Early Detection of Alzheimer's Disease from Multi-Channel EEG Signals by Means of Neural Synchrony Measurement Techniques in Optimize Frequency Bands"* (In Preparation)

Conference Papers

Shamaila Iram, Dhiya Al-Jumeily, Abir Hussain, Paul Fergus, and David Lamb, *"On the Early Detection of Neurodegenerative Diseases from Gait and EEG Signals: A Machine Learning Approach",* BCS International IT Conference, Abu Dhabi, UAE, 2013.

Shamaila Iram, Dhiya Al-Jumeily, Paul Fergus, Martin Randles, Abir Hussain, *"Computational Data Analysis for Movement Signals Based on Statistical Pattern Recognition Techniques for Neurodegenerative Diseases",* The 13th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, 25th - 26th June, Liverpool, UK, 2012.

Shamaila Iram, Dhiya Al-Jumeily, Paul Fergus, Martin Randles*, "E-Health: The Potential of Linked Data and Stream Reasoning for Personalised Healthcare"*, in Developments in E-systems Engineering (DeSE), IEEE Society Dubai, 6th - 8th December 2011.

Shamaila Iram, D. Al-Jumeily, and J. Lunn, *"An Integrated Web-Based e-Assessment Tool"* in Developments in E-systems Engineering (DeSE), IEEE Society, 2011, pp. 271-275

Dhiya Al-Jumeily, **Shamaila Iram**, Francois Vialatte Paul Fergus, "A Novel Method to Analyze EEG Synchrony for the Early Diagnosis of Alzheimer's Disease in Optimized Frequency Bands", accepted for publication in CCNC 2014 - Consumer eHealth Platforms Services and Applications-America - CeHPSA 2014

Shamaila Iram, Dhiya Al-Jumeily, Paul Fergus, Abir Hussain, "Exploring the Hidden Challenges Associated with the Evaluation of Multi-class Datasets using Multiple Classifiers", International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), IEEE, 2014

Table of Contents

List of Tables

List of Figures

Abbreviations

AD, PD, HD	Alzheimer's Disease, Parkinson's Disease, Huntington's Disease,
ALS, CO	Amyotrophic Lateral Sclerosis, Control subjects
ANN	Artificial Neural Network
BPNN	Back propagation neural network
BSS	Blind Source Separation
CSP	Common Spatial Patterns
DCT	Discrete Cosine Transform
EEG	Electroencephalogram
GRNN	General Regression Neural Network
LDC	Linear Discriminant Classifier
Loglc	Logistic Linear Classifier
MCI	Mild Cognitive Impairment
MiAD	Mild Alzheimer's Disease
MLP	Multi-Layer Perceptron
MMSE	Mini-Mental Score Evaluation
MSC	Magnitude Squared Coherence
NB	Naïve Bayesian
NDDs	Neurodegenerative Diseases
NMC	Nearest Mean Classifier
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
polyc	Polynomial Classifier
QDC	Quadratic Discriminant Classifier
RBFNN	Radial basis function neural network
ROC	Receiver Operating Characteristic
SVC	Support Vector Classifier
treec	Binary Decision Tree Classifier
UDC	Uncorrelated Normal Density based Classifier
xcorr	Cross Correlation

Chapter 1 Introduction

1.1 Background

Advancements in machine learning provoke new challenges by integrating data mining with biomedical sciences in the area of computer science. This emergent research line provides a multidisciplinary approach to combine engineering, mathematical analysis, computational simulation, and neuro-computing to solve complex problems in medical science. One of the most significant applications of machine learning is data mining. Data mining provides a solution to find out the relationships between multiple features, ultimately, improving the efficiency of systems and designs of the machines. Data mining techniques provide computer based information systems to find out data patterns, generate information for the hidden relationships and discover knowledge that unveils significant findings that cannot be accessible by traditional computer based systems.

The types of machine learning can be supervised, unsupervised or reinforced. In supervised learning, the labels for each class are provided for the classifier at the training stage. In unsupervised learning, also known as clustering, the class labels are not known, but the classifier is asked to group the instances into groups where they display the same pattern of features, and hence each cluster may represent one class. In reinforcement learning, the classifier makes a classification of each instance and is given a score after each classification, to reflect how well it classified the instance. The classifier then adjust its future actions accordingly [1].

Neurodegenerative diseases (NDDs) are accompanied by the deterioration of functional neurons in the central nervous system. These include Parkinson's, Alzheimer's, Huntington's, and Amyotrophic Lateral Sclerosis among others. The progression of these diseases can be divided into three well recognized stages; retrogenesis, cognitive impairment and gait impairment. Retrogenesis is the initial stage of any NDD which starts with the malfunctioning of cholinergic system of basal fore brain that further extends to Entorhinal Cortex and Hippocampus [2]. As a result of retrogenesis, the memory of the patients severely affected due to the accumulation of pathological neurofibrillary plaques and tangles in the entorhinal cortex (EC), hippocampus, caudate, substantia nigra parts of the brain [3]. This stage is known as "Cognitive Impairment". Finally, a patient cannot maintain his/her healthy, normal gait due to disturbances in cortico-cortical and cortico-subcortical connections in the brain, e.g., frontal connection with parietal lobes and frontal lobes with basal ganglia, respectively [4].

Early detection/diagnosis of life threatening and irreversible diseases such as neurodegenerative diseases (Alzheimer's, Parkinson's, Huntington's, and Amyotrophic Lateral Sclerosis) is an area of great interest for researchers from different academic backgrounds. Diagnosing NDDs at earlier stage is hard where symptoms are often dismissed as normal consequences of aging. Moreover, the situation becomes more challenging where the symptoms or data patterns of different NDDs turn out to be similar and discrimination among these diseases becomes as crucial as the treatment itself. In this research work, we claim the significance of analysing gait signals for discriminating movement disorders in different NDDs for accurate diagnosis and in time treatment of the patients as well as the early diagnosis of these diseases using EEG signals.

One significant tool for the discrimination of different NDDs is "Gait Signals". Hausdorff *et al* [5] suggested that the understanding of relationship between loss of motor neurons and the perturbation in the stability of stride-to-stride dynamics can help us to monitor neurodegenerative diseases progression and in assessing potential therapeutic interventions. Furthermore, they claimed a reduced stride-interval correlation with aging in Huntington's disease. Later in 2010, the same gait signals are used by Yunfeng and Krishnan [6] to estimate the probability density functions (PDFs) of stride intervals and its two sub-phases

for Parkinson's disease. Moreover, Masood *et al.* used the same dataset of gait signals for the discrimination of different NDDs [7].

The other important tool for the early diagnosis and treatment of neurological and neurodegenerative diseases is the electroencephalogram (EEG) signals. Hans Berger for the first time measured the first EEG in humans and still nowadays EEG is extensively used to evaluate neurological diseases [8]. The EEG signals originate from the cerebral cortex and evoke by auditory and somatosensory stimuli.

The goal, here, is to implement a data mining approach with innovative ideas, using gait and EEG signals as a discriminative and diagnostic tool, to design a diagnostic and therapeutic system for the early diagnosis of life threatening diseases such as Alzheimer's, Parkinson's, Huntington and Amyotrophic lateral sclerosis, etc.

1.2 Problem Statement/Motivations

This section provides a detail ensight of the challenges and the problems need to be looked into, from two different perspectives—challenges with the early diagnosis of NDDs and issues related to machine learning.

1.2.1 Issues with the Early Diagnosis of NDDs

Neurodegenerative diseases, especially Alzheimer's disease is the most prevalent form of dementia and according to statistics, 5-10% of the population is affected by these diseases above the age of 65 [9]. The clinical symptoms of the disease are characterized by progressive amnesia, linking it with the continuous and gradual loss of cognitive power and finally, paralyzing the person by affecting the motor neurons. NDDs triggered by the deterioration of neuronal cells due to accumulation of neurofibrillary tangles and pathological proteins (such as α-synuclein, tau proteins etc.) and also the senile plaques in cortio-cortical and cortio-sub cortical parts of the brain [3]. Since the occurrence of memory loss can be related to one of the aging factors, the ability to predict or diagnose a NDD turns out to be impossible at an earlier stage.

Advances in medicine and healthier lifestyle choices are allowing people to live longer [10]. However, as this shift continues, so does an increase in age-related neurodegenerative diseases, such as Alzheimer's and dementia [11]. Currently, treating neurodegenerative diseases, places considerable pressure on national healthcare systems [12]. Many believe that significant increases will be unsustainable [13]. The solution is not obvious; however, approaches centred on early detection, and management are likely to yield some interesting results. Nonetheless, early detection of neurodegenerative diseases is still a major unresolved and significant area of concern for national healthcare services, globally. Neurodegenerative diseases are one of the leading causes of death, even in developed countries. A progressive central nervous system disorder leads towards severe neurodegenerative diseases like Alzheimer's, Parkinson's, Huntington's, and Amyotrophic Lateral Sclerosis. Due to the insidious onset and gradual progression of pathological changes, it is crucial to divide the evolution of neurodegenerative diseases into different stages in order to detect their symptoms earlier.

One possible approach is to build on the advances made in e-Health systems to improve the detection, diagnoses and treatment of such diseases to support disease management and integrated care strategies [14]. This will allow physicians to incorporate information and communication technologies into the decision-making process to enhance the diagnosis of such diseases and inform treatment strategies [15]. The research agenda is timely, given that conclusive diagnosis of these diseases is currently only possible posthumously, by direct examination of the affected brain tissue after the death of a patient [16]. Compounding the problem further, obvious symptoms of neurodegenerative diseases are only visible during the advanced stages of the illness (i.e., gait impairment) when no possible cure is available. This often leaves the patient in a miserable condition awaiting his or her death. Clearly, new approaches are required to detect the early onset of symptoms associated with such diseases to either prevent or mitigate disease progression [17].

At a basic level, it allows individual patterns or features within the data to be explicitly associated with particular diseases. For example, abnormal and chaotic body movements caused by damage to neurons can be associated with Huntington's disease. At a more advanced level, the similarities between different neurodegenerative diseases need to be clearly defined. This will allow a patient's unique needs to be considered when deciding on

appropriate treatments. While, having similar characteristics, different neurological diseases cause atrophy in different parts of the brain; Huntington's disease causes damage to the caudate, Parkinson's damages the substantia nigra, and Amyotrophic lateral sclerosis (ALS) damages the lower motor and pyramidal neurons, resulting in severe damage to body movement. Furthermore, there is a need to take into account other features directly related to diseases such as age, gender and so on. Clearly, focusing on a single correlation is unlikely to identify a particular neurodegenerative disease. Solutions that are designed to make correlations between multiple patterns or features within the data are likely to be particularly effective in identifying specific neurodegenerative diseases.

The loss of cognitive power is generally associated with a decrease of functional synchronization of different parts of the brain. Hence, loss of functional interaction between cortical areas could be considered a possible symptom of any NDD. Finding the synchronization in terms of coherence and correlation can possibly provide significant information in the early diagnosis of NDDs. However, the compactness of EEG signals because of different frequency bands makes this task less straightforward. More research is still required to find out the exact role of each frequency band in the early diagnosis of these diseases.

1.2.2 Issues related to Machine Learning Approach

In the context of machine learning, data mining offers many challenges that needed to be considered to get optimal results from a classifier. These factors can affect the mining process in terms of computation time, extraction and selection of appropriate features and implementation of new approaches that can help us to get expected results. This section, briefly states those challenges:

- **Skewed datasets:** Learning of a classifier from an imbalanced datasets usually generates biased results. In this case, a classifier becomes more sensitive (highly trained) for the majority class and less sensitive (less trained) for minority class. Ultimately, the results obtained from such classification make the situation more complicated, especially, when the data is being processed from a real time

environment—biomedical, genetics, radar signals, intrusion detection, risk management and credit card scoring [18].

- **Handling missing data:** There are many reasons behind missing entries in a dataset. A damage in the remote sensor network, failure of gene microarray to yield gene expression, finger prints, dust or manufacturing defects, missing applicable tests while diagnosing patients, can be some of the reasons of missing entries in a dataset, as described by Marlin in [19]. Feature extraction and feature classification based on these datasets that lead to unreliable results. Problems of missing data should be incorporated before starting a computation process.
- **Multiclass datasets:** The problem of skewed datasets becomes even more complicated when it comes to multiclass datasets. Practically speaking, in real world environments, mostly the datasets come from a multiclass domain, for instance, protein fold classification [20]. These multiclass datasets pose new challenges as compared to simple two-class problems. Zhou *et al.* in their paper [21] argue that handling multiclass datasets is much harder than handling two-class problem domains. Jeopardizing the problem further, almost all classifier evaluation techniques are designed for two-class problems and become unfit for multiclass problems.
- **Extraction and selection of relevant features:** The accuracy of a classifier is directly dependent on the variables that are provided for the classification. The analysis of gait as well as EEG signals and extraction of relevant information is not an easy task. Gait signals may be contaminated with other muscle movement signals or by the environmental data. Similarly, EEG signals may contain the signals of eye movements or externally generated signals (power line, electrode movement, etc.). In the presence of these artifacts, discrimination or classification leads to wrong results. This problem motivates some *preprocessing* steps to get clean signals before classification. Once the signals are extracted, the next challenge is the selection of the most relevant features among others. This not only saves computation time but also reduces the complexity of the system.
- **Data Filtering:** Previous studies focus on the analysis of compact EEG signals without filtering them into narrow frequency bands. This does not provide optimal information about the frequency band which is more important in detecting Alzheimer's (or other NDDs) at its earlier stage.

- **Selection of a Classifier:** In the field of pattern recognition, the main focus is the successful classification of the features with the maximum possible accuracy rate. A classifier with a specific set of features may or may not be an appropriate option for another set of features. Moreover, different classification algorithms achieve different degrees of success for different kinds of applications [22]. In this case, selection of an appropriate classifier becomes a challenging task. Indeed, further research is still required to generalize the performance of the classifiers.

1.3 Aims and Objectives

The main aim of this research work is to provide an early diagnostic and therapeutic system for NDDs using machine learning techniques. More precisely, it focuses on the study of gait as well as EEG signal processing and classification techniques to propose new methods that can help clinicians for the early diagnosis of these diseases.

Following are the main objectives to achieve this aim:

- Collection of gait signals for different neurodegenerative diseases such as Parkinson's, Hartington's, and Amyotrophic lateral sclerosis and also for control subjects;
- Removing artifacts and handling missing entries of datasets and addressing the issues with imbalanced datasets and proposing oversampling and under-sampling methods to handle this;
- Classification and evaluation of multiclass datasets using *PRTools* (a *Matlab* integrated pattern recognition tool) using a set of linear, nonlinear and Bayes normal classifiers;
- Performance evaluation of classifiers from two different perspectives, *i.e.,* Visualization Techniques (ROC analysis and Reject Curves) and Statistical techniques (Confusion Matrix, Precision, Recall, Sensitivity, specificity and F-Measure);
- Collection of EEG signals of Alzheimer's patients and control subjects using 32-channel electrodes and filtering of EEG signals into different *optimized frequency bands* to improve the detection accuracy;

- Compute the neural synchronization of Alzheimer's patients and healthy persons by using *neural synchrony* measurement *techniques (Phase synchrony, Coherence, and Correlation);*
- Propose a novel method to compute the *synchronization* of neuronal activities using *Principal Component Analysis (PCA).* Also, measure and compare the perturbation in the functional brain activities of Alzheimer's patients with the neuronal activities of control subjects by applying the *Wilcoxon rank-sum test;*
- Develop and implement the *Gram-Schmidt orthogonalization* process for constructing an orthogonal basis for a Euclidean space to find out the best features that can act as *biomarkers* for the early detection of Alzheimer's

1.4 Novel Contributions

Based on the developmental stages of NDDs, this research work has been divided into two parts. Part I is based on the discrimination analysis of gait signals of different NDD patients while Part II represents the analysis of EEG signals of Alzheimer's patients as well as control subjects. Novel contributions of the research work are presented below:

Part I-Discrimination of gait signals: during the analysis of gait signals of different NDDs and control subjects, we propose the following methods to handle the challenges that are described in section 1.2:

- The possible solutions of imbalanced datasets are provided in terms of *under-sampling* and *over-sampling*. We intend to produce more pseudo-data to solve the problem of getting biased results due to different numbers of entries in each set.
- In our research work, we intend to select all significant features that have direct or indirect impact on the progression of NDDs. For instance, age, gender, weight of the subjects, BMI factor and exact level of severity of the disease are important features that are being neglected in the available literature.
- For the discrimination of gait signals of different NDDs that arose by similar causes, we intend to implement a wide variety of classifiers. They belong to linear, non-linear and Bayes normal classifiers.

- Due to multiclass datasets, it is impossible to check the accuracy of the classifier using a single evaluation technique. It is intended to represent results using different evaluation techniques—*visualization* and *statistical* techniques.
- In this research work, I intend to propose a *classifier fusion strategy* from *base-level* classifiers to get higher accuracy rates and to resolve the ambiguity of selection of a classifier, for a particular dataset.

Part II-Computation of EEG signals: this part is based on the rectification of EEG signals to find out those significant features that can help clinicians to diagnose Alzheimer's at earlier stage. Following are the novel contributions of our research work for processing the EEG signals:

- In order to find out those hidden patterns that are usually neglected during compact EEG processing, we intend to divide each EEG signal from each channel into five narrow frequency bands – delta (δ), theta (θ), alpha (α), beta (β) and gamma (γ). After that, each data band is used to measure the synchronization in each part of the brain. These frequency bands can help to extract the most significant features that can later be used as biomarkers for the early diagnosis of Alzheimer's.
- We propose two novel methods to compute EEG signals; one is called *PCA based synchrony measure* while the other is called *Average synchrony measure*. In the PCA based method, we intend to eliminate all redundant information that can be a base of providing biased results. The results of these methods are compared using the *Wilcoxon rank-sum test.*

Part III-Benefits of Computing Gait and EEG Signals: In this research work, two different kinds of signals are computed – Gait and EEG signals. Gait signals are particularly used for the exact diagnosis of a specific NDD for the accurate and in time treatment of the patients. For instance, diagnosis time is of vital importance in the treatment of a disease especially for chronic diseases. The main challenge with NDDs is that they all pose the same symptoms at the final stage—gait disorder. At this stage, it is very difficult to discriminate a specific neurodegenerative disease with a non-invasive method. Automatic classification of gait patterns by statistical pattern recognition techniques will help to solve this problem by discriminating different NDDs according to their data patterns. This will save the time of

doctors/practitioners not only for diagnosing exact NDDs but also their timely treatment of the disease.

On the other hand, EEG signals provide significant information for the early detection of NDDs. EEG generates by bioelectric phenomenon that is stimulated from the cerebral cortex by auditory and somatosensory stimuli is further investigated, explored and interpreted to understand the brain functionality and to identify different pathologies. This process particularly helps to find out those diseases that are impossible to be cured by medication, for instance, the neurodegenerative diseases. This phenomenon also helps for the early detection of life threatening diseases and widely used in the clinical studies.

1.5 Thesis Organization

This thesis is organized as follows:

Introduction (Chapter 1): this chapter highlights the importance of diagnosing NDDs at earlier stage and also presents the potential of the machine learning approach to achieve this target. In addition, the issues related to early diagnosis are discussed in detail. This chapter clearly advocates the development of new approaches and methods in machine learning to get optimal benefits from it, in the area of computer science.

Background (Chapter 2): this chapter elaborates different types of NDDs, their developmental stages and their symptoms at each stage. Moreover, a background of signal classification and signal processing is presented in it. It highlights the differences between supervised and unsupervised machine learning. Furthermore, different types of supervised machine learning techniques are discussed in detail.

A Strategic Framework for the Early Detection of NDDs (Chapter 3): this chapter presents an explanation of each module of the proposed framework, from data collection to its processing and then concluding the final results. Moreover, it gives an idea of the tools and techniques that are used to complete this research work.

Assessment of Gait Dynamics (Chapter 4): this chapter demonstrates the assessment of gait signals of different NDDs and control subjects. It also presents the possible solutions of

skewed datasets, missing data entries, multiclass pattern recognition, and discrimination among similar diseases. Moreover, it discusses eleven different classification techniques for discriminating the gait signals of NDDs and control subjects. It also presents various performance evaluation techniques to measure the accuracy of each classifier.

Classifier Fusion Strategy (Chapter 5): it presents our novel idea to combine *base-level* classifiers to achieve higher accuracy rate. Moreover, it highlights six different combining rules (product, maximum, minimum, mean, median and voting) and their importance in combining the classifiers. The main purpose is to check if the new approach shows superior performance compared to the *stand-alone* classifiers. Instead of looking for better classifiers and more appropriate set of features, this chapter provides an insight of looking at the best set of classifiers and the best combination method.

Neural Synchrony Measurement (Chapter 6): this chapter demonstrates the significance of computing EEG signals to measure neural synchronization of Alzheimer's disease and control subjects. It presents our novel methods to apply three neural synchrony measurement techniques (*phase synchrony, cross correlation and MS coherence*) on three different datasets of EEG signals. These methods are; *PCA based synchrony measure* and *Average synchrony measure*. The *Wilcoxon ranksum (Mann-Whitney) test* is used to compare the results of these two methods. Later, *Gram Schmidt orthogonalization* is applied with the *"n-probe"* function to get the most important features that can help clinicians for the classification and also for the early diagnosis of AD.

Conclusions and Future Work (Chapter 7): this chapter provides the summary and conclusions extracted from the whole research work. It also highlights the novel contributions and the limitations of the work and discusses some probable future directions.

Chapter 2 Neurodegenerative Diseases and Machine Learning

2.1 Introduction

This chapter provides a brief overview of neurodegenerative diseases, their development stages and importance of their early detection. It also gives a detailed insight of signal processing in terms of feature extraction, feature selection and feature classification. A brief description of gait and EEG signals is also provided because of the major role they play in the early detection/diagnosis of Alzheimer's and other neurodegenerative diseases. This chapter also presents a great deal of information about different kinds of classification algorithms. As such, it details the different processing steps of the early detection of neurodegenerative diseases, that is, measurements of gait and brain activity, preprocessing, feature extraction and classification.

2.2 Neurodegenerative Diseases (NDDs)

Neurodegenerative diseases is an umbrella term used to describe medical conditions that directly affect the neurons within the brain [23]. These include Parkinson's, Alzheimer's, Huntington's, and Amyotrophic Lateral Sclerosis among others. Patients suffering with these kinds of disease, experience a cognitive decline over a long period and symptoms include gait abnormalities; problems with speech, and memory loss due to progressive cognitive deterioration [24]. Given the fact that people are now living longer, neurodegenerative

diseases have become more prevalent in developed countries and this is placing a major economic burden on health care services. For example, in 2005 it was estimated to cost USD 315 billion to treat the 29.3 million people suffering with dementia [25]. In 2009, it was estimated to cost USD 422 billion to treat 34 million people with dementia around the world [26]. As the disease progresses the patient starts acting like a child because the degenerative mechanism reverses the process of neurons development which ultimately results in the form of functional disturbances, behavioural change, disabilities, cognitive and neurological disorders.

The rapid development of e-Health systems focused on improving the pattern of diagnosing and treating disease with the help of disease management or integrated care strategies [14]. The decision process taken by the physicians during diagnosis and treatment may be further improved through the implementation of information technology [15]. Unfortunately, conclusive diagnosis of NDDs is only possible posthumously, by direct examining the affected brain tissues after the death of a patient [16]. Obvious symptoms of these diseases are only visible at the last stage (Gait Impairment), when no remedy could be effective and the patient is left in a miserable condition waiting for his/her death. To gain a better understanding of neurodegenerative diseases it is worth considering some of these in more detail.

2.2.1 Alzheimer's disease (AD)

Alzheimer's is a neurodegenerative disease which poses the greatest growing challenge among the aging population [27]. The results from a recent survey show that while cancer and heart disease have typically been the top priorities in healthcare, Alzheimer's has become just as important in recent years [28]. This disease is so far proven incurable and irreversible. The exact cause is unknown and there is no evidence to suggest whether the disease or the build-up proteins is the root cause.

The changes in the brain that accompany these symptoms are "tangles" and "plaque" of a toxic protein—Amyloid Beta (Aβ). These pathological neurofibrillary tangles accumulate in the entorhinal cortex and hippocampus parts of the brain that are responsible for the short term and the long term memory of a person [29]. Neuroscientists have reported that in order

to keep memory alive the communication between these two parts is very essential and any hurdle between these two regions breaks the circuit and leads towards memory disturbance and eventually memory loss [3].

2.2.2 Parkinson disease (PD)

Parkinson is another neurodegenerative disease, first described by James Parkinson in 1817 [30]. More than 2% of the population over 65 years and approximately 5–20/100,000 individuals per year are affected by this disease, indicating its prevalence and incidence rate linked with aging [31]. According to a UK health economic report, the total cost needed to be invested for the treatment of an individual is £5993 which is a huge economic burden and a probable threat is the increment in this cost in the coming years [32].

Parkinson's disease is characterized by the dopaminergic deterioration process of the nerve cells of substantia nigra [33], a part of the brain responsible for the production of "dopamine"—a chemical which works as a neurotransmitter for controlling movements in different parts of the body. The degenerative process starts from the base of the brain, leading to the destruction of olfactory bulbs, followed by the lower brain stem and subsequently susbstantia nigra and mid brain [34]. Eventually, it destroys the limbic system and frontal neocortex resulting in cognitive and psychiatric symptoms.

2.2.3 Huntington's disease (HD)

Huntington's has first discovered by George Huntington in 1872, and is a devastating degenerative neuropsychiatric disorder [35]. According to a report, the overall prevalence of Huntington's disease is 8 out of 10,000 people in caucasian populations [36]. So far, no preventive measures have been discovered for this fatal disease.

A specific part of the brain, PolyQ, has the Huntington's gene with 11-34 repeated sections of glutamine—responsible for the production of cytoplasmic protein called Huntington. When the PolyQ region generates more sections of glutamine, a mutant Huntington protein is produced which is the actual cause of Huntington's disease [37]. This disease is an incurable

hyperkinetic motor disorder. The primary symptoms of this disease are jerky and shaky movements called chorea [38].

2.2.4 Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic Lateral Sclerosis (ALS) also known as Lou Gehrig's disease is another neurological disease that degenerate both the lower motor neurons (LMN) and upper motor neurons (UMN). This male dominant disease has been mostly seen in individuals, 40 to 70 years old [39]. This disease, either in sporadic or familial forms, occur with an estimated incidence of 0.4 to 1.8 per 100,000, quite uniformly throughout the world [40].

To date, it is believed that the actual cause of this disease is a mutant gene, superoxide dismutase (SODI) that affects the motor neurons of the brains. Moreover, the toxicity in Cerebrospinal fluid (CSF) is also considered a cause of neuron degeneration [41]. Motor neurons are the nerve cells that are responsible for voluntary movement of the muscles [42]. Weaknesses in arms and leg muscles are the earlier symptoms of ALS which leads to sever attack to chest muscles, leaving patients unable to breathe [43].

This brief introduction reveals that different neurological diseases cause atrophy in different parts of the brain; Alzheimer's causes deterioration in cortex and hipcampus, Huntington's disease causes damage in caudate, Parkinson's in substantia nigra and Amyotrophic lateral sclerosis (ALS) damages the lower motor and pyramidal neurons, resulting in a severe damage to body movement.

2.3 The Developmental stages of NDDs

This section explains the development stages of NDDs, and the symptoms that appear at each stage. We also discuss how we can detect neurological diseases from the pre-clinical stage to their last stage. The development cycle of neurodegenerative diseases is divided into three main stages:

1. Retrogenesis;
2. Cognitive Impairment;

3. Gait Abnormality;

Retrogenesis: The starting point of NDDs is a malfunctioning of the cholinergic system of the basal fore brain, which further extends to the Entorhinal Cortex and the Hippocampus that are responsible for the short and the long term memory [2]. These changes in the brain usually start 10-20 years in advance and the first visible sign of NDDs is forgetfulness or some problems in short term memory, e.g., forgetting the place for eye-glasses, everyday objects, misplacing the keys, etc. Symptoms may include enhanced memory loss, attention loss, difficulties in recognizing the family members, needing help in getting dressed and also gait problems.

The disease with its progression starts affecting the cerebral cortex resulting in the form of further decrease in cognitive power. This stage is linked with the clinical diagnosis of NDDs in patients which include confusing among familiar places, losing decision power, misplacing valuable things, mood and personality changes, childish actions in office, increased anxiety, loss of spontaneity and sense of initiatives [2, 44].

Further atrophy in the affected area of the cerebral cortex results in the form of serious problems with language, sensory neurons and reasoning. Patients show serious attitude towards wandering and agitation. Symptoms may include enhanced memory loss, attention loss, difficulties in recognizing family members, needing help in getting dressed and also gait problems.

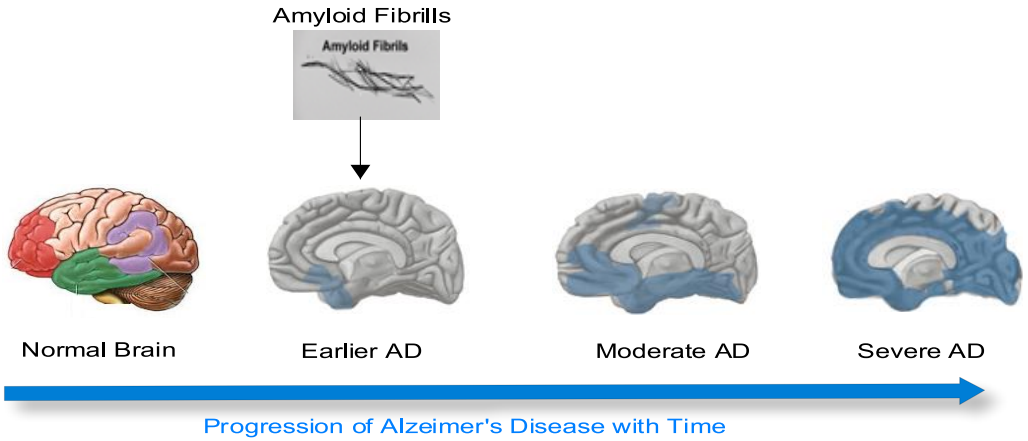


Figure 2-1: Manifestation of pathology and its progression in AD

Figure 2-1, shows the process of retrogenesis where the darker areas depict the affected parts of the brain. Similarly, this process of retrogenesis has been elaborated in Table 2-1, which shows the process of normal human brain development compared to the deterioration of brain cells due to neurodegenerative diseases. Here, the upward arrow indicates the development process and the downward arrow shows the destruction of brain parts.

Table 2-1: Brain Development vs. Brain Deterioration

Human Development Stages Vs. Alzheimer's Stages			
Developmental Stages	**Acquired Abilities**	**Alzheimer's Stages**	**Lost Abilities**
Adolescence-to- Puberty	Work nicely without help Develop working skills Manage routine works accurately	Preclinical-to- Early Stage	Work with less confidence Losing focus on skills Minor mistakes in work
Mid Childhood –to-adolescence	Get good memory Try to learn complex tasks Managing with clothing and food Good understanding	Early Stage-to- Mild Stage	Forgetting little things Cannot handle complex tasks Difficulty in managing food and getting dressed
Early Childhood-to- Mid Childhood	Walk steadily Try to do small tasks Manage to put on cloths Taking shower on their own	Mild Stage-to- Moderate Stage	Disturbance in walking Cannot perform small tasks Cannot take shower on their own
Infancy-to- Early Childhood	Holding up head Trying to sit Smile Shaky walk Try to speak	Moderate Stage-to- Severe Stage	Speaking problems Cannot walk Loss of memory Cannot hold-up their head

Cognitive Impairment: There is a very close relationship between neuro-degeneration and toxic proteins. This stage is accompanied with the accumulation of pathological

neurofibrillary plaques and tangles in the entorhinal cortex (EC), hippocampus, caudate, substantia nigra parts of the brain. These proteins play a pathogenic role in the progression of NDDs which results in the form of neurons degeneration and memory impairments. The Entorhinal Cortex (EC) is that part of the brain which gets affected due to Alzheimer's. Neuroscientists have reported that in order to keep memory alive the communication between the Entorhinal Cortex (EC) and the Hippocampus is very essential and any hurdle between these two regions breaks the circuit and leads towards memory disturbance and memory loss. It is concluded that EC is the main hub which is more vulnerable to NDDs and these diseases propagate with the network of neurons [3].

Our research work shows that accumulation of these pathological proteins is another factor, which could help with the early prediction of Alzheimer's and other neurodegenerative diseases.

Gait Abnormality: Predicting a disturbance in gait activity indicates a disturbance in cognitive functions. Scherdera *et al* [45] have proposed a term "Last-in-First-out" which refers to the phenomenon that the neural circuits that mature late in the developmental life cycle are more vulnerable to neuro-degeneration and this concept helps in early prediction of any kind of dementia (Neurodegenerative diseases). Zhu *et al* [46] stated that a healthy gait pattern requires input not only from the neurological system associated with motor and sensory neurons but also from cortical processes such as judgment, planning and a spatial awareness. Higher level gait disturbances are under consideration these days, which are closely related to disturbances in cortico-cortical and cortico-subcortical connections, e.g., the frontal connection with parietal lobes and frontal lobes with basal ganglia, respectively [4]. Disturbances in cognitive function have a direct link with higher level gait disturbances and it is one of the main symptoms of brain disease. Figure 2-2 elaborates the relationship between neurological diseases and their effects on body movements.

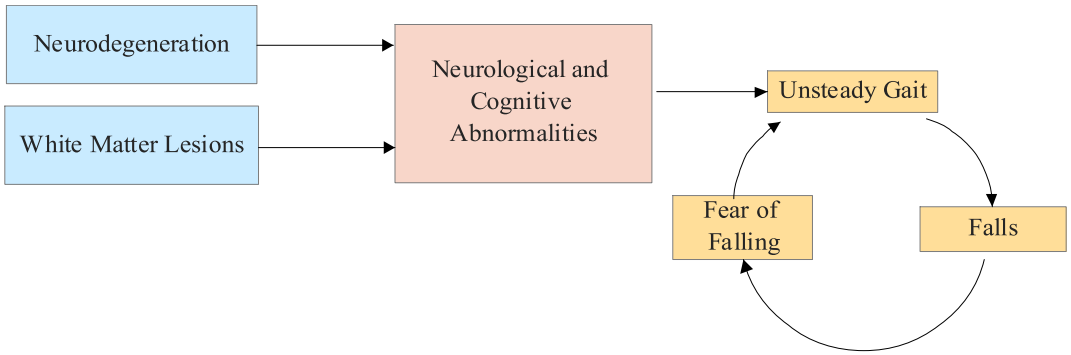


Figure 2-2: Relationship between cerebral pathology and gait disorder [47]

2.4 Bio-Signal Processing

Signal processing is the process of modelling, detection, identification and utilization of patterns and structures in a signal. Random signals are processed through statistical models of signal processing, which are used for decision making systems, extracting the relevant information from noisy, distorted and incomplete signals. A signal describes the information through variation of quantity which reflects the properties, characteristics, state, the course of action and the information about a source and that information may be processed directly by humans or machines for the purpose of decision, forecasting, control, investigation, research and further exploration of an object [48].

Biomedical signal processing centres on the acquisition of vital signals extracted from biological and physiological systems. These signals help us to obtain information about the current state of living systems, and therefore, their monitoring and interpretation have significant diagnostic value for clinicians as well for researchers to extract information related to human health and diseases.

Biomedical signal processing depends on the knowledge of their origin, nature of the signals, their properties and their complexities which come along signals. They have to be clearly examined to be processed non-invasively and indirectly due to their underling complex biological structure. In addition, the extracted signals are not always ready to be used because of noise. These unwanted signals are sometimes due to malfunctioning of the equipment or sometimes due to other body signals that create a hindrance in obtaining the required results.

Consequently, pre-processing of the signals is required to get the required set of data for further experiments.

This section briefly introduces the signals we have used for our research work, i.e., gait and EEG signals. It also provides information about feature extraction and feature selection techniques and also explains feature classification.

2.4.1 Gait and EEG Signals

A cyclic movement of the feet in which one or the other alternate strikes to ground is called gait [49], and the measures obtained by the stride-to-stride movements of the feet are called gait signals [50]. Hausdorff *et al* in [50] suggested that the understanding of the relationship between loss of motor neurons and the perturbation in the stability of stride-to-stride dynamics can help us to monitor neurodegenerative diseases progression and in assessing potential therapeutic interventions. Gait cycle duration is also referred as the stride time, i.e., fluctuation from one stride to the next in a complex manner. Due to intact neuronal control the fluctuation magnitude of the strides in control subjects is relatively small (~2%).

A variation of the surface potential on the scalp reflects the functional activity of the brain. This surface potential of the brain is collected by electrodes, attached on the scalp. The voltage between the electrodes is measured and ultimately this is filtered, amplified and the recorded data is collected, which commonly is known as EEG. EEGs are used as a method of investigating mental processes to investigate any perturbation in the brain activity. The EEG is roughly defined as the mean electrical activity in the brain at different sites of the head [51]. More specifically, it can be defined as the extracellular current flows of a large number of neurons.

2.4.2 Feature Extraction and Feature Selection

The pattern recognition process consists of two steps; feature extraction and feature classification. A feature is one particular aspect of an instance that can assist in grouping it to a particular class. In other words, features are synonymous of input variables or the attributes of a dataset that provide good representation of a specific domain, related to the available

measurements [52]. In the case of medical diagnosis, these features can be the symptoms of a disease. The features can be qualitative or quantitative as shown in the Figure 2-3.

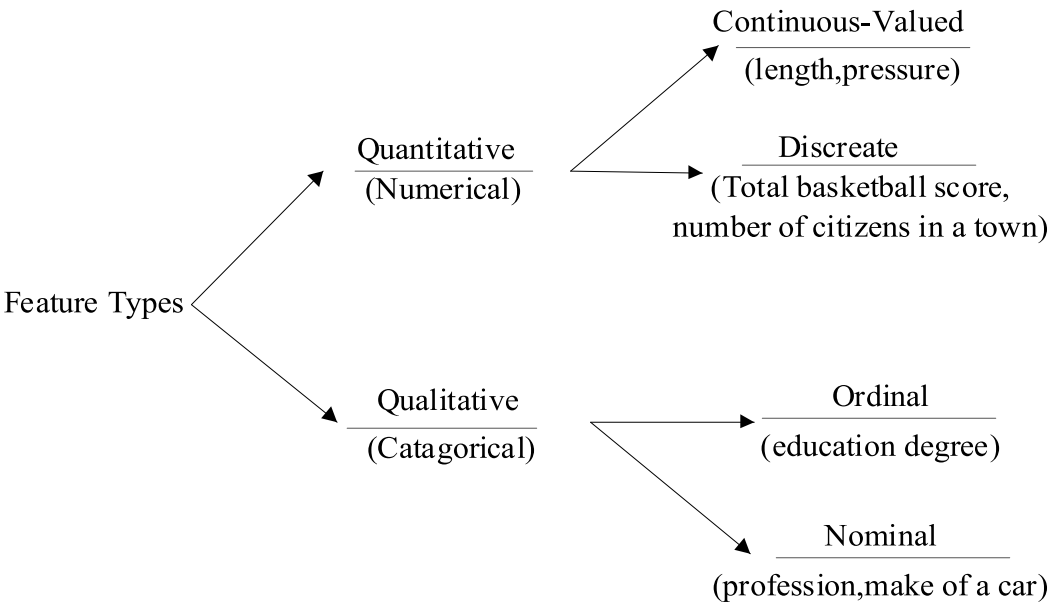


Figure 2-3: Types of Features [53]

Feature extraction consists of finding a set of measurements or block of information to present the properties of the signal [54]. These features are the basic index of detection, classification and regression in the field of biomedical signal processing and also in data analysis.

The expression of the features can be binary, categorical or continuous. For instance, they can be the physical condition of the patient (age, health status, family history), position of the electrode on the scalp to get EEG signals, or may be EEG signal descriptor (frequency, voltage, amplitude, phase, etc.) [54]. The performance of the pattern recognition system depends on the features we select and also on the classification algorithms.

This process also removes erroneously recorded signals caused by sensor malfunctioning and noise that can have a negative effect on signal classification. This process can be defined using the following mathematical formula and the process is illustrated in Figure 2-4.

$$\begin{cases} I(t) \cong S(t) \\ F(t) = S(t) - N(t) \end{cases} \tag{2.1}$$

Where *I(t)* is the data retrieved from the data source, that is mapped to some signal *S(t)* and the inherited noise found in the signal is defined as *N(t)*. Consequently, the filtered value can be defined as the signal *S(t)* - the noise value *N(t)*.

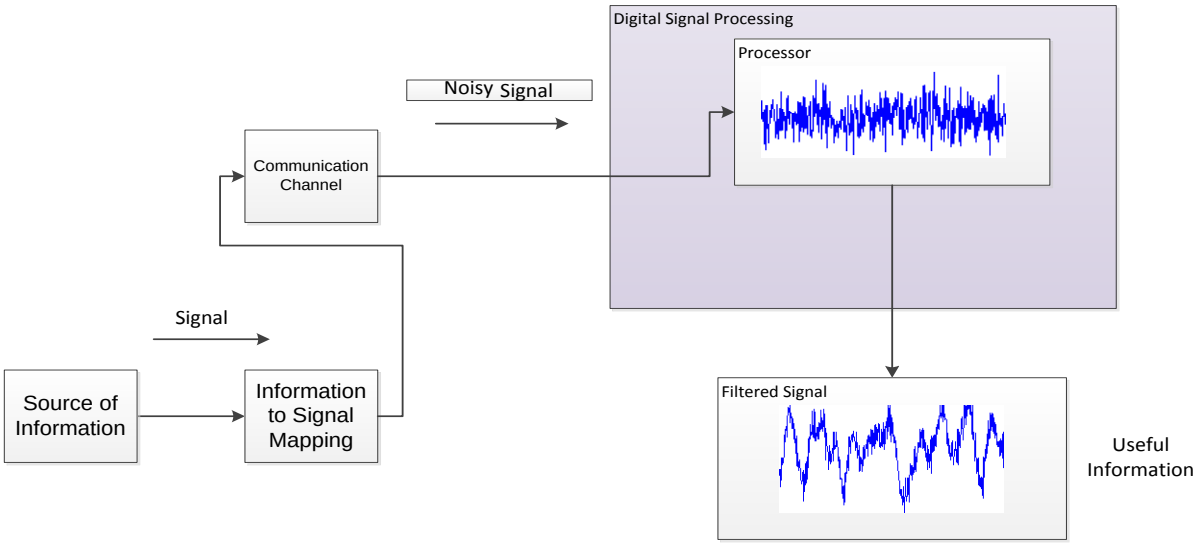


Figure 2-4: Feature Extraction and Noise Reduction Process

Feature selection, on the other hand, is the process of identification and removal of irrelevant as well as redundant features from the datasets [55]. The existing datasets may have hundreds and thousands of features. Some of them could be totally irrelevant or some others may have redundant information. This can lead to more complications and also to the increased processing time of classification. This is also effective in handle multi-dimensional data, which ultimately enables data mining algorithms to work more efficiently and effectively. Different methods are available to handle this issue. More details are available in [55].

2.5 Pattern Recognition

The two major types of learning are supervised learning and unsupervised learning. The main problem with unsupervised learning is the recognition of structure of the data, that is, to know whether there are groups in the data or not. Also, what characteristics make the object similar within the group and different across the groups. Clustering is the best option for unsupervised learning. In clustering, there is no labelling of the data.

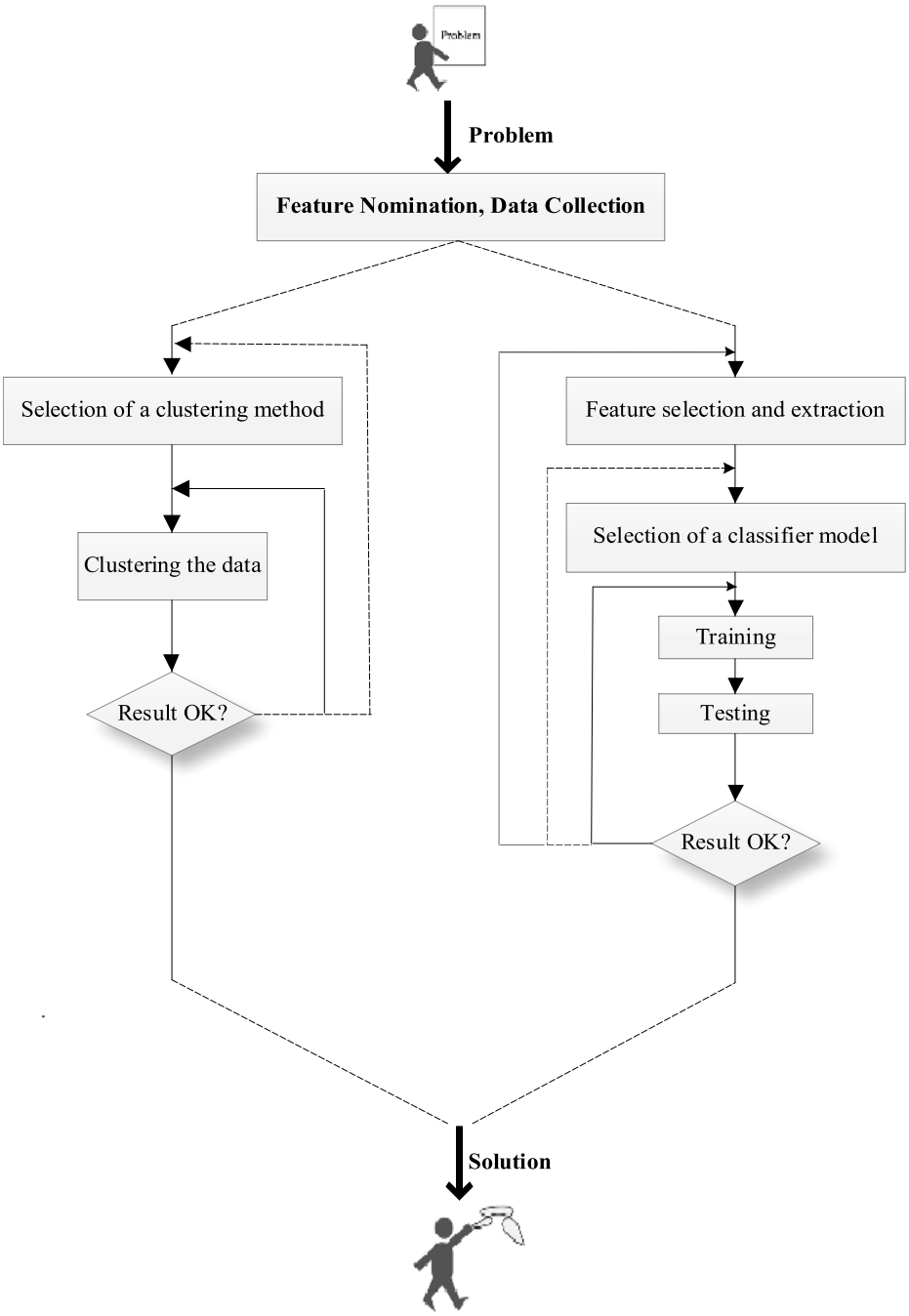


Figure 2-5: The pattern recognition cycle [53]

On the other hand, in supervised learning which is also known as classification, each object in the dataset has pre-assigned labels. The label for each object is provided at the learning stage while the testing stage recognises the particular class of an object.

Figure 2-5, presents a basic cycle of pattern recognition technique where a user comes with a problem and a set of data. The data is visualized and a possible technique is investigated to analyse the problem and finally the user is provided with a possible solution.

2.5.1 Pattern Classification/ Feature Classification

Feature classification, generally known as machine learning, is the automatic assignment of a class to the feature vector that has been previously extracted from the signals. The algorithms used for this classification are known as "classifiers". Classifiers are able to learn how to identify the class of a feature vector by providing it a training dataset. The training set constitutes the feature vectors already labelled with the exact class label. An important thing about the classifiers is that the learner does not know which action is to be taken to get better results rather it has to be discovered which algorithm is best, by trying out different classifiers. The reason is that almost every dataset comes with different specifications. The advantage and disadvantage of this technique is that there is no ground truth against which the results are to be compared. The users have to do some subjective estimation of the results to conclude their effectiveness.

The remainder of this section provides an overview of well-known classification algorithms. These classifiers are divided into four categories; logic or rule based classifiers, rule learners classifiers, perceptrons and statistical learning classifiers.

2.5.1.1 Logic Based (Symbolic) Classifiers

The decision tree and rule based classifiers are the main logic based classifiers.

Decision Trees

Decision trees classify the instances by sorting them based on their feature values. The node of tree is called a feature of an instance that has to be classified, while each branch represents a value that the node assumes. The classification of the instances starts from the root node and sorted based on their feature values. The branches then lead either to other features or end in leaf nodes, which are the classes. The root node divides the training data into possible branches. There are numerous methods that help to find out the root node such as information

gain and gini index. However Murthy has narrated that there is no single criterion which should be used to divide the dataset [1].

To avoid the overfitting of training data, two common approaches are usually considered:

1) Training of the classifier should stop before it reaches the point where it perfectly fits the training data;

2) Pruning of the induced decision tree. The most effective way is to pre-prune the decision tree before it grows to its full size. This can be accomplished by a threshold test for the feature quality matric. Else, in the post-pruning method, a check of the tree's performance is made if necessary, pruned.

Decisions trees are usually univariate because at each internal node they split the dataset based on a single feature. However, there are a few other methods that are constructed on multivariate features [56] to improve the classification accuracy by creating new binary features with logical operators such as conjunction, negation, and disjunction.

The main advantage of using decision trees is their comprehensibility. The classification of an instance to a particular class is easily understandable. Also, another aspect is that the decision tree works better for discrete/categorical features.

Rule Learner

Rules can be derived from decision trees by creating a separate rule for each path taken from the root to each leaf node [57]. A training dataset can also be used to generate rules using a variety of rule-based algorithms. Algorithms that are used to construct a rule are called "separate-and-conquer" algorithms or "covering" algorithms. The only difference between rule learner and the decision tree is the former evaluates the quality of the set of instances while the decision tree evaluates the value of each feature that is to be tested. Further advancements in the rule learner added additional features (characteristics) to avoid "over-fitting" by stopping the specialization process with the use of quality measures or by generalizing overly specialized rules in a separate pruning phase [58].

For a rule induction system, it is very important to generate decision rules with high probability or reliability. A "Rule quality" function is mostly used to measure these qualities

such as the J-measure. In case of a conflict when multiple rules are agreed by the example to be classified, a rule quality measure is associated which each rule for a final decision. These rules can be statistical or empirical. RIPPER is a rule-based learner which works with the process of growing and repeating. The growing phase is more restrictive as compare to the repeating phase in order to fit the training data as well as to avoid over-fitting.

Rules are normally more comprehensive than decision trees for learning a binary problem, since with rule learners, only the rules for the positive class needed to be learnt. If a multi-class dataset is to be classified then the rule based learner must be run separately for each class. One disadvantage of this is that for each class a separate rule is needed and that could be inconsistent or incomplete but these problems are not common with the decision tree algorithms. Moreover decision tree algorithms work more efficiently as compared to rule based learners. Rule based learners work on the principle of separate and conquers while decision tree works on divide and conquer rules. Flach and Lavrac [59] suggested that the classification accuracy of rule based learners can be improved by combining features of users from their background knowledge as well as by automatic feature construction algorithms.

2.5.1.2 Artificial Neural Network (Perceptron based technique)

An artificial neural network (ANN) is a machine learning technique based on the connections of neurons in our brain that mimic the learning capabilities from experience. Neurons are simple processing unit cells and exist in millions in the brain. Each neuron is connected to many thousands of other neurons.

As a neural network is trained from past data, they are trained to generate output based on the information extracted from the previous training dataset [60]. The common way of learning in the perceptron algorithms is they go through the training dataset again and again until they find an output vector which is correct for all training sets. Later on, the learned weight matrix is used to classify the test data [1].

There are several advantages of using ANNs: 1) without making prior assumption of the function, they can easily adjust with the datasets, 2) being a universal function approximator, they can easily approximate any function with arbitrary accuracy, 3) they are a non-linear model, hence can be used for most complex real world application, 4) there are used in many

critical infrastructures like industry, business, fault detection, science, bankruptcy prediction, hand writing recognition, and bio informatics [60].

Some commonly used neural networks based on single layer or multi-layer perceptron are discussed below.

Back propagation neural network (BPNN)

This is a simple and effective model, which is also known as the feed forward back propagation neural network. It is based on three layers, input, hidden, and output layers. During training, the data is fed into the input layer, which is then propagated to the hidden and finally the output layer, called forward pass. The weight is calculated and adjusted on these layers on input, output and hidden layer to generate output value of the resulting sum. The actual and target values are compared and the calculated error is then propagated back to the hidden layer. This is used to update the weight of each node again. This is called backward pass or learning. This cycle keeps working until the error is acceptable. This model can be used for the test data which does not need any modification in the weight matrices. The input layer receives the test data and the feed forward network then generates the results based on the trained network [41]. Figure 2-6 shows the learning process of a neural network.

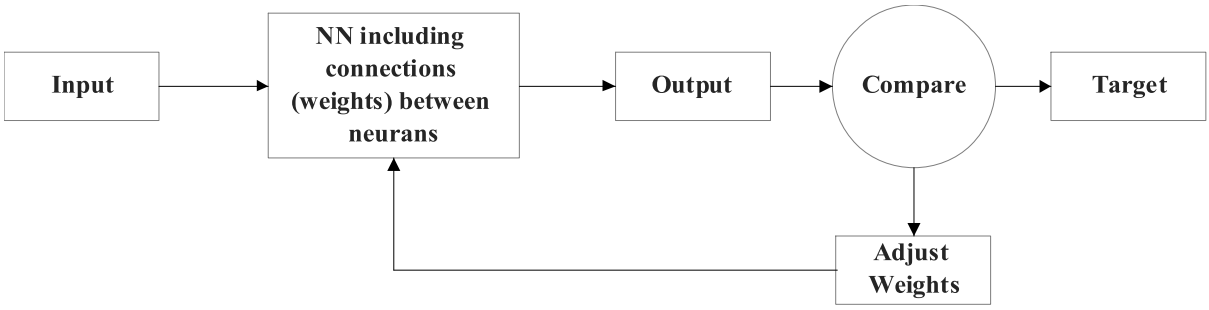


Figure 2-6: Learning process of BNNP [61] referred by [1]

BNNP is a robust neural network and widely used in many applications but it still has many limitations like; training time of the classifier is very high due to the number of input and target pairs. Also the internal mapping of the trained classifier on the test data work as a black box which is difficult to understand and does not provide any confirmation if it can provide all acceptable solutions [62]. Determining the size of the hidden layer is also a problem; the

exact number of neurons should be known to make a perfect approximation as well as for better generalization capabilities. On the other hand a large number of nodes can lead to the overfitting problem [1].

Radial Basis Function Neural Network (RBFNN)

It is a multi-layer neural network based on an input layer, a kernel (hidden layer), and an output layer. Kernel basis functions in the kernel layer are called radial basis functions. It is different from the BPNN in its training algorithms. Also the kernel units calculate the output of RBFNN as a linear combination of radial basis functions. It is considered as a Multi-Layer Perceptron (MLP) because the parametric statistical distribution model and non-parametric statistical distribution are combined in a serial sequence [60].

Some advantages of using RBFNN over BNNP are that can be trained the classifier quickly because a single hidden layer can be used for modelling any non-linear function. They have better mapping capabilities due to their simpler architecture. Due to these characteristics RBFNN is considered an interesting alternative for pattern classification.

General Regression Neural Network (GRNN)

GRNN is a feed-forward neural network, which uses non-linear regression functions for approximation of supervised data. It also constitutes three layers; input layer, hidden layer, and output layer. The input layer is linked with the output layer by direct mapping. Instead of using learning rate or momentum as a transfer function, GRNN for learning phase, uses smoothing parameters. The computation time for GRNN is remarkably less because of two reasons, mentioned in [60], 1) there is a one pass training of the data through the network, 2) a single smoothing factor is selected to optimize the transfer function for all nodes.

Due to above mentioned characteristics, the GRNN improves the learning process as well reduces the computational complexity. It uses the non-parametric estimator density function like the probabilistic neural network but the difference is that they are suitable for continuous values while PNN works better to find boundaries between categories of pattern.

There are other kinds of neural networks like the Probabilistic neural network (PNN) and the Complementary Neural Network (CMTNN). PNN is a type of radial basis networks, related

to the Bayesian decision rule and Parzen while CMTNN uses a pair of opposite feed forward back propagation neural network for classification. More details are available in [63].

2.5.1.3 Statistical Learning Algorithms

In contrast to the previously described learning algorithms, statistical algorithms assume that the process of learning task itself is riddled with uncertainty. Therefore, instead of assigning an instance definitely to one class or another, they are given a probability that they belong to a particular class. For instance, linear discriminant classifier (LDA) and also Fisher's discriminant classifier are used to find the linear combination of features that best describe the assignment of an object to a specific class [1]. It is considered that Bayesian networks are the most powerful and well-known representatives of statistical learning technique.

Bayesian Networks

A Bayesian network provides a graphical representation of presenting the probability relationships among a set of features/variables. It is a directed acyclic graph (DAG), where the nodes are random variables connected by directed arcs. An arc denotes a direction of casual influence. Any nodes that do not have an arc directed towards them are called "root nodes". These root nodes must be given a prior probability. All other nodes have an associated probability table. This table is filled with conditional probabilities, which state the probability of that random variable appearing, given observable evidence of other random variables. The evidence must be obtained from a node that is an immediate neighbour only, from another node directed towards the node.

The Bayesian network provides a simple and flexible method to solve a problem. In theoretical format it can be represented as [64]:

$$P(h|e) = \frac{p(e|h).p(h)}{p(e)} \tag{2.2}$$

Here $p(h)$ is the prior probability of the proposed hypothesis; $p(e)$ is the prior probability of evidence e; $p(e|h)$ is the probability of e given h, while $P(h|e)$ is the probability of h given e.

Naïve Bayes Classifiers

A Naïve Bayesian (NB) network is a basic and simple example of Bayesian network. It is composed of a cyclic graph in which the class variable is always the parent node, while the observable variables are its children. The model dictates that there are strong assumptions of independence between the children nodes, which means that there are no arcs between the children [1]. Although NB is considered a very simple and efficient classifier as it does not take much time for training, yet this simplicity comes with lower predictive accuracy.

One other advantage of NB is the handling of missing attributes in the datasets, while other Bayesian networks simple ignore the missing attributes of the data. Other classifiers use more sophisticated techniques such as model imputation or sample deletion, or else a more computational expensive technique such as expectation maximization [65].

Other advantages of using NB include its short computational time as less time is required for the training of the classifier. The authors in [1] also suggested that if the product form is converted into the sum model using some algorithms then significant computational advantages could be achieved. They also mentioned that numerical features are discretized during data pre-processing although numerical distributions can also be used to calculate the probabilities.

There are some disadvantages of using Bayesians network. For instance, BN handle discrete variables better than continuous variables. Also, BN are not suitable for datasets with a large set of features as for this kind of data a very large network is needed to be constructed which is simply not feasible in terms of time and space [1]. Another problem is the discretization of the numerical features before induction, in most of the cases.

Linear discriminant classifier (LDC)

This is a simple classifier, which works effectively even when the classes are not distributed normally [53]. A discriminant function which is obtained by monotonic transformation of posterior probabilities *p(wi|x)* consists of:

$$g_i\,(x) = \log[P(w_i)p(x|w_i)] \qquad i = 1, \dots, c \tag{2.3}$$

Where $p(w_i)$ is the prior probability of the class w_i and $p(x|w_i)$ is the class conditional probability density function.

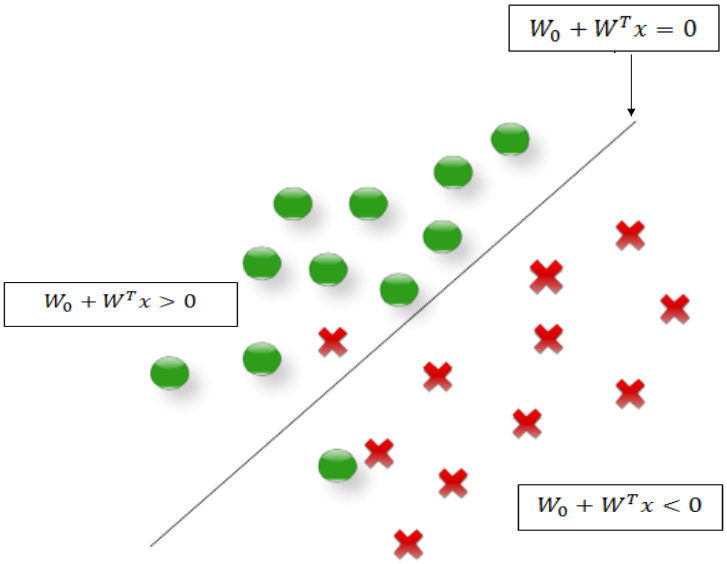


Figure 2-7: A hyperplane which separates two classes: the "Circles" and the "Crosses"

The separating hyperplane (as shown in Figure 2-7) works on the projection that maximizes the distance between the two classes means and minimizes the interclass variance [66]. In our case where we are using LDC for a multiclass classification, which means N>2 several hyperlanes will be used to separate the feature vector in four classes. The strategy, generally used in multiclass datasets is "One Versus the Rest" (OVR) strategy, which works by separating one class from the rest of the classes. The computational requirements for this technique are not very high. Also, it is a stable classifier because the results usually do not vary much by varying the training dataset.

Quadratic Discriminant Classifier (QDC-Bayes Normal)

This classifier produces non-linear decision boundaries between datasets of different classes. The Bayes rule is used to compute the posterior probabilities of a feature vector to decide which class it belongs to. It associates the feature vector to the class with the highest probability. QDC works by assuming a different normal distribution of data, providing quadratic decision boundaries, as the name of the classifier depicts [66].

Uncorrelated Normal Density based Classifier (UDC- Bayes Normal)

It works similarly like the quadratic classifier but the computation of a quadratic classifier between the classes in the dataset is done by assuming normal densities with uncorrelated features. The Quadric Bayes classifier takes decisions by assuming different normal distributions of data. It leads to quadratic decision boundaries, as the name of the classifier reveals.

A linear classifier predicts the class labels based on a weighted, linear combination of features or the variables of the objects [67]. Logistic, Fisher's, nearest means and polynomial are a few linear classifiers, available in PRTools [68].

Logistic Linear Classifier (loglc)

The logistic linear classifier computes the classification of a dataset by maximizing the likelihood criterion using the logistic (sigmoid) function [69]:

$$logistic = \frac{1}{1 + e^{-x}} \tag{2.4}$$

The only drawback with this function is that it does not perform well when the values of features exceed 1000.

Fisher's Discriminant (Minimum Least Square-fisherc)

By minimizing the errors in the least squares sense, this function finds a linear discriminant function between the classes in the dataset. This "one-against-all" strategy applies on all multi-class implementations, which also works for soft and target labels. Pseudo-Fisher procedures, based on pseudo-inverse, are used for high dimensional datasets or small sample sizes. This classifier does not use the prior probabilities stored in the datasets [70].

Nearest Mean Classifier (nmc)

The nearest mean classifier (nmc) is a plain nearest mean classifier which is sensitive to feature scaling but does not use any prior class probabilities, i.e., it is insensitive to class priors. In the nearest neighbor classifier the test data is classified according to the Euclidean distance between the test sample and the nearest trained sample and here again in the nearest

mean classifier the mean of the Euclidian distance of nearest trained class is computed to classify the sample data.

$$Euclidean\ distance = d(x,y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \tag{2.5}$$

Polynomial Classifier (polyc)

Polynomial classifier adds polynomial features to the datasets in order to run the untrained classifier. In this classifier, the combination of 2nd order terms may also be constructed but for higher orders no combinations are generated. This is also known as higher order neural network (HONN) [71].

Some non-linear classifiers that are selected from Prtools to manipulate our datasets are parzen, decision tree, support vector machine and k-nearest neighbor. A brief explanation of these classifiers is given below.

Parzen Classifier (parzenc)

It computes the optimum smoothing parameter between the classes in the datasets. The leave-one-out estimate classification errors and final classification is stored as mapping. Parazenc is unable to calculate the density estimate. Discrimination is produced for smoothing parameters without any learning process. Smoothing parameters may be scalar, vector or a matrix with objects and their features [67].

Binary Decision Tree Classifier (treec)

The Computation of a decision tree classifier is done out of a dataset using binary splitting. Classification of large datasets may cause some problems but the decision tree solves this problem. In this type of classification, the subjects' classes are decided on the basis of sequence of decision rules. A decision tree is constructed in two phases; the 'growth phase' (initial tree) and the 'prune phase' (sub tree) [72].

Support Vector Classifier (SVC)

Support Vector classifier (SVC) is optimized for a dataset by quadratic programming in which the non-linearity is determined by the kernel. If SVM models use sigmoid kernel then it behaves more or less like two-layer, perceptron neural network. There are four basic kernels; linear, polynomial, radial basis function (RBF) and sigmoid. In this type of classification training set is mapped by the function Ø into a higher dimension space. It finds a linear separating hyperplane with the maximum margin in higher dimension space [73].

There are two particularly attractive properties of SVMs:

1. A decision boundary, called a maximum margin separator is built so that the distance between points of different instances on either side is it is as large as possible. This helps in generalization.
2. The decision boundary is a linear separating hyperplane, but SVMs embed the data into a higher dimensional space, with a kernel trick. This allows data that is not separable in the original space, to be more easily separated in a higher-dimensional space.

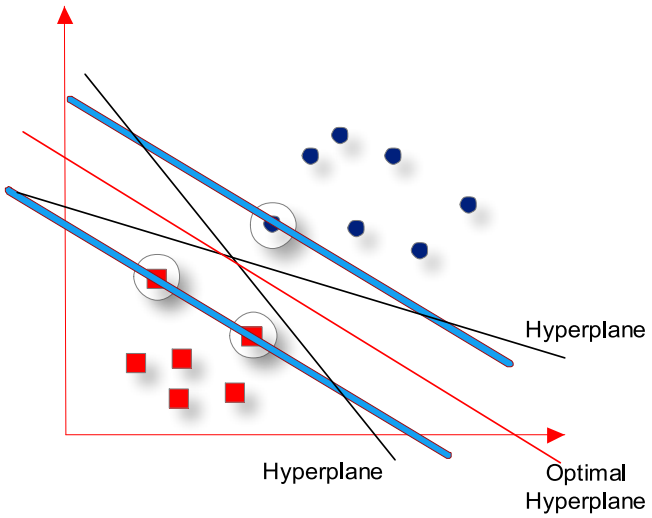


Figure 2-8: An example of Support Vector Machine [1]

Figure 2-8 above shows an example of a group of instances, with the optimal hyperplane separating them, and the maximum margin on either side of the hyperplane. The hyperplane lies midway between the two margins.

K-nearest Neighbor Classifier (knnc)

knnc and parzenc are similar in the sense that their build-up classifiers still use the training dataset and their parameters, while knnc classifies the object in a feature space with the nearest training parameters. It is also called instance based learning, where all commutations are postponed until the end of classification. Classification of the object is based on the neighbor's selection which is correctly classified at the time of training and that neighbor's class is assigned to the object [74].

2.6 Summary

This chapter has elaborated the neurodegenerative diseases along with their developmental stages, their symptoms and their effects on patients. Different kinds of neurodegenerative diseases and their causes have also been explained in this chapter.

It has provided a brief introduction of bio-signal processing with a further explanation of supervised and unsupervised machine learning. We have reviewed the process of machine learning, i.e., data preprocessing, feature extraction, feature extraction and feature classification. Then we have reviewed different classification algorithms, their advantages and disadvantages. We have divided the classifiers into different categories, 1) rule based, 2) neural networks, and 3) statistical learning algorithms. Statistical learning algorithms are further divided into, 1) linear algorithms, 2) non-linear algorithms and 3) density based (Bayes rule) algorithms.

This chapter has highlighted that there is a large number of classification algorithms available for the classification of data. Despite this large number of studies, the most appropriate algorithms, if any, have not been identified yet. Moreover, it explains the need to explore and/design more efficient algorithms, in terms of accuracy and efficiency for the early detection of neurological diseases.

Chapter 3 A strategic Framework for the Early Detection of NDDs

3.1 Introduction

This chapter provides an overview of the whole research work step by step. More precisely, this chapter focuses on the framework that we follow for the early detection of neurodegenerative diseases. It describes the different phases of the framework to analyse gait and EEG signals. A brief justification of the tools and techniques that we use for our research work is also provided in this chapter.

3.2 Approach Overview

The design goals provide the system requirements for a suitable scheme as described in this section. The principal goals are as follows:

- Access industry recognized gait and EEG datasets (this research work considers gait and impaired neuron symptoms or indicators as a biomarker to detect the occurrence of neurodegenerative diseases) for classification;
- A classification fusion strategy that combines state-of-the-art classifiers to improve early detection (details provided in Chapter 5);

- A system for medical practitioners that provides real-time symptomatic data and analysis of neurodegenerative diseases to support diagnosis and treatment strategies.

Our proposed framework portrays the methodology for developing a model for the early detection of neurodegenerative diseases using statistical pattern recognition techniques. Figure 3-1 demonstrates the proposed framework which incorporates several distinct processes; *Data Gathering, Feature Extraction, Feature Classification, and Decision Making*. This whole framework is divided into four main phases that are elaborated below.

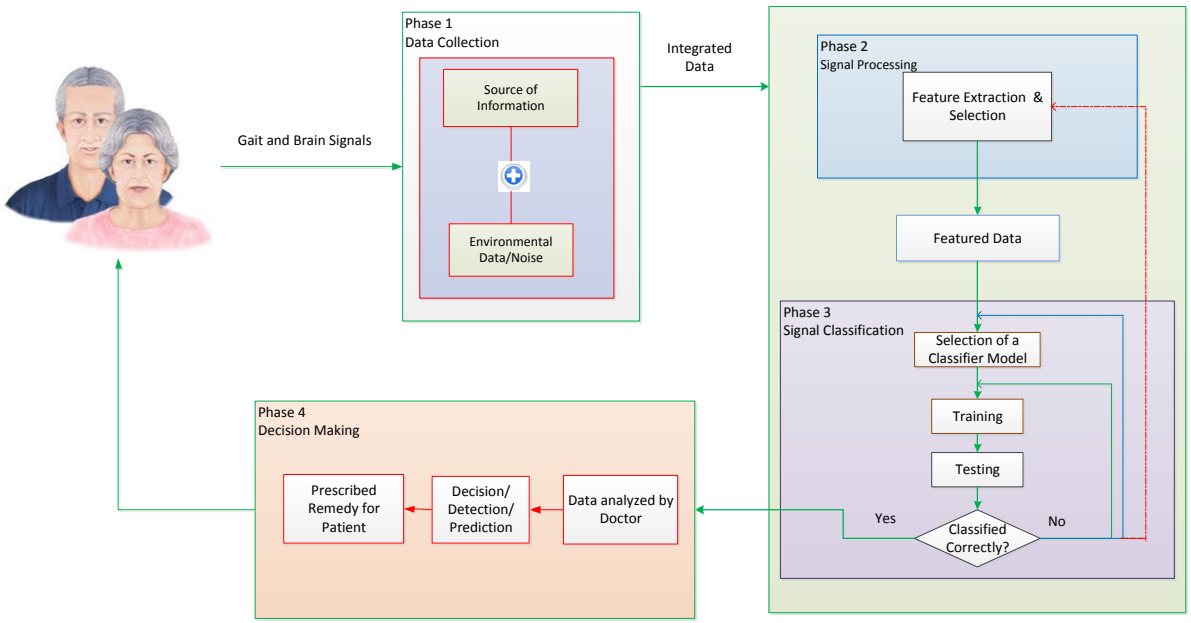


Figure 3-1: Proposed Framework for the Early Detection of Neurodegenerative Diseases

3.2.1 Phase 1: Data Collection and Data Integration

Data Gathering is the initial and the most challenging stage, where we get relevant data for both healthy persons and unhealthy persons. Data should be related to the appropriate domain, making sure that it provides the relevant information/ data patterns that are required for decision making. The most appropriate data that can provide a clue about neurodegenerative diseases is gait and brain data. These two are the most vulnerable parts of the body that get affected by neurological diseases.

Data integration is the process of merging data from multiple heterogeneous sources into one coherent database. This helps in providing each dataset, the same physical storage structure, naming convention, same unit of measurements, encoding structure, and also identical data type formats of attributes. This avoids any redundancy and inconsistency of the datasets thus ultimately improving the accuracy and efficiency of the mining process.

3.2.2 Phase 2: Signal Processing

This phase constitutes dimensionality reduction, feature extraction and feature selection. Noisy data come either due to faulty apparatus or due to variance in the datasets. This should be eliminated to get accurate results. Furthermore, once the data is cleaned and the required features are extracted from the datasets, the next step is the selection of appropriate features that best describe a particular disease.

3.2.3 Phase 3: Signal classification

This step is further divided into four steps; selection of a classifier according to datasets and their features, training is associated with assigning a "class label" to an object, entity, or to an event which is based on the measurements extracted from that particular object through any sensory system, testing a classifier on "test data" and finally checking the result if it is classified correctly or not. The plain, dashed and dotted lines in Figure 3-1, indicate that if the results are not according to expectation then the problem may be with the training of classifier, selection of the classifier or selection of relevant features, respectively.

3.2.4 Phase 4: Decision Making

Once the data is classified and evaluated, the next step is the interpretation of the patterns. This includes the interpretation of discovered patterns and also the visualisation. These results are then incorporated into a performance system so that appropriate action should be taken based on that knowledge.

Specifically speaking for neurodegenerative diseases, the decision cannot be taken simply by processing one kind of signals. As neurological diseases impair different parts of the body,

especially gait of the person and brain functions. So, the final decision can only be taken based on gait and EEG signals computation. Based on this critical situation, we divide our research work into two segments. First, the gait signals of different neurodegenerative patients like Parkinson's, Huntington's, and Amyotrophic Lateral Sclerosis (ALS) are analysed and compared with the gait patterns of healthy persons. The results are then evaluated using different evaluation techniques. Later, EEG signals of Alzheimer's patients are processed to find the difference in neural synchronization of different functional areas of the brain from healthy persons.

3.3 Statistical Tool Selection

Pattern recognition, these days, is gaining a tremendous attention in the medical field as it has proven more reliable in the prediction of clinical outcomes as compared to common clinical statistical tools [75]. For instance, Lin [76] designed a framework for the treatment of liver infection using classification and regression trees, Lee et al. [77] designed a system for the pulmonary nodules using feature selection and LDC. Similarly, Shao et al. [78], have successfully classified electromyography (EMG) signals (with 100% accuracy) using artificial neural networks (ANNs) for the identification of term and preterm labour of rats. Defending pattern recognition techniques, they claimed that several techniques like monitoring contraction by examiner, cervical state, intrauterine pressure (IUP), and tocodynamometry are subjective and do not provide accurate diagnosis or prediction of delivery time. Further arguing, they said, although few methods can assist to identify the oncoming signs yet none of the current method offers objective data processing that accurately predicts labour over a broad range of patients. They also highlighted the limitation of available technologies (multiple preterm labour symptoms, contraction >4 per hour, cervical ultrasonography, or fetal fibronection) in terms of less sensitivity and low positive prediction values. Similarly, Dan et al. [79] successfully classified Parkinson's disease by SVM classifiers using functional magnetic resonance imaging (fMRI) and structural images, as features. They obtained remarkable results with accuracy of 86.96%, sensitivity of 78.95%, and specificity of 92.59%.

Within supervised machine learning techniques data patterns can be classified using template matching, neural networks and statistical techniques [80]. One limitation with template matching is its inability to recognize patterns when they come from classes with large interclass variations. In our case, data is more sequential hence new datasets can be adjusted between the classes of the existing training sets. With neural networks, they behave like "black box" with excessively complex nonlinear input-output relationships, thus visual interpretation of the data itself becomes a challenge. In our case, the intension is to detect the anomalies in gait and EEG data patterns that need to retain maximum relation back to the actual physical measurements. On the other hand, in statistical analysis, presentation of each data pattern is held (in the form of a single point) in a multi-dimensional space, disjointing the regions for each class. Also, this approach retains the physical interpretation of the feature.

3.3.1 Selection of Matlab Tools

The literature survey has given us significant confidence to further explore pattern recognition techniques for the early detection of neurodegenerative diseases. However, the next challenge we face is the selection of classifiers that can be used for gait and EEG pattern identification. We intend to carry out computation on our dataset using PRTools, a Matlab toolbox used for pattern recognition [81]. We already evidence the classification of clinical data using PRTools in previous research findings [7], [82]. Complex bio-structured data patterns are analyzed using PRTools like tyrosine phosphoproteomic data from lung cancer [83] and also Gastric carcinoma and primary gastric lymphoma (PGL) in the stomach [84].

Other Matlab tool boxes that we have selected are the statistical toolbox to compute Principal Component Analysis (PCA). The signal processing toolbox is used to compute the coherence and correlation between EEG signals. Also, the communication toolbox is used to compute the phase synchronization of the signals.

3.4 Summary

This chapter has explained the framework; we have followed for the early detection of neurodegenerative diseases, in detail. It has highlighted the end-to-end process of data discovery, which incorporates several distinct processes; *Data Gathering*, *Feature Extraction*, and *Feature Evaluation*. Furthermore, it has stated the significant advantages of using statistical pattern recognition techniques. At the end, we have justified the use of PRTool for gait pattern recognition in neurodegenerative patients and healthy persons. We have mentioned other Matlab toolbox (communication, statistical, and signal processing) that have been considered to accomplish this research work.

Chapter 4 Assessment of Gait Dynamics

4.1 Introduction

In the previous chapters, the potential of machine learning technique is presented to overcome the problems of biomedical sciences. Understanding the importance of *data mining* that helps in the early detection of a disease, this chapter focuses on the early diagnoses of different NDDs using statistical pattern recognition techniques.

Following are the main challenges, which we intend to investigate in this chapter:

- Addressing the issues with imbalanced datasets;
- Handling missing observations—missing entries;
- Classification of multiclass datasets (4-classes—multiclass pattern recognition);
- Diagnosing movement disorders with similar symptoms but of different root causes.

We already have published our results, based on the findings of this chapter in [85]. Furthermore, the same techniques are then applied to develop a pattern of behaviour for the detection and identification of patterns, which are the results of an attack on a critical infrastructure [86]. Later on, working on the same line, these techniques are then further verified on the early detection of preterm births, results are published in [87].

4.2 Gait Signals

The analysis of walking pattern in humans, in order to check the abnormalities in the body, is called gait analysis. Gait pattern analysis plays a vital role in the identification of any neurodegenerative disease. As we already have explained, in Chapter 2, a healthy gait pattern requires a direct input from the neurological system of the brain. Hence, any perturbation in the brain has direct impact on the gait patterns of a person. To find out this disturbance and the difference from the gait patterns of control subjects, we analyse gait data (motion vectors in milliseconds) of different kinds of patients with neurodegenerative diseases; Parkinson's disease (PD), Huntington's disease (HD), and Amyotrophic Lateral Sclerosis (ALS).

Figure 4-1, demonstrates the stride time in CO, PD, HD and ALS. The Y-axis, in all four subjects, shows the mean value of the stride time, calculated by taking the average over five minute walk. It shows that the coefficient of variation (CV), a measure of stride-to-stride variability, is highest in PD and HD subjects, while smallest in CO subjects and higher in ALS subjects comparatively.

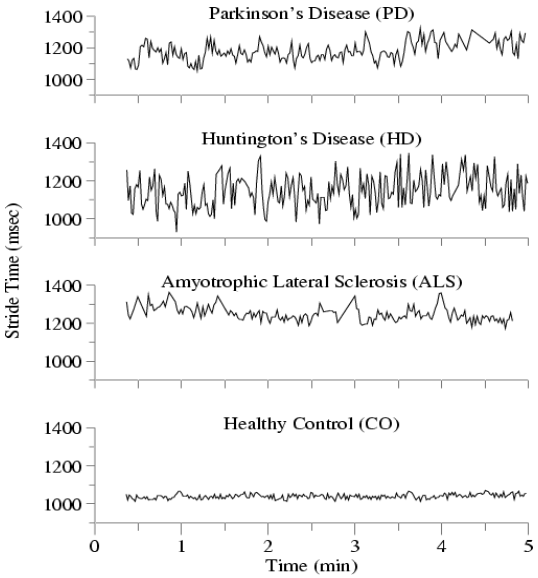


Figure 4-1: Gait cycle duration in CO, PD, HD and ALS [5]

4.2.1 Discrimination of different NDDs with Gait Pattern Analysis

Diagnosis time is of vital importance in the treatment of a disease especially for chronic diseases. The main challenge with NDDs is that they all pose the same symptoms at the final

stage—gait disorder. At this stage, it is very difficult to discriminate a specific neurodegenerative disease with a non-invasive method. Furthermore, each NDD has different root causes irrespective of their similar symptoms at the final stage. This has been elucidated in Figure 4-2, which demonstrates that each NDD shows symptoms' similarity at its final stage but the root cause of each disease is different. Different parts of the brain get affected due to different degenerative diseases; HD causes damage in caudate, PD in substantia nigra [88], ALS damages the lower motor and pyramidal neurons [42], while AD attacks the cortex and hippocampus part of the brain [89]. Moreover, the pathological proteins that are responsible for these diseases are also of different kinds; Amyloid Beta Protein (Aβ protein) for Alzheimer's, α-synuclein for Parkinson's, polyQ mutant Huntington protein for HD and SOD1 (superoxide dismutase1 gene) toxicity for ALS [90].

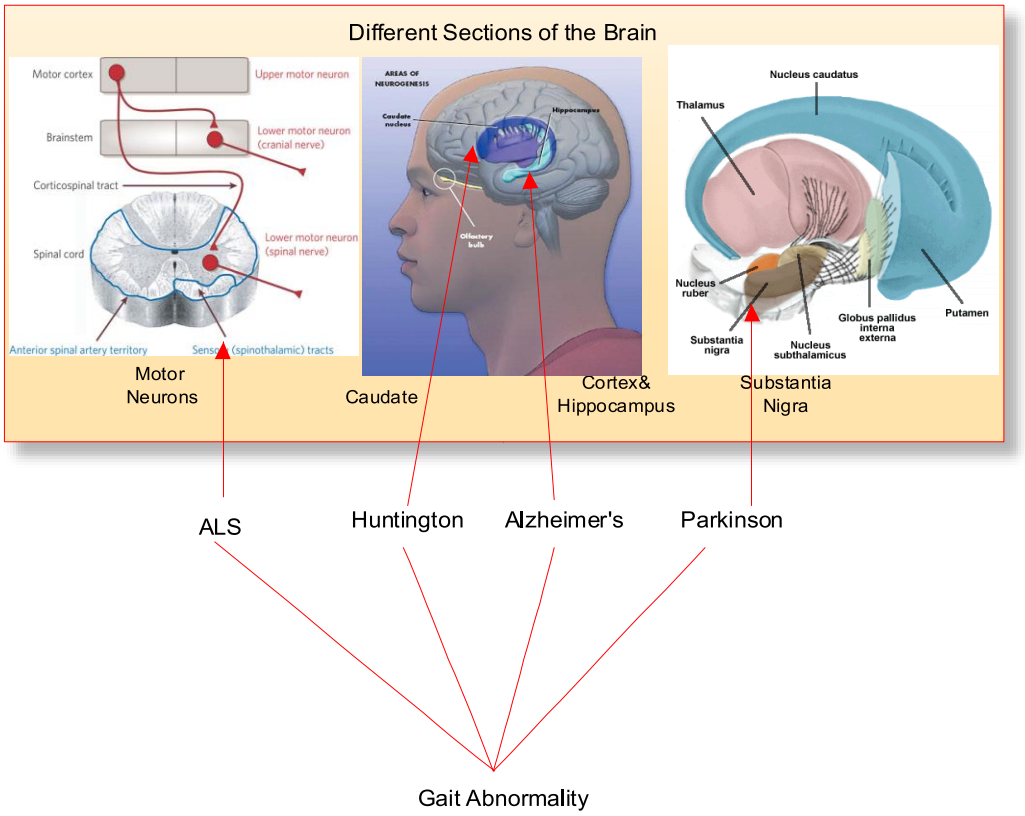


Figure 4-2: Association of Gait disorders with different NDDs.

Automatic classification of gait patterns by statistical pattern recognition techniques will help to solve this problem by discriminating different NDDs according to their data patterns. This

will save the time of doctors/practitioners not only for diagnosing exact NDDs but also their timely treatment of the disease.

4.3 Data Collection and Data Description

The data collection is the first step towards *data pre-processing* and *data classification*. Sotiris suggests the "brute force method", if an expert's analysis is not available for the selection of specific attributes or features [1]. This method encourages the *"trial and error"* process, where we test everything in the hope that the right information (attributes, features) can be isolated.

For the analysis of gait patterns, we have collected gait signals from *Physionet*[1] for NDDs patients and CO subjects. There are 16 healthy control subjects (14 Females, 2 Males) and 20 Huntington's (14 Females, 6 Males), 13 ALS (3 Females, 10 Males) and 15 Parkinson's patients (5 Females, 10 Males) with movement problems. All of them are at their final stage of the disease. We have collected data for left and right foot stride signals with time in milliseconds (total time 10 sec). Temporal parameters of the gait are measured by using force sensitive insoles that were placed in the subject's shoe. The raw data were obtained using force-sensitive resistors, with the output roughly proportional to the force under the foot. Stride-to-stride measures of footfall contact times were derived from these signals. The data were then sampled at 300 Hz by an analog-to-digital converter and stored into an ankle-worn recorder.

For the subjects with Parkinson's disease, the severity of the disease is calculated with the Hohn and Yahr score ($1.5 \leq \text{Severity} \leq 4$); a higher score indicates more advanced disease. For the subjects with Huntington's disease, the severity of the disease is calculated with the Functional Capacity Measure ($1 \leq \text{Severity} \leq 12$); a lower score indicates more advanced functional impairment. For the subjects with Amyotrophic Lateral Sclerosis, the measure here is the time since the onset of the disease ($1 \leq \text{Severity} \leq 54$). For the control subjects, an arbitrary "0" is used.

[1] www.physionet.org

Table 4-1, demonstrates the comparison of average values, fluctuation magnitude, and fluctuation dynamics of gait rhythms of control, PD, HD and ALS subjects.

Table 4-1: Gait Rhythm Dynamics [5]

Parameters	**Control**	**PD**	**HD**	**ALS**
Age (Range,yr)	20–74	44–80	29–71	36–70
Stride time, ms	1,091 ± 23	1,18 ± 30	1,138 ± 38	1,370 ± 61
Speed, ms	1.35± 0.04	1.00 ± 0.05	1.15 ± 0.008	1.02 ± 0.07
Stride time CV, (%)	2.3±0.1	4.4±0.6	7.6 ±1.2	4.5±0.6
Stride time $SD_{detrended}$, ms	27±2	52±6	120±25	65±10
α (Wilcoxon Test)	0.91±0.05	0.82±0.06	0.60±0.04	0.74±0.07
Autocorrelation	5.9±0.4	7.2±1.6	3.2±0.5	4.2±0.6
Nonstationarity index	0.67±0.02	0.64±0.03	0.54±0.03	0.69±0.05

4.4 Previous Findings in Early Detection and Signal Classifications

Scientists have proposed different methods for early detection of neurodegenerative diseases such as examining cognitive decline, using biomarkers, or through the presence of metabolites or genes [91]. However, in recent years, early detection and neuroimaging techniques, including genetic analysis, are techniques that are commonly used to detect potentially life-threatening diseases like cancer, cystic fibrosis, and neurological diseases [92]. Mini-Mental Score Evaluation (MMSE) and symptom's quantification are other well-known techniques commonly used to diagnose neurodegenerative diseases [93].

Nonetheless, the use of computer algorithms and visualization techniques are considered fundamental to support the early detection process. One example of this is the Common Spatial Patterns (CSP) algorithm proposed by Woon *et al.* [91] that has been successfully used to study Alzheimer's. CSP, which belongs to an adverse class of algorithms known as Blind Source Separation (BSS), incorporates significant properties of class labeling and dimensionality reduction. Moreover, this classification algorithm performs signal separation to rank and order the relevant separated components found within the data. This technique has already been practiced in the diagnosis of cognitive disorders like schizophrenia and depression [94]. The only problem with CSP is that while ranking and ordering the separated components, it also separates the relevant and interesting components of the signal.

Classification efficiency has been addressed by Mantzaris [95] for osteoporosis risk factor prediction with the multi-layer perceptron (MLP) and the Probabilistic Neural Networks (PNNs). MLPs are feed forward networks and work with the back-propagation learning rules and widely used in medical data processing. PNN is another type of feed forward networks consists of three layers; input layer, radial basis and a competitive layer and it works on Parzen's Probabilistic Density Function (PDF) [96]. In terms of overall performance, PNN networks perform slightly better than MLP networks. However, in his research work, the author could not address the overfitting issue of MLP with normal and pathological data patterns—some of the relationships that seem statistically significant might be due to noise.

Due to high inter-subject variability between neurodegenerative patients, from mild-to-moderate and from moderate-to-severe, it is difficult to determine the appropriate features to classify data accurately. This problem is further exacerbated when a large number of patients are used. However, Latchoumane *et al.,* have addressed this issue by analyzing EEG (electroencephalogram) signals using Multi-way Array Decomposition (MAD), which is a supervised learning process for evaluating multidimensional and multivariate data like EEG [97]. The MAD approach analyses time, frequency, and electrode signal domains simultaneously. This technique has also been used by Acar *et al.,* in studies on epileptic seizures [98].

The Parallel Factor Analysis (PARFAC) model has also been used to extract the multilinear interaction between groups, frequency, and space in EEG signals [97]. The PARAFAC model

is associated with the multilinear version of the bilinear factor models [99]. This technique is useful for analyzing spatial-frequency characteristics for correct classification of subjects.

The primary goal of such algorithms is to extract meaning from potentially huge amounts of data. In other words, to characterise features associated with particular neurodegenerative diseases. This has led to a great deal of work in feature extraction within medical datasets. One example of this is the Discrete Cosine Transform (DCT) algorithm that decreases the number of features and the computation time when processing signals [7]. DCT is used to calculate the trapped zone, under the curve, in special bands. These are described as features and used to evaluate different classifiers for neurodegenerative diseases, like Huntington's disease, Parkinson's disease and Amyotrophic Lateral Sclerosis. The results show that the Quadratic Bayes Normal Classifier is better at identifying different neurodegenerative diseases compared to others. However, they have only evaluated this approach using two feature datasets.

Similar algorithms have been used to predict heart disease using Decision Trees, Naïve Bayes and neural networks [100]. The results show that using the lift chart for prediction and non-prediction, the Naïve Bayes algorithm predicted more heart disease patients than both the Neural Network and Decision Tree approaches. While these are interesting results, only three data mining techniques were compared. A much wider study is required to determine whether other techniques work better.

Joshi *et al.,* have performed such a study where different data mining techniques are compared for the early detection of Alzheimer's [101]. Using data, collected from patients suffering with Alzheimer's, Joshi *et al.,* were able to identify the various stages of Alzheimer's using machine learning, neural networks, multilayer perceptrons, including the coactive neuro-fuzzy inference system (CANFIS) and Genetic Algorithms. The results showed that CANFIS produced the best classification accuracy result (99.55%) as compared to C4.5 (a decision tree algorithm).

Other algorithms, such as dissimilarity based classification techniques, have proven to be very useful for analyzing medical data sets. For example, algorithms, such as the k-nearest neighbour classifier (k-NN), and Linear and Quadratic normal density based classifiers, have been extensively used to classify seismic signals [102]. Nonetheless, the results have shown

that Bayesian (normal density based) classifiers outperform the k-NN classifier, when a large number of prototypes are provided.

While these approaches provide obvious benefits, current applications for classifying medical data are still lacking consistency in terms of revealing hidden significant information, especially from real-time clinical data. The main limitation with the approaches described is that they only consider a small number of classifiers. Furthermore, many of them fail to include relevant and important features, such as age and gender that can have a significant impact on results. Moreover, overall accuracy depends on a single set of variables while other variables could potentially have more impact on the performance evaluation [103].

The approach posited in this paper considers all renowned classification algorithms and uses a large-scale feature set. Each variable in the array has its own significant relationship with the progression of specific diseases. Moreover, rather than relying on base-level classifiers, a new strategy is described based on the fusion of classifiers in Chapter 5. In this way, it is possible to explore any new dimensions that may emerge from the results.

4.5 Data Pre-processing

Once the data is collected, the next step is "*data preprocessing",* before analysing and evaluating the data. Incomplete, inaccurate and contaminated data analysis can lead to inappropriate and below quality results. Therefore, a crucial and primary task is to identify the limitations and insufficiencies of the datasets.

Although, Physionet is a *NIH (National Institute of Health Sciences, USA)* funded, reliable online data repository that researchers and medical doctors are using since 1999, yet aimed at confirming the quality of the data before starting our project. We have contacted the administration of Physionet and got a positive reply from *George B. Moody* (Harvard-MIT Division of Health Sciences and Technology, Cambridge, USA). He claimed that particular precautionary measures are considered before collecting the data to make it noise free but some datasets (especially for ambulatory subjects) need some preprocessing steps to clean them further.

Our concern is with a database containing gait data for different neurodegenerative diseases (PD, HD, and ALS) and also for control subjects. Carefully visualizing the datasets, we find the following limitations that needed to be considered before applying classification algorithms:

- Number of subjects are different in each group—Imbalanced Datasets
- Missing observations in the datasets for few subjects—Handling Missing Entries
- Extracting meaningful and relevant features that can act as a biomarker for the early detection of NDDs— Feature Extraction

4.5.1 Imbalanced Datasets

The following section helps us to understand the issues related to imbalanced datasets and their possible solutions. It also highlights the importance of using re-sampling techniques in the field of medicine.

- Issues with Imbalanced Datasets

Learning from imbalanced datasets is an important and controversial topic that is addressed in our research work. These kinds of datasets usually generate biased results [104]. For instance, imagine a medical dataset with 50 true negative values (majority class) and 20 true positive values (minority class). If half is selected for training and the remainder for testing (25 healthy and 10 sick persons), we find that the accuracy is 90%. The result suggests that the classifier performs reasonably well. However, what happens, when all the negative values are accurately identified (healthy persons) and only 5 out of the 10 positive values (sick persons) are classified correctly. In this situation, the classifier is more sensitive to detecting the majority class patterns but less sensitive to detecting the minority class patterns. This is caused because the training data is imbalanced. In other words, the classifier concludes that 5 out of the 10 unhealthy people are healthy when this is not the case. These kinds of results ultimately cause more destruction if data comes from real time environments, such as biomedical, genetics, radar signals, intrusion detection, risk management and credit card scoring [18].

- Previous Findings in Imbalanced Datasets

Advocating the resampling technique, Xiong *at al.*, [105] narrated that training a classifier with an imbalanced positive and negative dataset in machine learning results in poor classification performance. To classify the horizontal gene transfer (HGT) for the detection of microbial genome diversification, they selected Synthetic Minority Over-sampling Technique (SMOTE) to generate more patterns for HGT in genome. They get remarkably less mean error rate using SVM classifier as compared to the previous findings. Working on the same line, using SMOTE for oversampling and SVC for classification, Tao *et al.*, [106] regenerated Curvelet-transformation textural features together with morphological features to classify the patients with lung cancer. Results revealed that accuracy based on cross-evaluation for the original unbalanced data and balanced data was 80% and 97%, respectively.

Majid and Andreas [107] used Synthetic Protein Sequence Oversampling (SPSO) method to create protein sequences of the minor class and get better accuracy and Matthew's correlation coefficient than imbalanced datasets. Chia-Yun *et al.*, in their paper [108], claimed that the ability of the predictive modeling methods is adversely affected if the datasets are imbalanced. They recommend the oversampling method to overcome the overfitting of the classifier (Support Vector Machine) to classify the compounds for cytotoxicity with respect to the Jurkat cell line. Compared to previous results in the literature, the SVM models built from oversampled data sets exhibited better predictive abilities for the training and external test sets.

Similarly, Xuan *et al.*, [109] suggested a novel method of over-sampling the imbalanced datasets called safe-SMOTE, for the classification of two gene expression datasets of cancer, i.e., colon-cancer and leukemia. They showed that the sensitivity with the oversampling technique increased from 81.82% to 90.50%. Also, the G-mean value of the control increased from 85.45% to 86.04%.

- Possible Solutions of skewed datasets

In order to solve the imbalanced dataset problem it is necessary to resample datasets. Different resampling techniques are available to achieve this, that include *under sampling*

and *over sampling* [110]. Under sampling is a technique where we reduce the number of patterns within the majority class dataset to make it equivalent to other classes. In over sampling, more data is generated within the minority class.

- Re-Sampling of NDDs Datasets

As already mentioned that the dataset, we analyse in this research work, has 16 CO, 15 PD, 20 HD, and 13 ALS subjects. Given the number of data subjects, there is already a small number of subjects in each database; hence under sampling is not a good idea for such datasets.

To go for the over sampling technique, we note the upper and lower limit of the foot stride intervals for all subjects in a dataset, for left and right feet gait data. First, the *mean* is taken for the right and left feet gait rhythm in term of stride interval (time from initial contact of one foot to the subsequent contact of the same foot), of each subject in one dataset. Then the maximum and minimum rage of right and left feet motion values are calculated to find the upper and lower limit. We generate random values of motion vectors between those minimum and maximum values for each database (CO, PD, and ALS) in order to get equal data patterns in each class, i.e., 20 per class (equal to HD—maximum number of subjects in this database). Now we have 20 subjects in each class of neurodegenerative diseases.

4.5.2 Feature Extraction

To avoid the ambiguity of noisy data and also to remove the missing observations from the datasets, it is crucial to extract the relevant and appropriate information from the datasets. We take the mean of all 3000 stride intervals of 10 sec. for left and right feet strides of each subject in one dataset for all four databases. The other features that we extract from the datasets are age, height, weight, speed, time, and BMI factor.

Figure 4-3, demonstrates a collection of prior knowledge (age, gender, height, weight, BMI, walking speed and time) with empirical knowledge (sensory measurements for right and left feet signals) to get a posterior knowledge to recognize gait patterns of a diseased and healthy person.

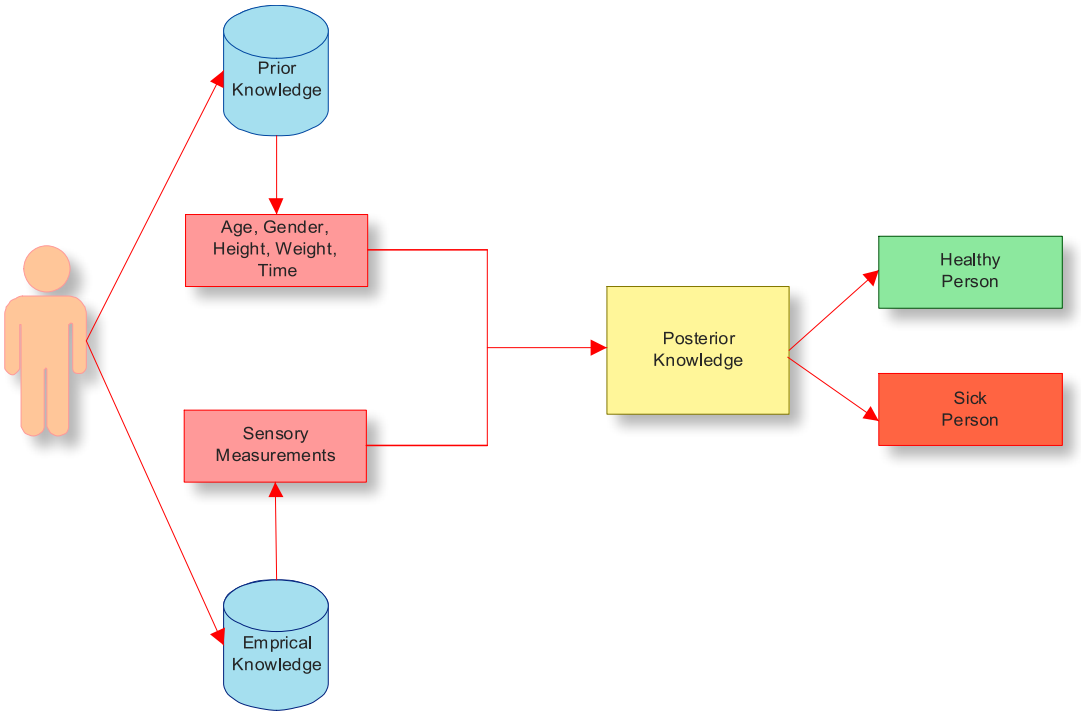


Figure 4-3: Extraction of Posterior Knowledge from Prior and Empirical Knowledge

A set of eight features that are used to classify the gait data of neurodegenerative diseases and healthy persons is given in Table 4-2. It shows the *mean* of maximum and minimum vales of all features in 4 datasets.

Table 4-2: A Set of Eight Feature used for Classification

Right Feet Signals	Left Feet Signals	Age	Height	Weight	Time	Walking Speed	BMI
Motion Vector Stride time in ms (-0.6739-0.5411)	Motion Vector Stride time in ms (-0.5421-0.2069)	20-80 (Years)	1.57-2.13 (Meters)	40.82-117.5 (Kg)	10 Sec	0.5-1.82 (m/sec)	14.4-37.1 (weight(kg) / height²(m²))

Similarly, Table 4-3, Table 4-4, Table 4-5, and Table 4-6, provides detail information of datasets for CO, HD, PD and ALS subjects respectively.

Table 4-3: Dataset for CO Subjects with Extracted Features

Right Feet Signals	Left Feet Signals	Age	Height	Weight	Time	Walking Speed	BMI
Motion Vector Stride time in ms (-0.6739-0.5411)	Motion Vector Stride time in ms (-0.5421-0.2069)	20-74 (Years)	1.67-1.94 (Meters)	50-95 (Kg)	10 Sec	0.91-1.54 (m/sec)	14.9-25.2 (weight(kg) / height²(m²))

Table 4-4: Dataset for HD Subjects with Extracted Features

Right Feet Signals	Left Feet Signals	Age	Height	Weight	Time	Walking Speed	BMI
Motion Vector Stride time in ms (-0.5576-0.6634)	Motion Vector Stride time in ms (-0.1.750-0.5834)	29-71 (Years)	1.57-2 (Meters)	45-102(Kg)	10 Sec	0.56-1.82 (m/sec)	16.2-32.2 (weight(kg) / height²(m²))

Table 4-5: Dataset for PD Subjects with Extracted Features

Right Feet Signals	Left Feet Signals	Age	Height	Weight	Time	Walking Speed	BMI
Motion Vector Stride time in ms (-0.9942-0.3693)	Motion Vector Stride time in ms (-0.8065-0.78)	44-80 (Years)	1.67-2.13 (Meters)	43-100 (Kg)	10 Sec	0.5-1.33 (m/sec)	14.5-26.6 (weight(kg) / height²(m²))

Table 4-6: Dataset for ALS Subjects with Extracted Features

Right Feet Signals	Left Feet Signals	Age	Height	Weight	Time	Walking Speed	BMI
Motion Vector Stride time in ms (-0.9958-0.1645)	Motion Vector Stride time in ms (-1.1159-0.2155)	36-70 (Years)	1.57-1.88 (Meters)	40.82-117.5 (Kg)	10 Sec	0.77-1.302 (m/sec)	16.6-37.1 (weight(kg) / height²(m²))

Before features classification, we have analysed the gait signals, to find out any relation of gait of a person with neurological disturbances. Neurophysiological changes that are associated with aging, affect the locomotor system's ability to generate stride-intervals correlations. To test the hypothesis, stride intervals correlation with neurological functions that altered with aging, we have analyzed the gait signals of control healthy persons and persons with neurological diseases. Computation of left and right foot stride signals revealed that the duration of gait cycle fluctuates from one stride to the next in a complex fashion. Figure 4-4, shows some "noisy" variations with the stride signals of diseased person that present some fractal property while this variability is not attributed with the normal gait of

the person, Figure 4-5. Nonetheless, these fluctuations in movement signals may also appear with normal aging which indicates neurological changes in the brain of healthy persons as well, but not due to degeneration process.

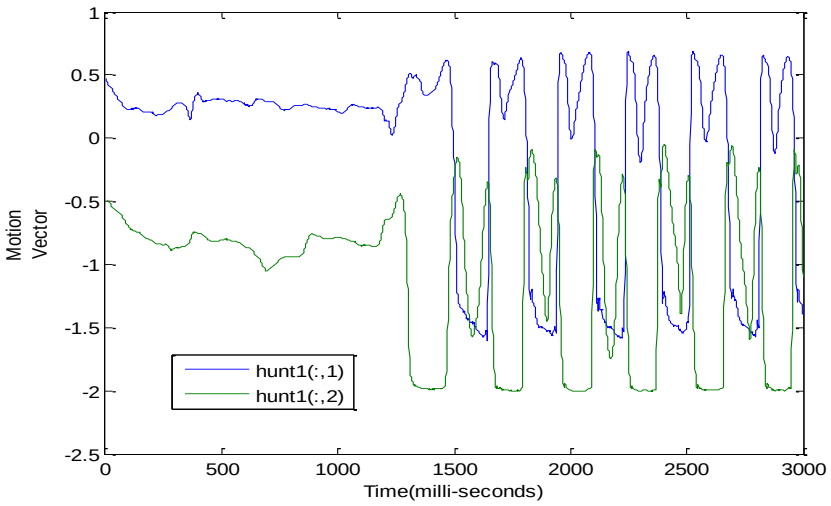


Figure 4-4: Signal Analysis for a Neurodegenerative Person

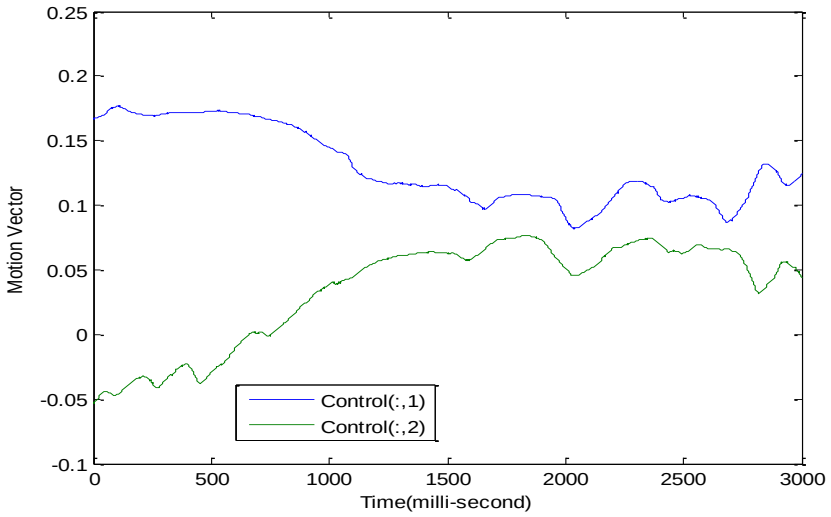


Figure 4-5: Signal Analysis for a Healthy Person

4.6 Signal Classification

The dataset containing the eight features described in the previous section provides the feature set required to diagnose neurodegenerative diseases accurately. More specifically this dataset is used to *select a classifier*, *train it*, *test it* and finally *evaluate* the result to determine if the correct classification is performed.

The computation is directly proportional to the number of features considered in the dataset. Figure 4-6, demonstrates a *scatter plot* using only three selected features and shows the complexity of classification of gait patterns for each subject. In this instance, *Feature 1* and *Feature 2* are associated with "right and left foot movement signals" while *Feature 3* represents "age" which is considered an important factor in disease progression.

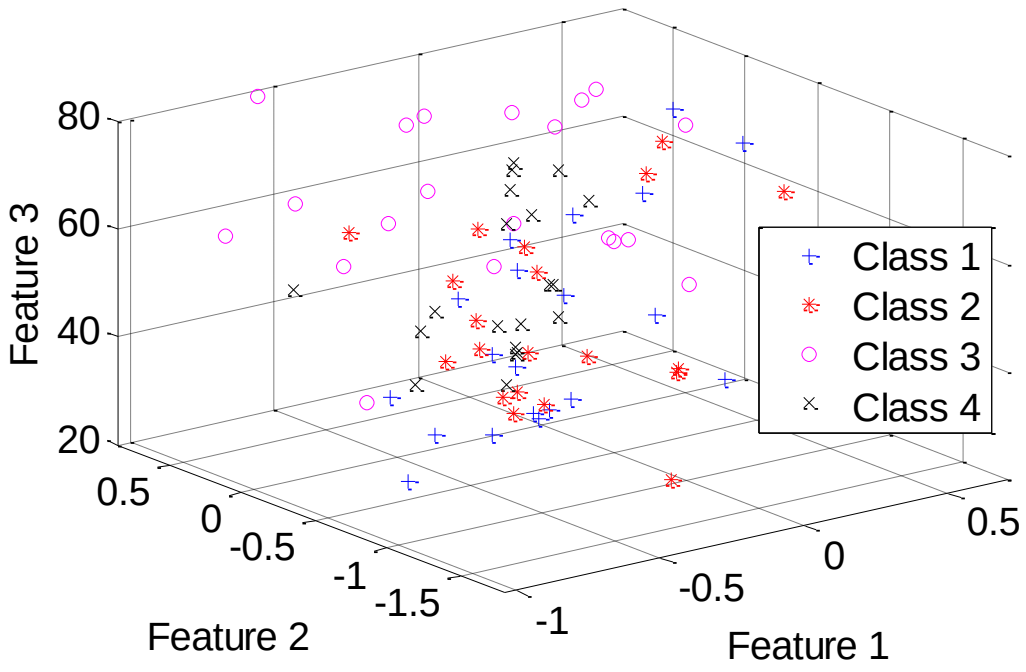


Figure 4-6: 3-Dimensional Scatter Plot of Selected Features

Using the defined feature set, several classifiers have been evaluated for consideration in the final *classifier fusion strategy*. The principle goal is to use classifiers that perform the best. Each class is labelled with a specific *label* before classification. In this particular case, class

of healthy subjects is labelled with “Class 1”, similarly HD, PD and ALS classes are labelled with “Class 2”, “Class 3” and “Class 4” respectively.

The classifiers considered are the, Linear Discriminant Classifier (ldc), Quadratic Discriminant Classifier (qdc) and the Quadratic Bayes Normal Classifier (udc) for density based classification. For Linear Classification, an additional four classifiers are selected, which are the Logistic linear (loglc), Fisher’s (fisherc), Nearest Means (nmc) and the Polynomial (polyc). A linear classifier predicts the class labels based on a weighted linear combination of features or the pre-defined variables. The Parzen (parzenc), Decision Tree (treec), Support Vector Machine (svc) and k-Nearest Neighbour (knnc) classifiers have been selected for non-linear classification of our datasets.

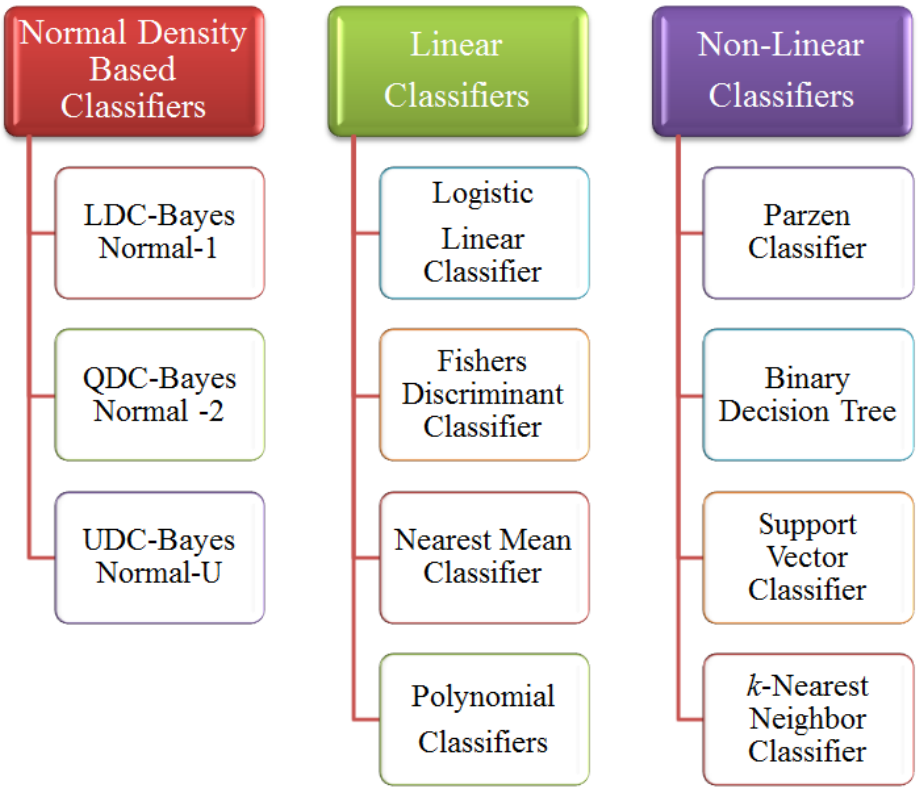


Figure 4-7: List of 11 Classifiers used from PRTools [68]

First, we apply all the Bayes classifiers on our dataset. Figure 4-8, shows the mapping of all these mentioned classifiers on scattered plot of 2-features’ dataset. The black line indicated the linear discriminant classifier (*LDC*) which is denoted by Bayes Normal-1 in the figure. The red line indicates the Bayes Normal-2, which is quadratic discriminant classifier (*QDC*).

Similarly, the Bayes Normal-3 which represents quadratic classifier for uncorrelated densities (*UDC*) is denoted in by blue line.

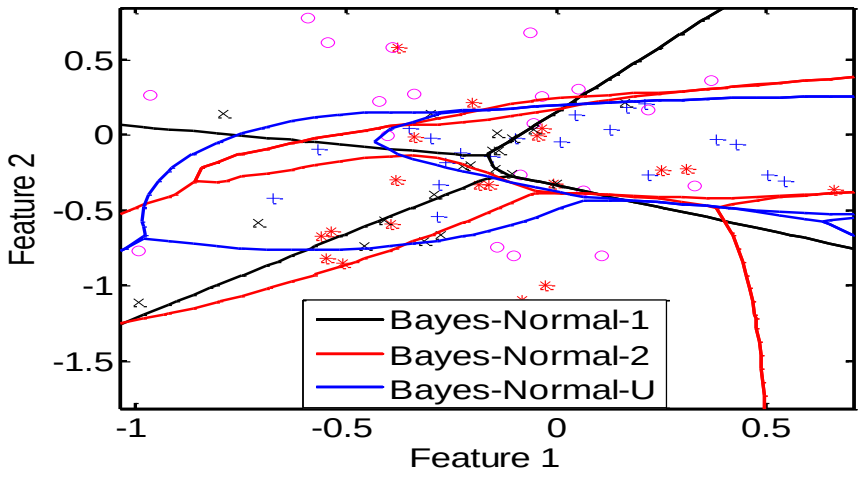
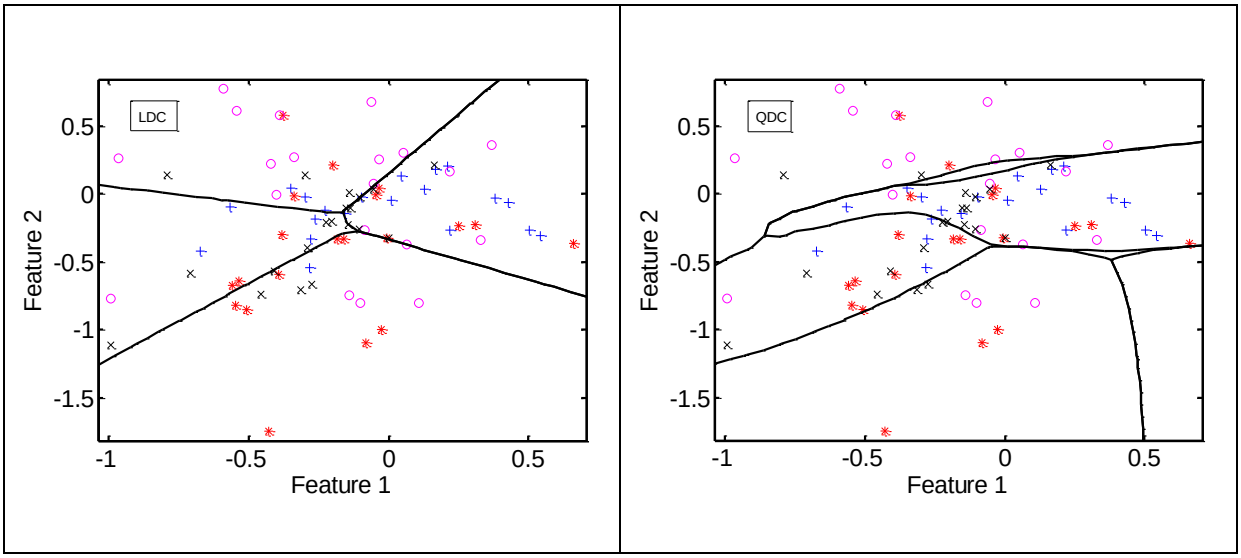
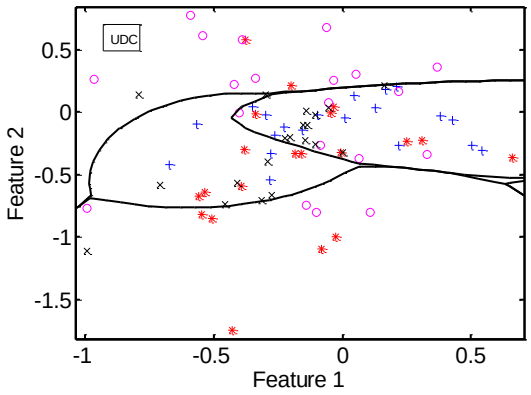


Figure 4-8: Class Mapping for Bayes Normal Classifiers

Table 4-7, shows the mapping results of each individual classifier (*LDC, QDC,UDC*) that belong to Bayes classifiers separately on a dataset of gait signals for neurodegenerative patients and healthy persons.

Table 4-7: Class Mappings of Individual Bayes Normal Classifier





A set of linear classifiers which consists of Logistic linear (loglc), Fisher's (fisherc), Nearest Means (nmc) and the Polynomial (polyc) is applied on the data set. Figure 4-9, shows the mapping of all these mentioned classifiers on a scattered plot of 2-features dataset. In the figure the black line shows the logistic linear classifier, the red line indicates the fisher classifier, the blue line to nearest mean classifier while the pink line to polynomial classifier.

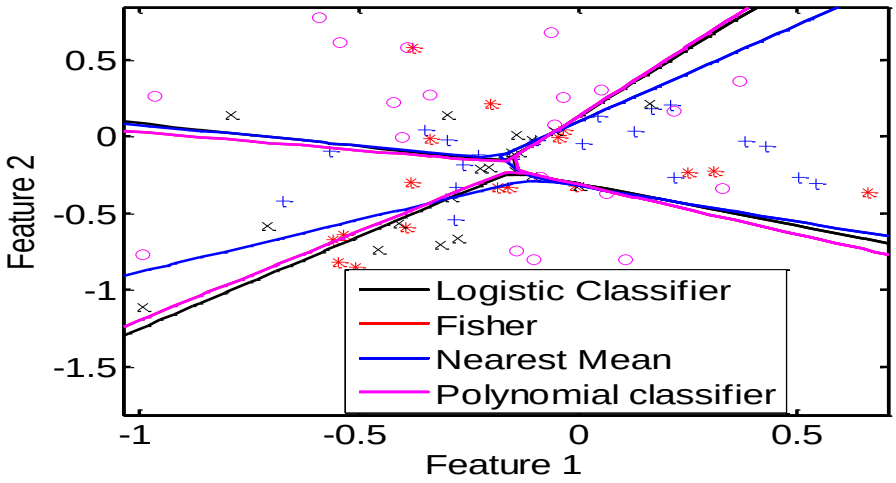


Figure 4-9: Class Mapping for Linear Classifiers

Table 4-8: Class Mappings of Individual Linear Classifier

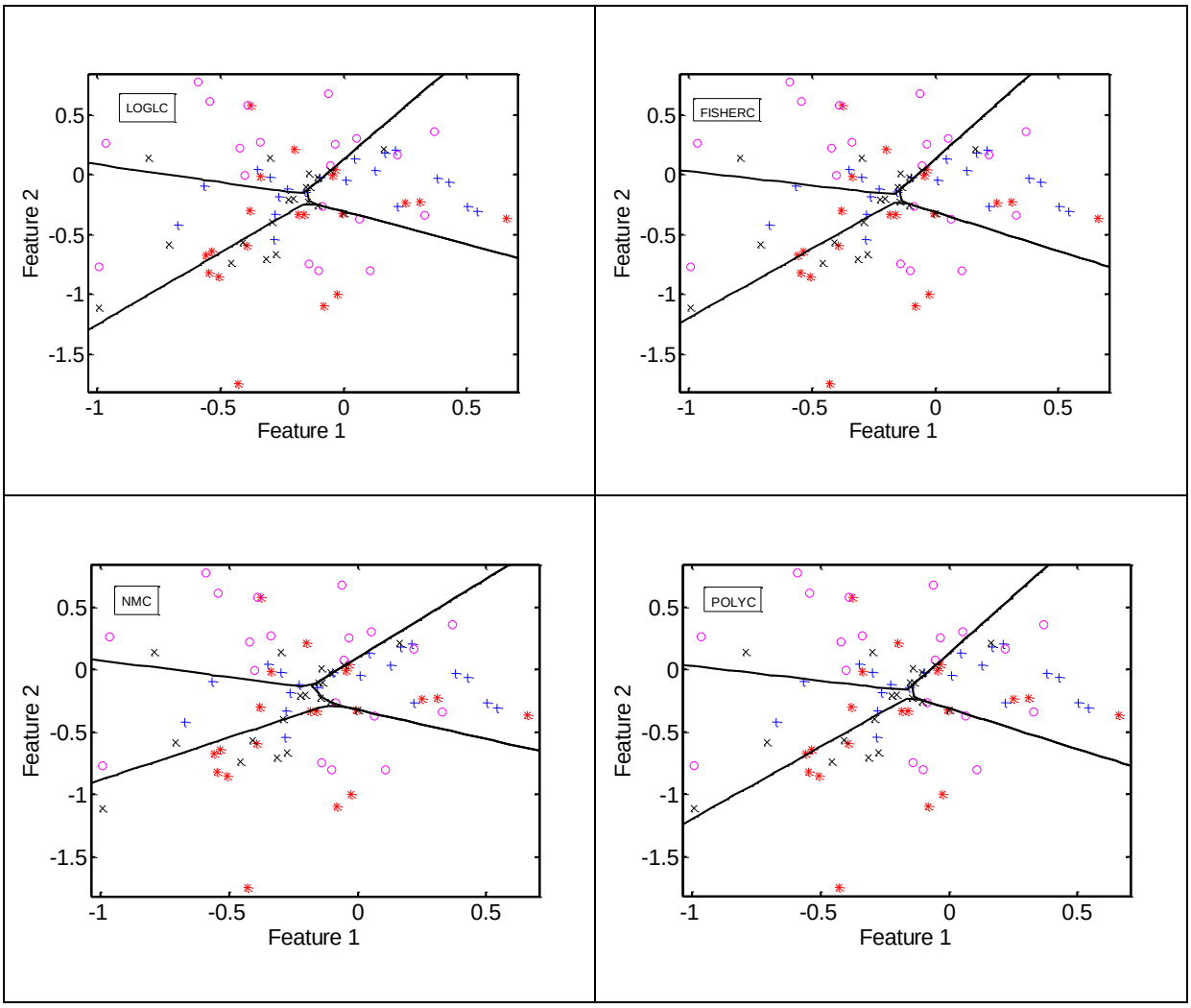


Table 4-8, shows the mapping results of each linear classifier (*LOGLC*, *FISHERC*, *NMC*, *POLYC*) separately on a dataset of gait signals for neurodegenerative patients and healthy persons.

Similarly, a set of nonlinear classifiers, Parzen (parzenc), Decision Tree (treec), Support Vector Machine (svc) and k-Nearest Neighbour (knnc), is applied on the dataset. Figure 4-10, shows the mapping of non-linear classifiers on a scattered plot of 2-feature's dataset.

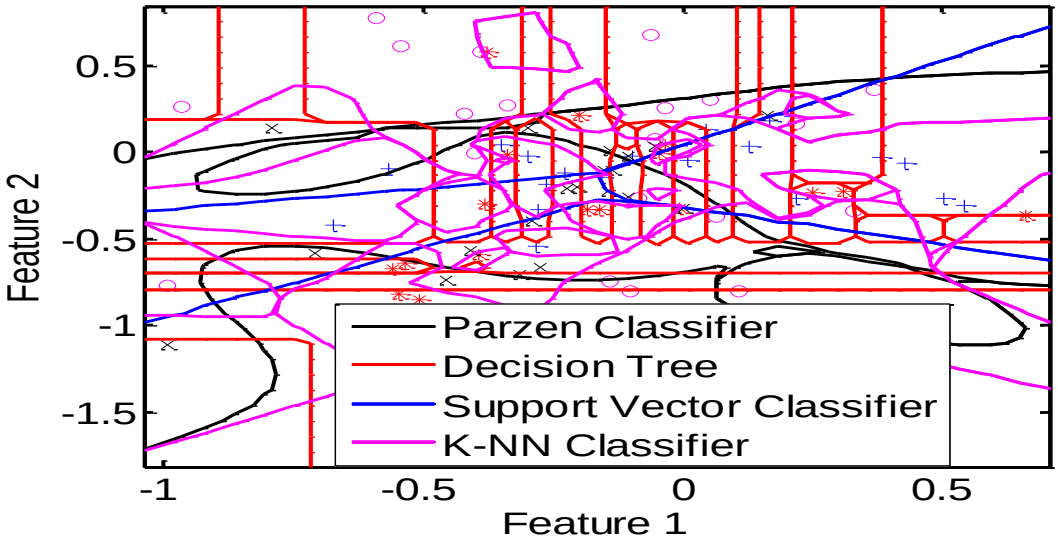
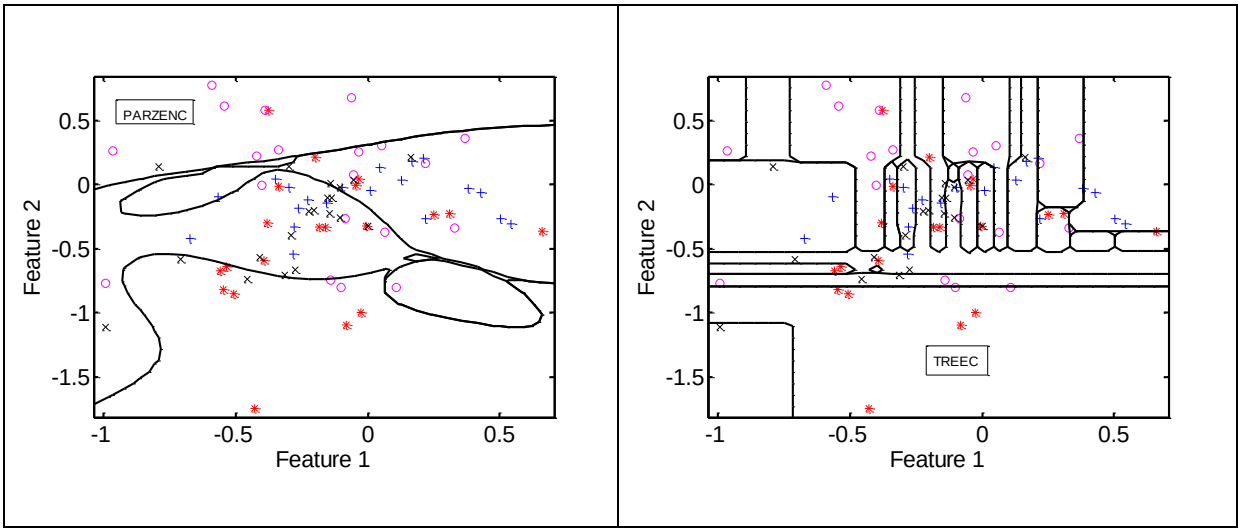
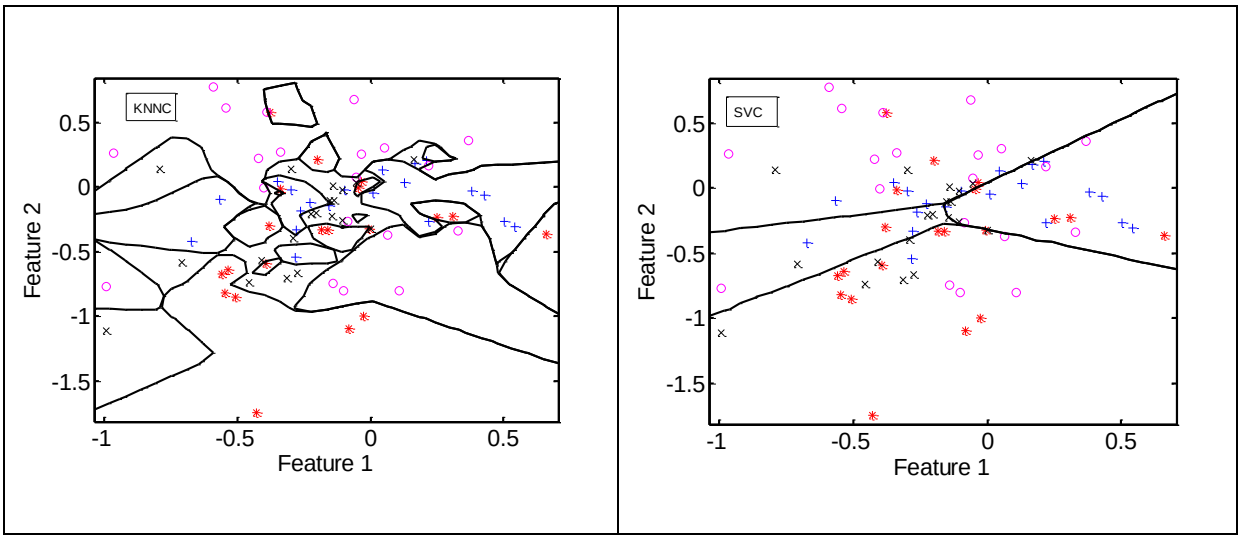


Figure 4-10: Class Mappings for Non-linear Classifiers

Table 4-9, shows the mapping results of each nonlinear classifier (*PARZENC*, *TREEC*, *SVC*, *KNNC*) separately on a dataset of gait signals for neurodegenerative patients and healthy persons.

Table 4-9: Class Mappings of Individual Nonlinear Classifiers





4.7 Performance Evaluation and Results

Performance evaluation of a classifier is mostly done by a parameter called decision threshold t ($0 \leq t \leq 1$), which decides the final class membership of a given object [111]. A class with higher posterior probability of this threshold is assigned to a particular object. This threshold value may vary for imbalanced datasets or multiclass datasets.

State-of-the-art study reveals that one particular performance measure may evaluate a classifier on a single perspective while fails to measure on another [112]. Although researchers have been evaluating classification algorithms by various techniques, yet there is no single authorized criterion that outperforms others.

More specifically, we are presenting two different kinds of measures to demonstrate and then to compare the performance evaluation results:

A. *Visualization:* representing the possible outcome of true and false values of a classifier in the form of graphs; Reject and ROC curves;
B. *Statistical Analysis:* to compare the evaluation results by mathematical formulas such as classification accuracy (Confusion Matrix), Precision, Recall, Sensitivity, Specificity and F-Measure.

***Confusion Matrix*:** determines the distribution of errors across all classes [113]. The estimate of the classifier is calculated as the trace of the matrix divided by the total number of entries. The formula is given below:

$$\text{Confusion Matrix} = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision: is a function of true positive and the objects that are misclassified as positive *i.e.* false positive.

$$Precision = \frac{TP}{TP + FP}$$

***Recall/Sensitivity and Specificity*:** Recall presents a function of correctly classified object *i.e.* true positive and the objects that are classified incorrectly *i.e. false negative*. Specificity describes the results in term of true *negative values*.

Both *Recall and Precision* are relevant to each other. *Precision* is the fraction of retrieved information relevant to the search while *Recall* is the fraction of the information related to search query that is retrieved successfully [114]. The formulas are given below:

$$Recall/Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

F-Measure: is another common evaluation metrics that combines precision and recall into a single value. The formula is as:

$$F - Measure = 2X\frac{RecallxPrecision}{Recall + Precision}$$

ROC and Reject Curve: Here, we have visualized Receiver Operating Curve (ROC) for Error Type I and Error Type II [115]. The curve is drawn for "False Positive" and "False Negative" values. Another common approach is the Reject curve which works on reducing the error cost by turning them into a rejection [116]. In this case, the objects close to the decision boundaries are not classified.

$$Type\ I\ Error = \frac{FN}{TP + FN}$$

$$Type\ II\ Error = \frac{FP}{TN + FP}$$

The confusion matrix results (in terms of accuracy) are described in Table 4-10, Table 4-11, Table 4-12, for Bayes Normal, linear and nonlinear classifiers respectively. They present the screen shorts of the confusion matrix along with truly classified subjects and their accuracy percentage.

Table 4-10: Results of Confusion Matrix for Bayes Normal Classifiers

Bayesian Normal Density Based Classification			
Classifiers	**Confusion Matrix**	**Truly Classified**	**Accuracy (%)**
Linear Discriminant Analysis (ldc)	True Labels | Estimated Labels 1 2 3 4 | Totals 1 | 7 1 1 1 | 10 2 | 3 4 2 1 | 10 3 | 0 1 8 1 | 10 4 | 1 1 2 6 | 10 Totals | 11 7 13 9 | 40	Healthy=7 Huntington=4 Parkinson=8 ALS=6	7+4+8+6=25 25/40=62.5%
Quadratic Discriminant Analysis (qdc)	True Labels | Estimated Labels 1 2 3 4 | Totals 1 | 2 3 1 4 | 10 2 | 2 5 2 1 | 10 3 | 0 0 9 1 | 10 4 | 0 1 2 7 | 10 Totals | 4 9 14 13 | 40	Healthy=2 Huntington=5 Parkinson=9 ALS=7	2+5+9+7=23 23/40=57.5%
Quadratic Bayes Normal Classifier (udc)	True Labels | Estimated Labels 1 2 3 4 | Totals 1 | 7 1 2 0 | 10 2 | 1 5 3 1 | 10 3 | 0 1 9 0 | 10 4 | 0 2 3 5 | 10 Totals | 8 9 17 6 | 40	Healthy=7 Huntington=5 Parkinson=9 ALS=5	7+5+9+5=26 26/40=65%

Table 4-11: Results of Confusion Matrix for Linear Classifiers

Linear Classification			
Classifiers	**Confusion Matrix**	**Truly Classified**	**Accuracy (%)**
Logistic Linear Classifier (loglc)	True Labels | Estimated Labels 1 2 3 4 | Totals 1 | 4 4 0 2 | 10 2 | 1 3 3 3 | 10 3 | 0 1 7 2 | 10 4 | 3 0 0 7 | 10 Totals | 8 8 10 14 | 40	Healthy=4 Huntington=3 Parkinson=7 ALS=7	4+3+7+7=21 21/40=52.5%
Fisher's Discriminant (Minimum Least Square-fisherc)	True Labels | Estimated Labels 1 2 3 4 | Totals 1 | 6 3 1 0 | 10 2 | 2 4 1 3 | 10 3 | 0 3 5 2 | 10 4 | 2 0 0 8 | 10 Totals | 10 10 7 13 | 40	Healthy=6 Huntington=4 Parkinson=5 ALS=8	6+4+5+8=23 23/40=57.5%

Nearest Mean Classifier (nmc)	True | Estimated Labels Labels | 1 2 3 4 | Totals 1 | 4 3 3 0 | 10 2 | 2 3 1 4 | 10 3 | 0 2 7 1 | 10 4 | 2 3 2 3 | 10 Totals | 8 11 13 8 | 40	Healthy=4 Huntington=3 Parkinson=7 ALS=3	4+3+7+3=17 17/40=42.5%
Polynomial Classifier (polyc)	True | Estimated Labels Labels | 1 2 3 4 | Totals 1 | 6 2 1 1 | 10 2 | 3 3 0 4 | 10 3 | 0 4 6 0 | 10 4 | 0 0 2 8 | 10 Totals | 9 9 9 13 | 40	Healthy=6 Huntington=3 Parkinson=6 ALS=8	2+3+6+8=23 23/40=57.5%

Table 4-12: Results of Confusion Matrix for Nonlinear Classifiers

Nonlinear Classification			
Classifiers	**Confusion Matrix**	**Truly Classified**	**Accuracy (%)**
Parzen Classifier (parzenc)	True | Estimated Labels Labels | 1 2 3 4 | Totals 1 | 7 2 0 1 | 10 2 | 3 5 1 1 | 10 3 | 1 0 8 1 | 10 4 | 2 3 1 4 | 10 Totals | 13 10 10 7 | 40	Healthy=4 Huntington=3 Parkinson=7 ALS=7	7+5+8+4=24 24/40=60%
Binary Decision Tree Classifier (treec)	True | Estimated Labels Labels | 1 2 3 4 | Totals 1 | 5 1 2 2 | 10 2 | 4 2 3 1 | 10 3 | 4 3 1 2 | 10 4 | 1 1 4 4 | 10 Totals | 14 7 10 9 | 40	Healthy=6 Huntington=4 Parkinson=5 ALS=8	5+2+1+4=12 12/40=30%
Support Vector Classifier (SVC)	True | Estimated Labels Labels | 1 2 3 4 | Totals 1 | 5 0 1 4 | 10 2 | 4 1 1 4 | 10 3 | 1 0 8 1 | 10 4 | 1 0 3 6 | 10 Totals | 11 1 13 15 | 40	Healthy=4 Huntington=3 Parkinson=7 ALS=3	5+1+8+6=20 20/40=50%
k-nearest Neighbor Classifier (knnc)	True | Estimated Labels Labels | 1 2 3 4 | Totals 1 | 7 1 1 1 | 10 2 | 4 1 3 2 | 10 3 | 5 0 4 1 | 10 4 | 2 5 2 1 | 10 Totals | 18 7 10 5 | 40	Healthy=6 Huntington=3 Parkinson=6 ALS=8	7+1+4+1=13 13/40=32.5%

The results reveal that some of the classifiers have shown comparatively lower error rate especially the Quadratic Classifier for uncorrelated variables (*UDC*), which belongs to the group of normal density based classification techniques (Bayes Normal Classifiers). *UDC* has classified *7/10 healthy, 5/10 Huntington, 9/10 Parkinson and 5/10 ALS* correctly. Three classifiers that have given comparatively better results are *LDC*, *UDC* and *PARZEN* with 62.5%, 65%, and 60% accurate results, respectively. The accuracy of each classifier is represented as a percentage and is illustrated in Figure 4-11.

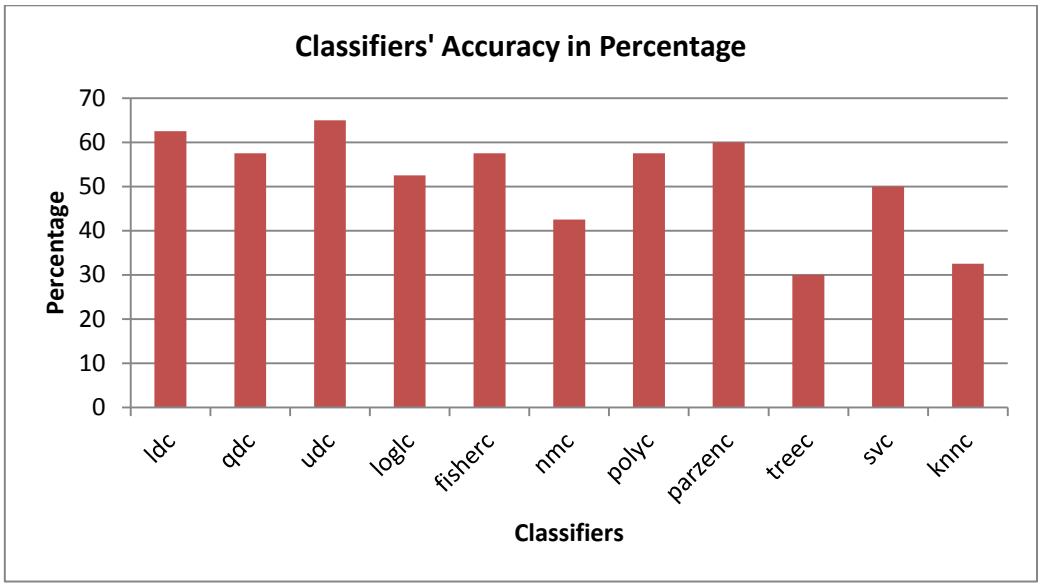


Figure 4-11: Classification Accuracy of Classifiers Tested

Furthermore, other evaluations measures are computed to analyse results from multiple perspectives. From now onwards, more focus will be on those classifiers that have given us better results i.e., LDC, UDC and PARZEN.

Table 4-13: Results of Various Evaluation Techniques

	Quadratic Bayes Normal Classifier (udc)	Linear Discriminant Classifier (ldc)	Parzen Classifier (parzenc)
Confusion Matrix	$\frac{7+5+9+5}{40}$ 65%	$\frac{7+4+8+6}{40}$ 62.5%	$\frac{7+5+8+4}{40}$ 60%
Precision	$\frac{5+9+5}{(5+9+5)+(3)}$ 86.36%	$\frac{4+8+6}{(4+8+6)+(3)}$ *exactly* 85.71%	$\frac{(3+7+7)}{(3+7+7)+(3)}$ 85%
Recall/Sensitivity	$\frac{5+9+5}{(5+9+5)+(5+1+5)}$ 63%	$\frac{4+8+6}{(4+8+6)+(6+2+4)}$ 60%	$\frac{5+8+4}{(5+8+4)+(7+3+3)}$ 56%
Specificity	$\frac{7}{7+3}$ 70%	$\frac{7}{7+3}$ 70%	$\frac{4}{4+6}$ 40%
F-Measure	$2X\frac{0..5418}{1.49}$ 72.72%	$2X\frac{0..51}{1.45}$ 70%	$2X\frac{0.476}{1.41}$ 67%

Type I Error	$\frac{5+1+5}{(5+9+5)+(5+1+5)}$ 36.6%	$\frac{6+2+4}{(4+8+6)+(6+2+4)}$ 40%	$\frac{7+3+3}{(3+7+7)+(7+3+3)}$ 43.3%
Type II Error	$\frac{3}{7+3}$ 30%	$\frac{3}{7+3}$ 30%	$\frac{6}{4+6}$ 60%

Figure 4-12, shows the results of the *ROC analysis* for the base-level classifiers, where the "Quadratic Normal Bayes Classifier" shows the least error rate compared to all other classifiers. In this case *Error I*, represents the "False Positive" values, while *Error II* presents the "False Negative" Values that show the system's failure to predict any disease and label the objects as healthy persons. As it can be noticed from Figure 4-12, the uncorrelated Quadratic Bayes Normal classifier generated less errors and produced better classification when benchmarked with the Bayes Normal-1 and Parzen Classifiers. This is because the Quadratic Bayes Normal classifier (Bayes Normal-U) uses uncorrelated variables.

Similarly in Figure 4-13, *Reject Curve* shows the least error rate for Quadratic Bayes Normal classifier (Bayes Normal-U), i.e., 0.3 while the error rate for Bayes Normal-1 and Parzen Classifiers is 0.45 and 0.5 respectively.

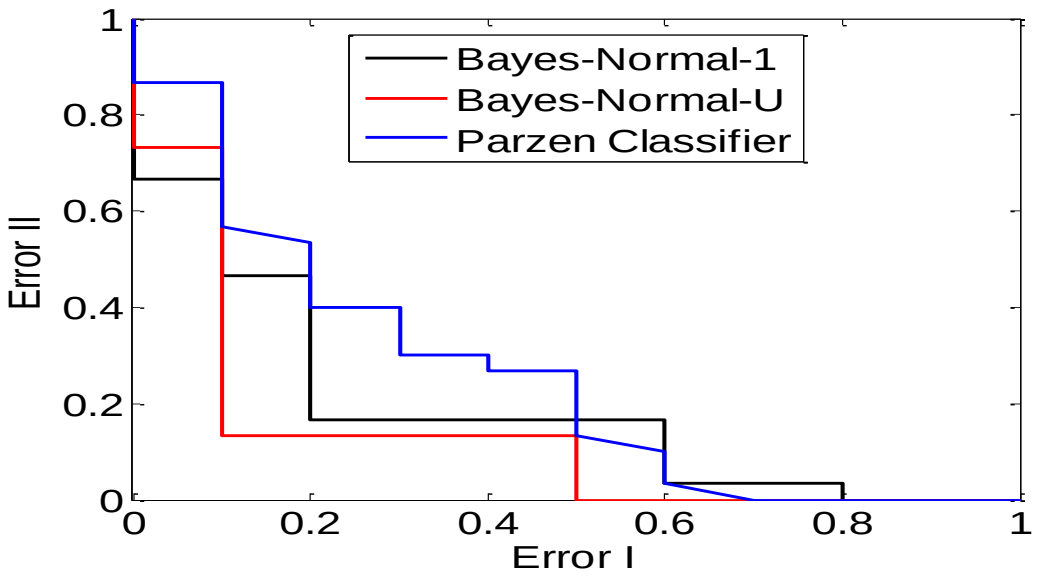


Figure 4-12: Receiver Operating Curve for Classifier's Evaluation

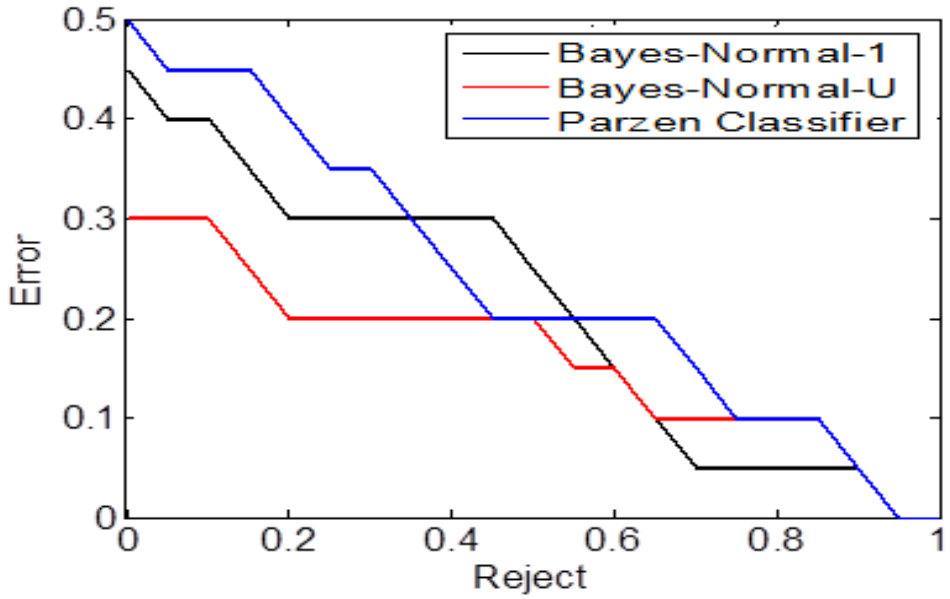


Figure 4-13: Reject Curve for Classifier's Evaluation

4.8 Results Discussion

As already mentioned, a set of 11 classifiers is tested that belong to different categories of classifiers. Three classifiers have given better results. Two of them belong to Bayes Normal classifiers; *UDC and LDC* while one belongs to nonlinear classifiers; *PARZEN*. Linear classifiers, surprisingly, could not provide better results. There might be two reasons behind it 1) linear classification is not a good option for multiclass datasets, 2) our datasets have a lot of inter subject variations. This leads to overlapping class distributions with in a feature space and ultimately gives a higher misclassification rate as with the case of *LOGLC, FISHERC, NMC* and the *POLYC*.

Given the nature of our database (multiclass datasets), it is very hard to compare the results with the previous findings, where most of the time, the analysis is for 2-class datasets. However, we elaborate our results from different perspectives. The *accuracy rate* (calculated from *confusion matrix*) is 65%, 62.5% and 60% for *UDC, LDC and PARZEN* classifiers.

Specifically speaking in term of sensitivity and specificity, it is actually quite tricky to calculate *true positive* and *false positive* values because of three different neurodegenerative

diseases. The sensitivity of UDC in term of (CO-HD), (CO-PD), and (CO-ALS) is 90%, 50% and 50% respectively. While the overall sensitivity for all 4-class datasets is 63%. Similarly, for LDC it is 40%, 80% and 60 for (CO-HD), (CO-PD), and (CO-ALS) respectively. Nonetheless, the overall percentage is 60%. In case of PARZEN, the overall sensitivity is 56%. Working on the same line, the specificity of PARZEN for individual disease can be calculated. However, the overall percentage is 70%, 70% and 40% for *UDC*, *LDC*, and *PARZEN* separately.

Similarly, precision of *UDC*, *LDC*, and *PARZEN* is 86.36%, 85.71%, and 85% respectively. *F-measure* is calculated through *sensitivity* and *specificity*, which is 72.72%, 70% and 67% for *UDC*, *LDC*, and *PARZEN* respectively. Furthermore, Table 4-13, also reveals the *false negative* and *false positive* values of all these classifiers.

Most of the literature surveyed only considers skewed datasets, where the number of healthy and diseased persons is not equal. This ultimately generates a biased result due to the dominating effect of the majority class. Even the ROC curves are hard to compare using different classifiers for different misclassification costs and class distribution. We have analyzed an equal number of objects for each class to avoid misclassification.

Secondly, some of the findings in the literature show higher accuracy in the classification of neurodegenerative diseases, particularly Masood *et al.* in [7] have used several alternative techniques and showed 86% accuracy rate. However, the feature set is much smaller as compared to the feature set used in this study. Also, they did not provide any solution of imbalanced datasets and used the skewed datasets for classification. Our research work has focused on eight variables (features) as input for our classifiers unlike previous work where only left and right feet signals are considered. Neurodegenerative diseases are more common in males as compared to females and they are closely linked to the age of the person. Therefore, we have also considered gender and age variables to produce results that are more reliable. Moreover, we have also considered the stage of the disease, patients walking speed and time that are other important input variables.

4.9 Comparison with SMOTE

A new oversampling method has been proposed and implemented in this research work and results have demonstrated that this method works comparatively better as compare to the SMOTE (Synthetic Minority Over-Sampling Technique) that is used in the previous literature. A comparison of the results is shown in Table 4-14.

Table 4-14: Oversampling Comparison with SMOTE Results

Classifiers	**SMOTE Results**	**Average Oversampling**
Linear Discriminant Analysis (ldc)	50%	62.5%
Quadratic Discriminant Analysis (qdc)	35%	57.5%
Quadratic Bayes Normal Classifier (udc)	53%	65%
Logistic Linear Classifier (loglc)	39%	52.5%
Fisher's Discriminant (Minimum Least Square-fisherc)	50%	57.5%
Nearest Mean Classifier (nmc)	32%	42.5%
Polynomial Classifier (polyc)	28%	57.5%
Parzen Classifier (parzenc)	46%	60%
Binary Decision Tree Classifier (treec)	42%	30%
Support Vector Classifier (SVC)	39%	50%
k-nearest Neighbor Classifier (knnc)	39%	32.5%

Due to the complexities of multi-class dataset (4-Dimensional), it is not easy to compare the results with the other oversampling techniques presented in the previous findings of literature survey (where binary class datasets are used in classification). I have computed SMOTE method for the same feature set and compared the results with Average Oversampling technique. Results have shown that in both cases the *UDC* classifier provides us better results but in case of SMOTE the accuracy rate for classification is 53% while in case of proposed oversampling method the accuracy rate is 65%. The same is the case with other classifiers where I get better results with proposed novel oversampling method as compare to SMOTE.

4.10 Summary

This chapter has presented the classification of gait signals of different neurodegenerative diseases using statistical pattern recognition techniques. It has addressed the main issue of *data preprocessing* i.e. skewed datasets, multiclass datasets, and missing entries. It focused on the detection of different neurodegenerative diseases from a database of 4-class datasets. A set of 11 *base-level* classifiers are tested on a feature set of 8 variables. Various evaluation techniques are used to find and compare the results. Results revealed that Bayes Normal classifiers (*UDC and LDC*) and a non-linear classifier *PARZEN* outperform others.

A novel idea of *combining classifiers*, from the *base-level* classifiers is presented in the next chapter. The main focus is to check if *combining classifiers* can increase the percentage of accuracy.

Chapter 5 Classifier Fusion Strategy

5.1 Introduction

In the previous chapter, we have tested various *base-level* classifiers on a feature set of 8 variables. From the pool of the classifiers those which provide complementary information, operating with high accuracy rate and lower error rate, are selected for our further research work. This chapter aims at presenting a novel approach for *combining classifiers* based on an evaluation methodology. This chapter also presents six different combining rules (product, maximum, minimum, mean, median and voting) and their importance in the combining of classifiers. The main purpose is to check if the new approach shows superior performance compared to the *stand-alone* classifiers. Instead of looking for better classifier and more appropriate set of features, this chapter provides an insight of looking at best set of classifiers and the best combination method.

5.2 Classifier Fusion Strategy

In the field of pattern recognition, the main focus is the successful classification of the features with maximum possible accuracy rate. Ahmed and Mohamed in their paper [22] narrates that a classifier with a specific set of features may or may not be an appropriate option for another set of features. Further augmenting their statement, they say that different classification algorithms achieve different degree of success for different kinds of

applications. Giving the reference of [117], they claim that *combining classifiers*, in an efficient way, can offer better complimentary information about the patterns to be classified than any single classifier. Combining classifiers provide more accurate decisions but at the expense of complexity.

Dietterich [118] suggested three reasons in favour of assembling the classifiers than choosing a single classifier.

1. Statistical Reason

Picking a single classifier increases the risk of making a bad choice to solve a problem. For instance, two classifiers at the same time are giving zero error rates. Nonetheless, the generalization performance of these two classifiers can be different even if the error rate is same. Hence, a safer option is, instead of picking one classifier, uses two of them and takes average of the output. This will decrease the risk of picking an inadequate classifier.

2. Computational Reason

An assembled classifier starts a searching process by running the local search from many different starting points. This solves the problem of local search of an individual classifier that may get stuck in local optima.

3. Representational Reason

The classifier space that is considered for pattern recognition does not always contain the optimal classifier. For instance, a set of linear classifiers is chosen for a dataset that can best recognize by nonlinear classifier. This way an optimal classifier can never be obtained. On the contrary, an ensemble of linear classifiers can approximate a decision boundary with any predefined accuracy.

Anil *et al*. [119] narrates some more benefits of combing classifiers for pattern recognition:

- By combining different classifiers together, an opportunity is provided to the designer to have an access of different classifiers that belong to different context and are developed for entirely different representation. For instance, face, voice as well as hand writing recognition.

- An ensemble classifier can also handle the multivariate training sets that are collected at different times and also in different environment. The training set may also have different features.
- Global as well as the local performance of each classifier is different from each other on the same set of training data. Each classifier occupies its own space in the feature space where it performs the best.
- By selecting a single network, as in case of neural networks where results vary by varying the initialization process due to the randomness inherent in the training procedures. Instead of discarding a single network all the networks take advantage to learn from training dataset, effectively.

Fusing classifiers together to seek better accuracy rate is a new approach in classification research that has not been fully explored yet. In many cases, improved classification results can be obtained by combining the output of single classifiers. Estimates of posterior class probabilities are improved when multiple classifiers are considered in parallel [120]. However, combining classifiers in a treelike structure, using weighted averages [121] is useful for analyzing real-time datasets. Fusing classifiers together in this way has already been successfully used within other domains, such as the identification and classification of remotely sensed images [122]. Clearly, these studies show that the accuracy and computational time of individual classifiers can potentially be improved when classifiers are combined [123].

In summary, even in the presence of different feature sets, different training sets, different training sessions or different classification methods, the final result depends on the output of combined classifiers hoping an improvement in the classification accuracy. If a fixed set of classifier (detail in Section 5.3) is selected then the main focus is on the combination function. Fixed combining rules can also be used by optimizing the input classifiers [119].

5.3 Classifiers Combining Rules—Combiners

After the selection of particular *base-level* classifiers, the next step is a search for a module that is needed to assemble the classifiers together, which is called the *combiner*. Combiners can be differentiated on the base of different characteristics— trainability, adaptivity, and

requirement of the output of individual classifier. Some combining techniques are adaptive in their nature. They work by evaluating the decisions of the individual classifiers depending on the input of individual classifier such as adaptive weighting [124], associative switch, mixture of local experts (MLE) [125], and hierarchical MLE.

Depending on the type of output from individual classifier, Xh et al. [126] grouped the expectation level in to three states: 1) *measurement (or confidence)*, 2) *rank*, 3) *abstract*. At the *measurement level* the output of the classifier is a numerical number which indicates the chances of given output belongs to a particular class. At the *rank level*, the choice of the class depends on the highest rank assigns by the classifier. However, it is not necessary that a highest rank is also the highest confidence level. At the *abstract level*, the decision is normally made on the base of unique class label or class labels. Further to his explanation, he added that the *confidence level* provides the highest information about the decision of a class while *abstract level* provides the least information.

A combination process consists of a set of individual classifiers (*base-level* classifiers) and a combining rule which combines the results of individual classifiers for a final decision. When and how the base-level classifiers will work together depends upon the combination scheme. According to Anil *et al.* [119] the combination schemes could be differentiated on the basis of their architecture, the characteristics of the combiner, and the selection of the individual classifies.

On the basis of the architecture, the combining schemes are divided into three categories, that are addressed in [119]; 1) *parallel*, 2) *cascading (or serial combination)*, 3) *hierarchical (tree like)*. In the *parallel* scheme, all the *base-level* classifiers are invoked separately and independently and later the results are combined by a combiner. In the *cascading* style, the individual classifiers are invoked in a linear sequence. For the sake of efficiency the cheap classifiers in term of computational time and measurement demands, are invoked first followed by the most accurate and the expensive one. In the *hierarchical* architecture, the *base-level* classifiers are combines into a decision tree like structure.

In our implementation of *classifier Fusion strategy*, *parallel* architecture is selected due to its simplicity, less computational time and also higher confidence level.

Once the posterior probability of all the classifier is computed, the next step is to combine them into a new set that can be used for maximum selection, for final classification. Robert and David in their paper [127] mention two sets of combining rules; 1) fixed combining rules, 2) trained combining rules.

5.3.1 Fixed Combining Rule

The fixed combing rules make sure that the classifier output is not just a number rather it should have a clear interpretation— class labels, distance and confidence level. The posterior probabilities are also considered the confidence. Following are the main fixed combining rules:

***Maximum*:** the *maximum* rule selects the outcome of the classifier producing the highest estimated confidence, which seems to noise sensitive. This apparently seems quite simple to select a classifier that is more confident on its output. However, this fails if the classifiers are overtrained. In that case the final decision is based on overconfidence, hence dominating the confidence without providing a better performance [128]. In addition, maximum rule fails for simple classifiers that are not sensitive for nuances hence better classifiers are required for detection.

***Median and Mean*:** they both average the posterior probability estimates thereby reducing the estimation error. This works well if all the *base-level* classifiers estimate more or less the same quantity.

***Minimum*:** the *minimum* rule selects the outcome of the classifier that has the least objection against a certain class. Likewise the maximum rule, it is hard to find the adequate situation where this rule performs the best.

***Product*:** it works by taking the product of posterior probabilities of each classifier.

***Majority/Voting*:** it counts the vote for each class over the input classifiers and selects the majority class. It simply coincides with the simple majority, normally (50% of the vote+1) in case of 2-class dataset.

5.3.2 Trained Combining Rule

Trained combining rules, on the other hand, train an arbitrary classifier using all the trained data in the intermediate space. The classifiers are usually trained as an output classifier, using the same training data set. The posterior probabilities are directly used for the building of the intermediate space. If the classes are not normally distributed then it is more advantageous to use the nonlinear rescaling.

5.3.3 Fixed vs. Trained combining Rules

This section provides a brief description about the advantages and disadvantages of fixed as well as trained combining rules:

- Fixed rules are simple to use and can be used without training of the classifier.
- Fixed rules require low memory space and less computational time while trained combining rules require more time as well as more memory space.
- Fixed rules are suitable for independent/ low correlated errors and exhibit similar performance. On the contrary, trained rules are suitable for classifiers that are correlated or exhibiting different performance.
- Flexibility of trained rules is better than fixed rules and also most of the time they perform better than fixed rules.

5.4 Implementation

Building on the previous set of results described previously, this section considers the three best performing classifiers for inclusion in the fusion classifier strategy. From the eleven classifiers tested, the Linear Discriminant Classifier (ldc), Quadratic Bayes Normal Classifier (udc) and the Parzen Classifier (parzenc) provide the best results with their accuracy in percentage being 62.5%, 65% and 60%, respectively. These base classifiers are selected and are included in the fusion strategy.

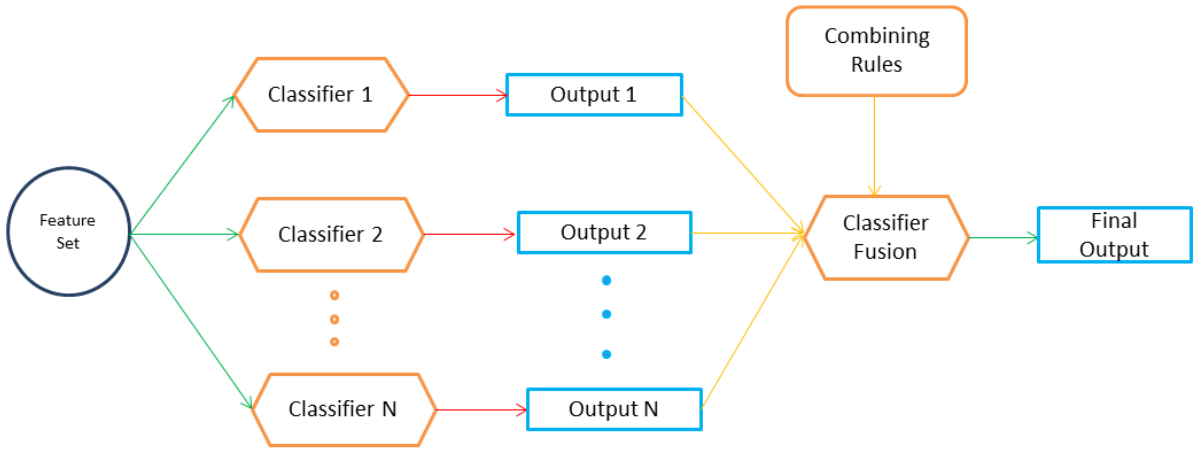


Figure 5-1: A Classifier Fusion Strategy

Figure 5-1, depicts a general framework of combining classifiers using combining rules. A feature set, related to most related attributes of the data, is provided to the classifiers. Resulting output are used to combine the classifiers using some combining rules.

The base classifier evaluation and fusion classifier strategy was implemented using Matlab. The code in Figure 5-2, begins by dividing the dataset into four classes (Huntington's, Parkinson's, Amyotrophic Lateral Sclerosis and healthy persons). Each class is assigned a label (1 for healthy, 2 for Huntington's, 3 for Parkinson's and 4 for Amyotrophic Lateral Sclerosis). Following the division of the dataset, it is randomly split into two equal parts; 50% for training the classifier and the rest are used for testing. Three untrained classifiers *(w1 = ldc, w2 = udc and w3 = parzenc)* are combined without rules into an array *(w)* were *w = [w1,w2,w3]*. The simultaneous training of a set of untrained classifiers is done using the training dataset and this results in a cell array *(v)* containing the trained classifiers.

```
1: a = dataset((data),genlab([20 20 20 20],[1;2;3;4]));
2: [trainset,testset] = gendat(a,0.50); %Divide dataset into two equal halves
3: w1 = ldc; % Untrained classifiers w1, w2 and w3
4: w2 = udc;
5: w3 = parzenc;
6: w = {w1,w2,w3}; % Combining Classifiers
7: v = trainset * w; %  % Training of Classifiers
```

Figure 5-2: Simultaneous Training of a Set of Three Base-Level Classifiers

The results for each of the classifiers are illustrated in Figure 5-3 below.

```
8:  disp([newline 'Errors for individual classifiers'])
9:  testc(testset,v); % Display errors

    Errors for individual classifiers

    Test results result for:

     clsf_1 : Bayes-Normal-1
     clsf_2 : Bayes-Normal-U
     clsf_3 : Parzen Classifier

        clsf_1     clsf_2     clsf_3

        0.425      0.350      0.475
```

Figure 5-3: Test Results Summary

Figure 5-4, shows how the base-level classifiers are combined into a cell array using a set of fixed combining rules. Six different rules were analyzed; *minimum selection (minc), maximum selection (maxc), median selection (medianc), mean selection (meanc), product combiner (prodc)* and *voting selection (votec).* The evaluation of a cell array of trained classifiers *(v)* and the untrained classifiers combined with rules *(vc)* is done by testing the (*testset*) set.

```
10: comb_base = [v{:}]; %Combining Classifiers into Cell array
11: wc = {prodc,meanc,medianc,maxc,minc,votec}; % Combining Rules into Cell array
12: vc = comb_base * wc; % Base Level classifiers combined by rules
13: testc(testset,vc); % Testing on 50% test data
```

Figure 5-4: Combing Base-Level Classifiers

The results for the classifier fusion strategy with combining rules are illustrated in below.

```
Test results result for:

 clsf_1 : Product combiner
 clsf_2 : Mean combiner
 clsf_3 : Median combiner
 clsf_4 : Maximum combiner
 clsf_5 : Minimum combiner
 clsf_6 : Voting combiner

    clsf_1  clsf_2  clsf_3  clsf_4  clsf_5  clsf_6

    0.375   0.375   0.450   0.375   0.475   0.350
```

Figure 5-5: Test Results Summary

Figure 5-6, describes the scenario and illustrates the simulated results obtained during the evaluation of gait signals using the eleven *base-level* classifiers. The results obtained from the three best performing classifiers are stored in a single array, and their error rates are computed. The *mean* error rate for the three classifiers is 0.42. The same three classifiers are

then combined into a cell array using six different rules. The *mean* error rate for the combined classifiers is 0.40, which is slightly less than the mean of the *base-level* classifiers.

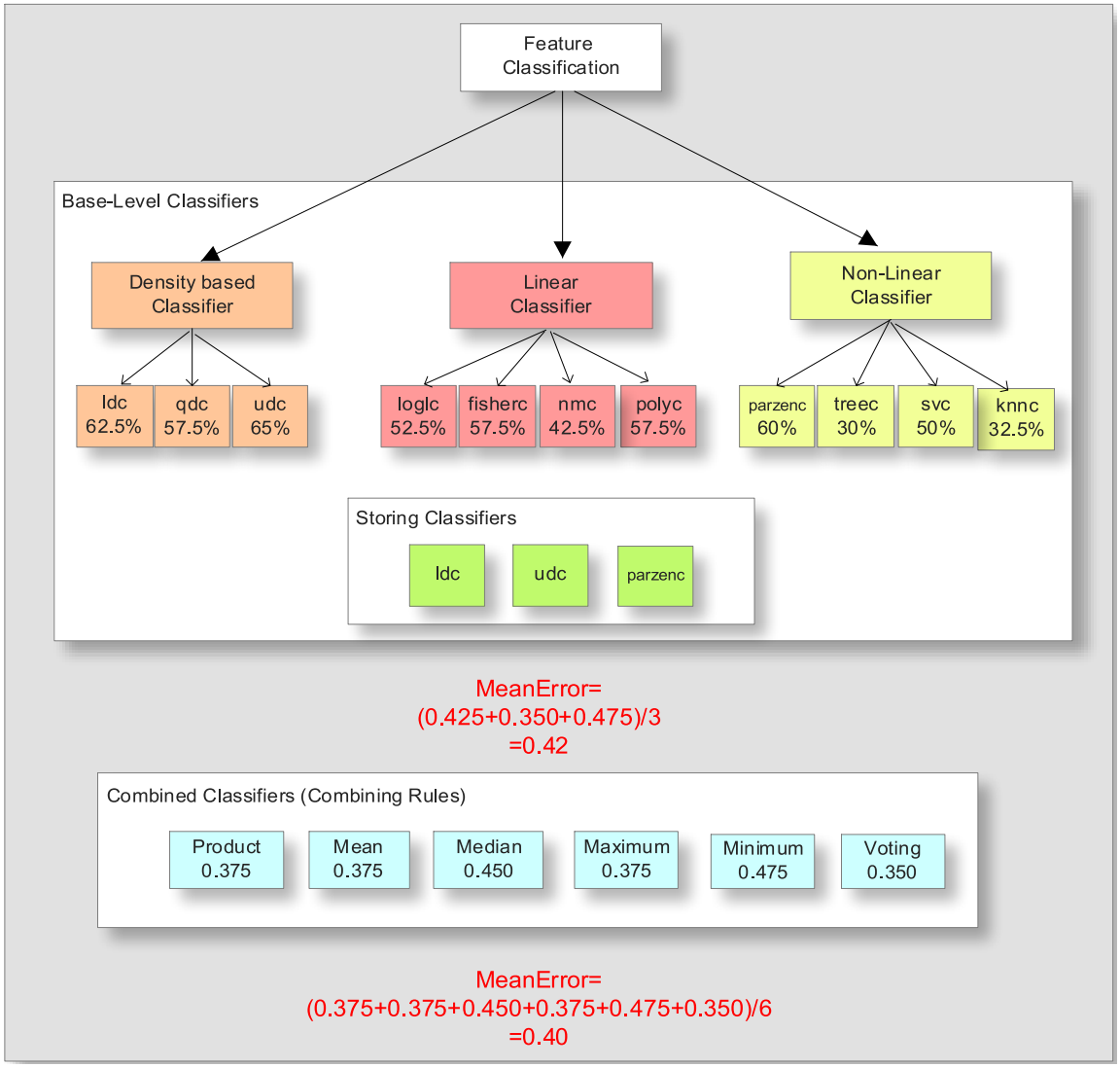
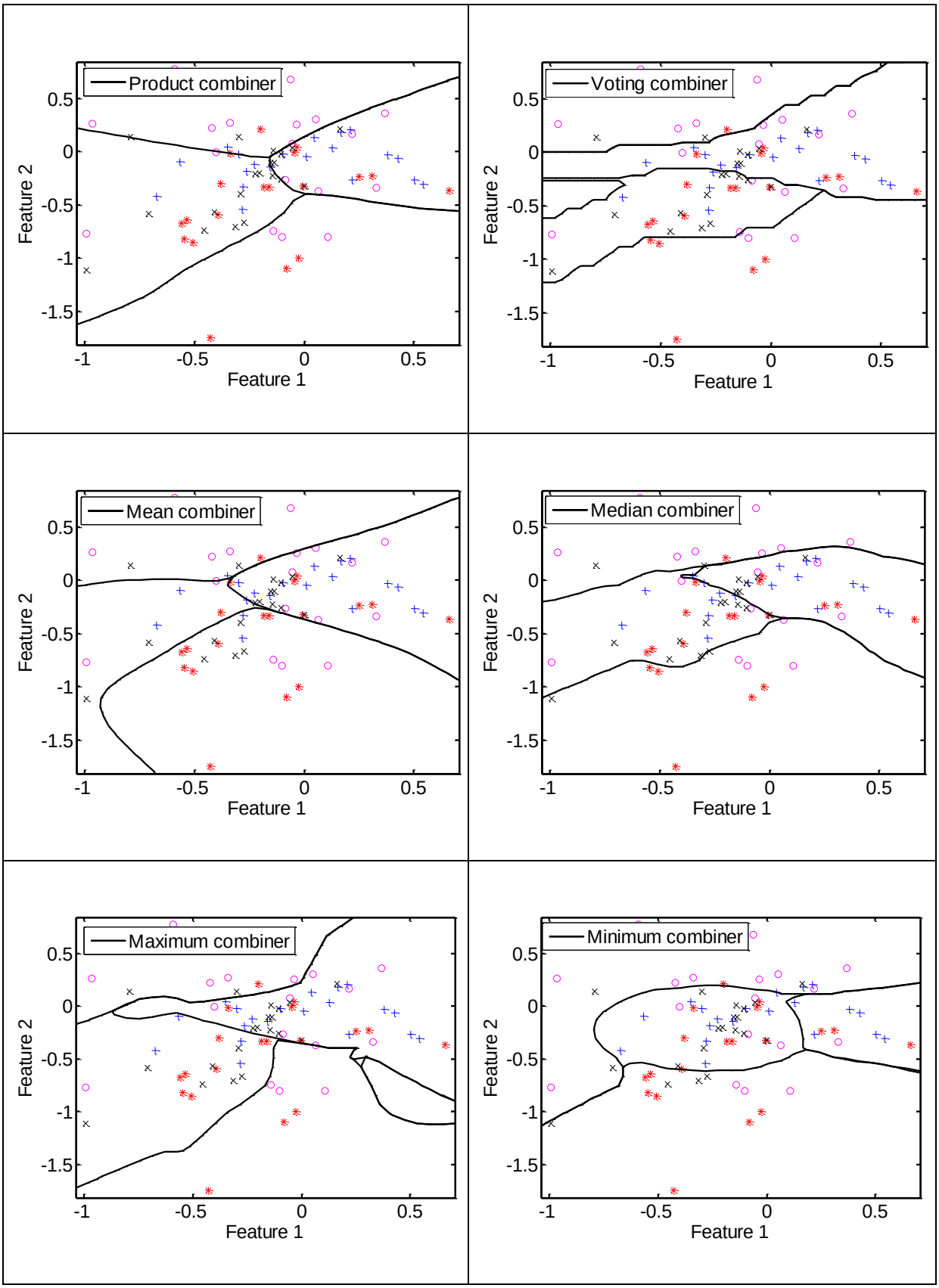


Figure 5-6: Summary of Feature Classification by Base-Level Classifiers and Fusion Strategy

Table 5-1, shows the mapping of each individual *combining rule* (minimum selection (minc), maximum selection (maxc), median selection (medianc), mean selection (meanc), product combiner (prodc) and voting selection (votec)) on a scattered plot of 2-features' dataset of gait signals of neurodegenerative patients and healthy persons.

Table 5-1: Class Mapping of Individual Combining Rule



5.5 The Evaluation and Discussion

This section presents the results of experiments performed on the classifier fusion strategy. For the evaluation of fusion strategy, the multiclass *ROC* (Receiver Operating Characteristic) analysis [115] technique and *Reject Curve* are used. This technique is useful for analyzing several different classes, four different classes here. First, the classifiers are evaluated in Matlab using the *"testc"* routine, which provides several performance estimates for a trained classifier on a test dataset. The *mean* value produced by the test results for individual classifiers is 0.42, which is an error rate. In comparison, the *mean* value for combined classifiers is 0.40, which is obtained by combining different classification rules. This has clearly shown that the combined classification technique (using combining rules) works better than the individual use of classifiers. Moreover, the results depict that the *voting* combination rule works more efficiently than other combining rules used. Using the voting combination rule, the prediction of the base-level classifiers is combined according to a static voting scheme, which does not change when changes to the training set are made [129].

Figure 5-7, presents the *ROC* analysis for fusion strategy. It shows the results when classifiers are combined using various combining rule algorithms that include *product, mean, median, maximum, minimum* and *voting* combining rules. As shown in Figure 5-7, the best result that produces the least error is the "Voting Combiner" with a value of 35.0% error rate. This is closely followed by the "Product Combiner," "Mean Combiner" and "Maximum Combiner." While other rules like "Median Combiner" and "Minimum Combiner" are showing 45.0% and 47.5% error respectively.

Similarly, Figure 5-8, shows the *Reject curve* for *fusion strategy* which indicates that the least error rate is for *voting* combining rule i.e. 0.35. On the other hand *product*, *mean* and *maximum* shows the same error rate which is 0.375. *Median a*nd the *minimum* have shown comparatively bad results i.e. 0.45 and 0.47 respectively.

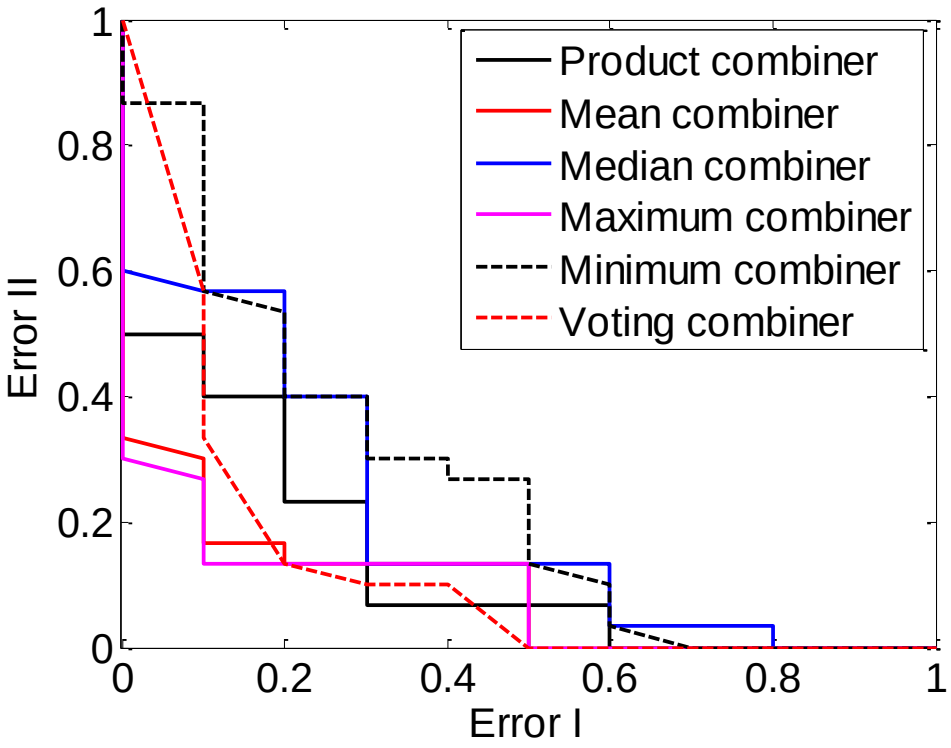


Figure 5-7: ROC Analysis to Compare Combining Techniques

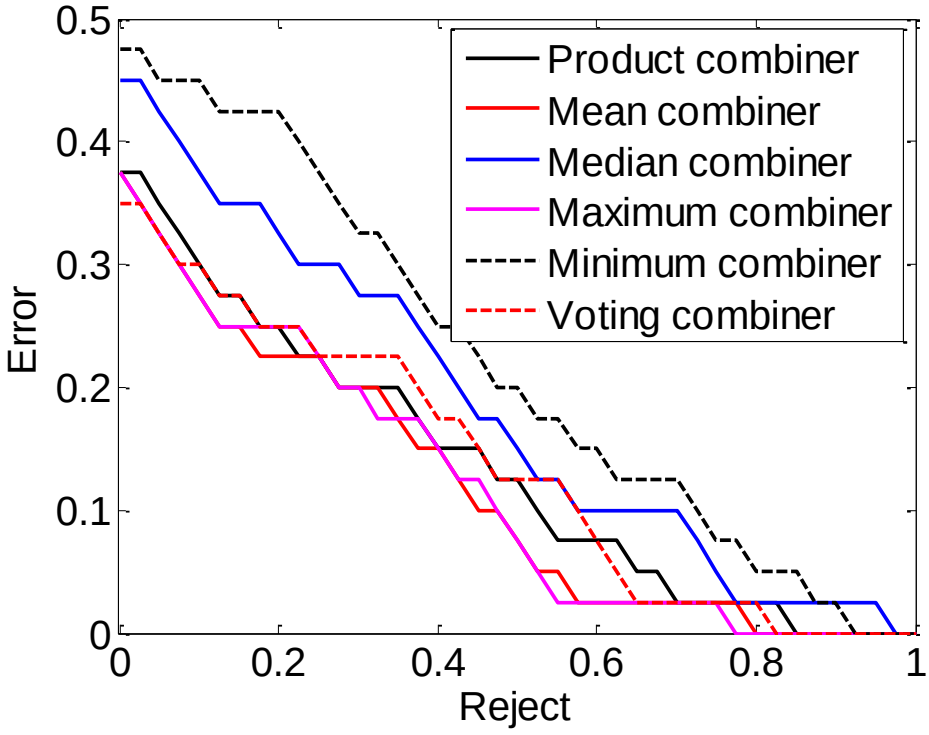


Figure 5-8: Reject Curve for Classifier Evaluation

5.6 Summary

In this chapter a novel idea of *classifiers fusion* is presented, implemented and then evaluated. From the literature, it has been revealed that combining *base-level* classifiers provide a better chance of getting accurate results. It also helps of avoiding those classifiers that are not suitable for a particular type of dataset; linear or nonlinear datasets. Total 6 combining rules are selected to implement the fusion strategy. They are minimum selection *(minc),* maximum selection *(maxc),* median selection *(medianc),* mean selection *(meanc),* product combiner *(prodc)* and voting selection *(votec)*). Total *mean* error rate for combining rule is 0.40 while that of *base-level* classifiers is 0.42. It has also been noted that voting combing rule has provided the highest accuracy rate as compare to other combining rules. Results revealed that the accuracy rate of combining classifiers with combining rule is greater than combining the *base-level* classifiers without any rule. Results are also shown in the form of ROC and *Reject curves*.

Chapter 6 Neural Synchrony Measurement

6.1 Introduction

An emergent research line arises from the integration between clinical neuroscience and computer science is called neuroengineering. It constitutes engineering, computational simulation, mathematical analysis, imaging techniques, and hardware based modelling to solve problems in clinical neuroscience. The goal is to apply engineering approach to further explore the research areas in neural functions and to use that knowledge in the diagnostic and therapeutic systems.

This chapter provides an importance of integrating computer science in clinical neuroscience for the early detection of neurological diseases. It presents the concept of neural synchronization to distinguish the neuronal activities in healthy persons as well as in mild Alzheimer's disease patients (MiAD). Electroencephalogram (EEG) is used as an important tool for the diagnosis of Alzheimer's disease. To detect perturbation in the EEG signals of MiAD patients, three neural synchrony measurement techniques; phase synchrony, magnitude squared coherence and cross correlation are applied on three different databases of MiAD patients and healthy subjects. At the end, all the aforementioned techniques are assessed by a statistical test (Mann-Whitney U test) to compare the results.

6.2 Neural Activities and Cerebral Cortex

Nervous system consists of neurons and non-neural cells. Neurons or nerve cells are used to transfer information to and from the brain and consequently a complex network is formed to perform the functions of nervous system. The classification of the neurons can be done on the basis of their morphology or functionality. On the basis of functionality, there are three types of neurons; sensory neurons, motor neurons, and interneurons. Sensory neurons are used to connect with the sensory receptors; motor neurons are connected to the muscles while interneurons are connected to the other neurons [130]. Neurons do not work in splendid isolation; they are interconnected to each other to form a network which is called neural network. Each part of the circuit is assigned a specific task.

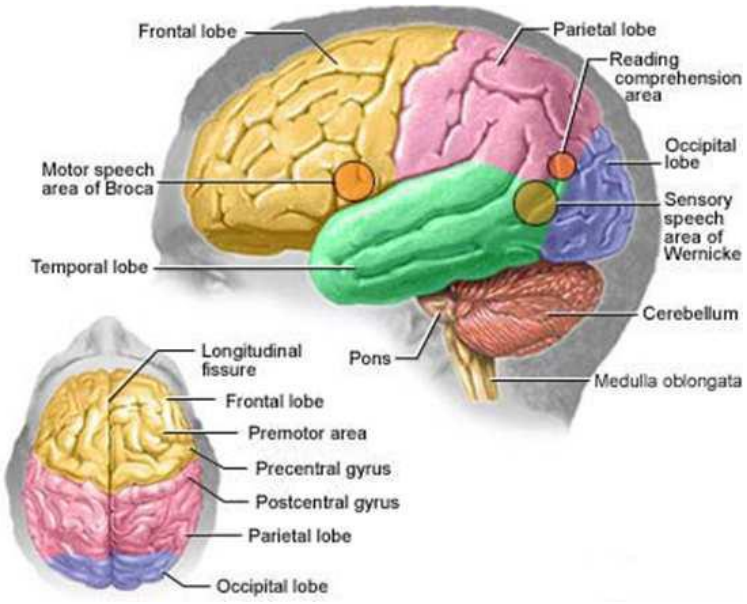


Figure 6-1: Cerebral Cortex and its Four Lobes [130]

The outermost layer of the cerebrum is called cerebral cortex which lies inside the brain. It plays an important role in language, consciousness, thought, attention, awareness, and memory. Cerebral hemisphere is divided into four lobes; frontal lobe, temporal lobe, parietal lobe/central lobe and occipital lobe, as shown in the Figure 6-1. Frontal lobe is associated with decision making, planning, problem solving, and motor speech. Temporal lobe is involved in language, hearing, emotion, sensory speech and memory. The parietal lobe which may also be called central lobe is involved in processing the sensory information of the body, reading comprehension and reception while occipital lobe responsible of vision [130].

6.3 EEG, Electrodes Placement, Data Collection

As we already have explained in Chapter 2 that EEG can be defined as the mean electrical activity in the brain at different sites of the head [51]. More specifically, it can be defined as the extracellular current flows of a large number of neurons. These electrical communications between the neurons are measured as a function of time. The change of potential in the neurons is measured when various neurons synchronously de- or hyperpolarized. The sum of the electrical potential, when cortical neurons simultaneously active is between 10µV to 150µV on the human scalp. The signal measured between two electrodes constitutes the EEG as shown in the Figure 6-2, this is discovered by Hans Berger in Jena in 1924 [8].

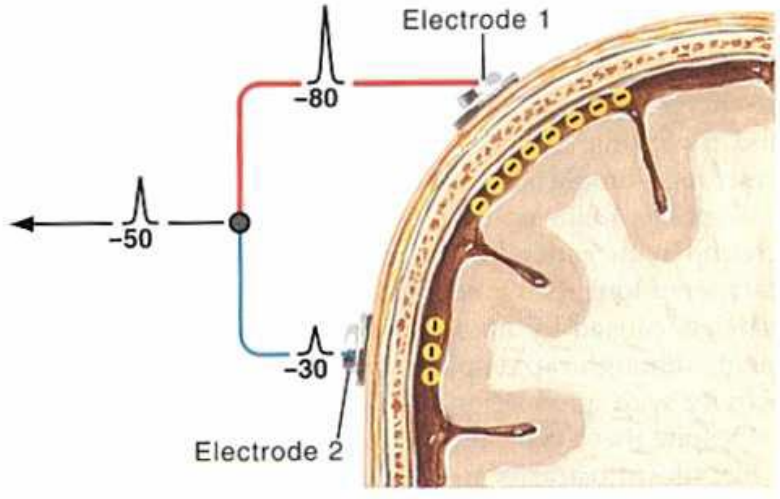


Figure 6-2: Collection of EEG from two electrodes [130]

This bioelectric phenomenon that is stimulated from the cerebral cortex by auditory and somatosensory stimuli is further investigated, explored and interpreted to understand the brain functionality and to identify different pathologies. This process particularly helps to find out those diseases that are impossible to be cured by medication, for instance, the neurodegenerative diseases. This phenomenon also helps for the early detection of life threatening diseases and widely used in the clinical studies.

Figure 6-3, shows the EEG signal in time and frequency domain. Time and frequency domain and the adoptive methods are the popular methods in EEG signal analysis for the early detection of neurodegenerative diseases [130].

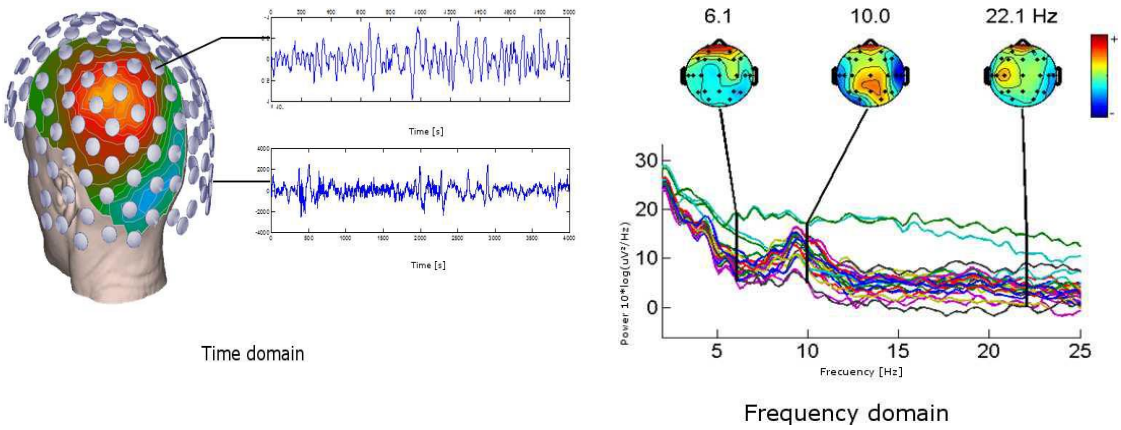
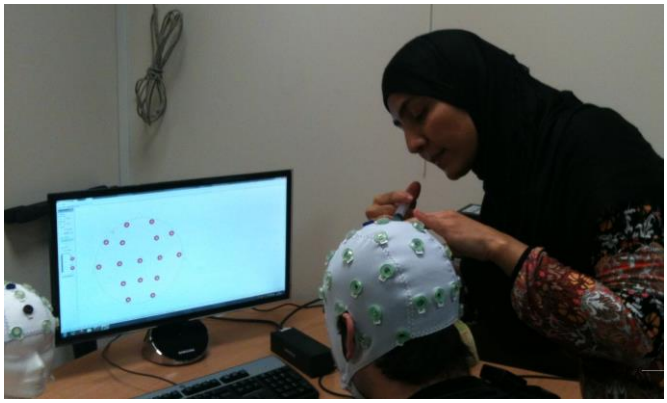
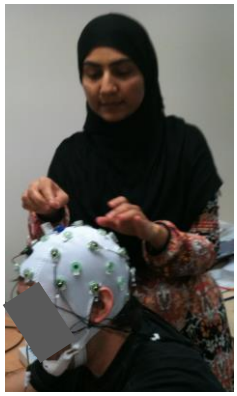


Figure 6-3: EEG signals in time and frequency domain [130]

For our early experiments to understand the structure and hidden patterns of EEG signals, we have collected EEG data from various healthy subjects at ESPCI PaisTech SIGMA laboratory, France. The age of the subjects was between 25 to 40 years.



(a) Gel injection on the scalp to increase conductivity



(b) Inserting the electrodes into the cap to receive EEG

Figure 6-4: Electrodes placement to Collect EEG from Healthy Subjects

Figure 6-4 (a), demonstrates the placement of electrodes cap on the scalp of our subject. The electrodes are placed according to international 10-20 system [131]. A gel is injected within these electrodes holes on the scalp to increase the conductivity between the scalp and the electrodes. After inserting the gel, the electrodes are placed on the cap to collect the EEG data as shown in Figure 6-4 (b). The EEG system, we use in *SIGMA lab* is *actiCap* EEG system with 16 electrodes, amplified by a *V-Amp 16* amplifier, both from *Brain Products*. The

electrodes are active electrodes and the data is filtered using the *Vision Recorder* software from Brain Products. The data is afterwards analysed using Matlab.

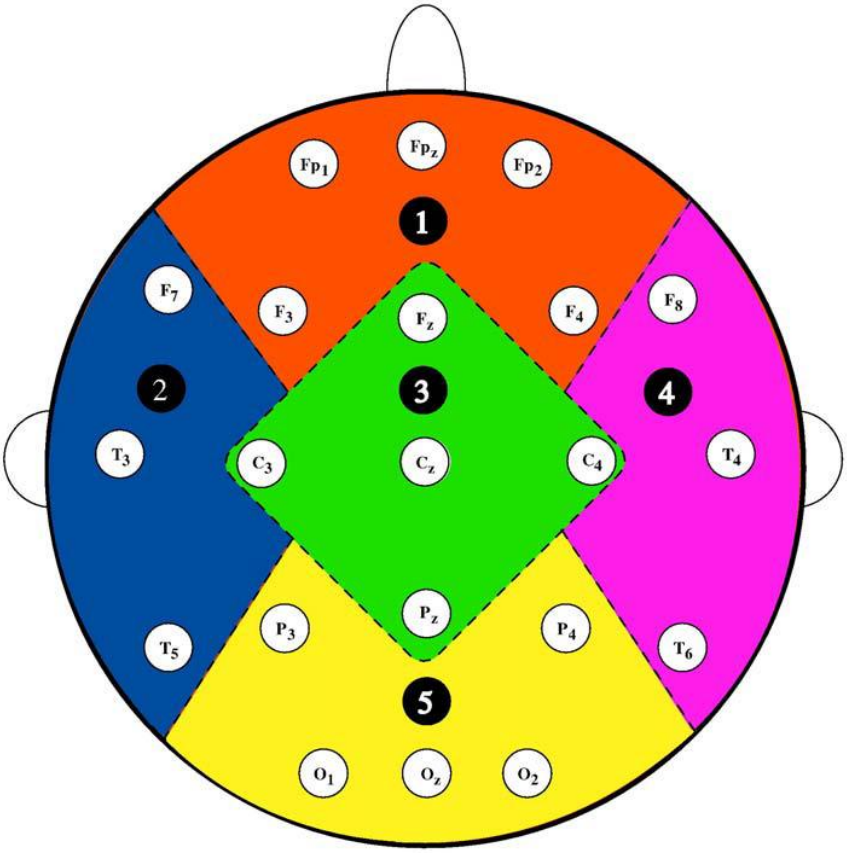


Figure 6-5: The 21 channels used for EEG recordings [132]

Figure 6-5, shows the position of the electrodes on the cap to receive EEG signals from the brain of the subjects. Different parts of the brain are differentiated with different colours, dotted lines and also with integers. For instance, integer '1' which is written on the orange part denotes to frontal part of the brain. This part includes five (5) channels—FP_1, FP_2, FP_z, F_3, and F_4. Left temporal is denoted by integer '2' and it is highlighted with blue colour. This part includes three (3) channels— F_7, T_3 and T_5. Similarly, central part of the brain is denoted by integer '3' and it is highlighted with green colour. This part has 5 channels—F_z, C_3, C_z, C_4, P_z. The 4^{th} part is right temporal which is denoted by integer '4' and is highlighted by pink colour. This part constitutes of three (3) channels—F_8, T_4, T_6. Similarly, the last part which is denoted by integer '5' is called occipital. This part is highlighted with yellow colour and it consists of five (5) channels—P_3, P_4, O_1, O_2, O_z.

6.4 Neural Synchronization

Synchronization, precisely speaking, is a coordination of "rhythmic oscillators" for a repetitive functional activity [133]. Whereas, neural synchronization is putatively considered a mechanism where brain regions simultaneously communicate each other to complete a specific task such as perception, cognition, and action [134, 135]. Any disturbance in the brain, caused by a disease or any other infection, can highly effect the synchronization of brain. Quantitative analysis of EEG signals provides a better insight of synchronization between different parts of brain. For instance, a less synchrony has been detected in the EEG signals of AD patients as compare to healthy persons [132].

Mild Cognitive Impairment (MCI) is characterized by impaired memory state of brain probably leading towards mild Alzheimer's disease (MiAD) or Alzheimer's disease (AD). This prodromal stage of AD is under a great influence of research since long time [2, 136, 137]. Statistics report that 6-25% of MCI are transformed to AD annually and 0.2-4% from healthy person to AD, [136, 138] revealing the fact that MCI is a transition state of MiAD and AD.

Loss of functional connectivity between cortical and hippocampus has long been an important focus of research to examine the cause of cognitive dysfunction in AD [139, 140]. Statistical analysis of interdependence among times series recorded from different brain areas, to study the functional interaction, is called "functional connectivity" [141]. Due to destructive characteristics of AD, it has also been characterized as a neocortical "disconnection syndrome" [142]. Brain's visualization, as a complex network of subsystems, has led us to find out the factors that can best identify functional disorders in brain [143]. There is now ample evidence that formation of dynamic links in term of synchronization constitutes the functional integration of brain [144-146].

Various synchrony measurement techniques have already been discussed to detect any perturbation in the EEG signals of AD patients [147]. For instance, both linear such as coherence and nonlinear such as phase synchronization methods are widely used to quantify synchronization in electroencephalographic signals [140, 148, 149]. A comparison of occipital inter-hemispheric coherence (IHCoh) for normal older adults and AD patients reveals a reduced occipital IHCoh both for lower and higher bands of alpha [150]. Almost

similar findings reported by Locatelli *et al.* [151] where a significant increase in delta coherence is noticed between frontal and posterior regions in AD patients while a decrease in alpha coherence is shown in temporo-parieto-occipital areas. Spontaneous phase synchronization of different brain regions is calculated by Kuramoto's parameter (ρ), which is particularly useful to measure multi-channel dataset [140].

Despite the considerable success of above mentioned techniques to analyze disruption in the EEG signals of Alzheimer's patients, further investigations are still required to fulfil the clinical requirements. For instance, in order to detect Alzheimer's at its earlier stage we need to focus on those areas where Alzheimer attacks at first and then we need to check its synchronization with rest of the brain regions. Furthermore, additional novel and comprehensive methods are still required to check the validity of aforementioned techniques on EEG signals to detect any perturbation in the brain signals of Alzheimer's patients.

6.5 Spatial-Spectral Analysis of EEG

Electroencephalogram (EEG) signals are considered a functional exam to evaluate cognitive disturbances and used as a diagnostic tool, especially when a diagnostic doubt exists even after the initial clinical procedures [147, 152]. A great deal of research has already been conducted to detect the fluctuations in (EEG) signals [132, 136, 139]. Alteration in the regional cerebral blood flow (rCBF) has been considered one of the causes of abnormality in EEG signals of AD [153, 154]. Studies on MCI have shown a decrease of alpha power [155, 156] and an increase of theta (4-8 Hz) power [157, 158] in cortio-cortical and subcortical parts of brain. Similarly, Babiloni e*t al.* in [136] have claimed the reduction of the synchronization likelihood both at inter-hemispherical (delta-beta2) and fronto-parietal (delta-gamma) electrodes.

Topographically analyzing the EEG signals, Micheal *et al.* [159] reported a less synchronization of upper alpha band between central and temporal cortex. In line, a correlation between higher low-frequency amplitude and alpha-beta activity at frontal region may reflect an early sign of cortical atrophy during the course of AD [160]. Similarly, perturbation in cholinergic inputs from basal forebrain to cortex and hippocampus indicates a decrease in cortical EEG coherence [161] that can be considered a biomarker for the early

detection of AD [136]. Moreover, a combination of multi-linear interaction within the tensor formed by (*subject x frequency x regions)* also provides a simple set of features for the interpretation and classification of AD at its early stage [97]. The concept of local and global methods is used to analyze synchronization between pairs of signals and entire EEG channels at the same time, respectively [132].

However, the studies so far, have provided a very limited regional comparison of brain. For instance, less synchronization has been reported between temporal and central regions [159] and also in fronto-parietal region [136]. Similarly, functional coupling of EEG rhythms by sensorimotor events is presented only in centro-parietal regions of brain [162]. A wider range of study is still required to analyze synchronization likelihood in all parts of brain (right temporal, left temporal, frontal, central and occipital) at the same time, on different sets of data to fully explore the progression of Alzheimer's disease in a patient.

6.6 Research Challenges

The above review suggests, firstly, that Spatial-Spectral Analysis of EEG signals can provide a measure of memory visualization. Secondly, neural synchrony measurement techniques have a potential to discriminate between AD patients and healthy subjects. What is still missing or ambiguous in the literature survey is the simultaneous comparison of all parts of brain with the right and left temporal (the most affected parts of brain) to analyze synchronization and also the implementation of new comprehensive methods to apply synchrony measurement techniques. Following paragraph highlights the importance and some novel contributions of our research work to analyse EEG signals:

- The amplitude range for the EEG signals vary from 29μV to 100μV while the frequency of EEG signals usually ranges from less than 1 Hz to 60 Hz [140]. Previous studies focus on the analysis of compact EEG signals without filtering them into narrow frequency bands. This lacks providing optimal information about the frequency band which is more important in detecting the Alzheimer's at its earlier stage. To detect the synchrony loss of EEG signals in Alzheimer's patients, we have filtered dataset of MiAD and control subjects into five narrow frequency bands (delta (δ: $1 \leq f \leq 4$ Hz), theta (θ: $4 \leq f \leq 8$ Hz), alpha (α: $8 \leq f \leq 12$ Hz), beta (β: $12 \leq f \leq 25$

Hz) and gamma (γ: $25 \leq f \leq 30$ Hz)). For each frequency band we have computed neural synchronization to compare all parts of brain (frontal, occipital and central) with left and right temporal.

- A high inter-subject variability has been seen in the EEG signals of AD patients, especially with different levels of severity and comorbidities [163-165]. In this situation the findings are not more reliable on a single set of data. Most of the existing studies focus on a single synchrony measure with a single set of data [166]. In this case it is hard to compare the results to conclude a single hypothesis. To extract a general set of feature we have analysed three different databases, each from one hospital at a time (detail in section 6.8). This will not only increase the validity of our research but will also provide more reliable findings that can later be used in clinical applications.
- Reducing features vector dimension, commonly known as feature reduction, will not only help us to get better results accuracy and a better speed of signal processing but will also avoid the classifiers to be over-fitted [167]. Analysing results without removing the redundant information or without eliminating the noise leads to non-reliable results. To remove the ambiguity of biased results due to "features redundancy" we have applied PCA (Principal Component Analysis) technique before applying synchrony measurement techniques.
- In this research work, we have proposed two novel methodologies to compute neural synchronization. One is named as *Average* synchrony measure. We have compared this technique to another proposed methodology named *PCA* based synchrony measure. Later, the results are compared using a statistical method called Wilcoxon–Mann–Whitney test.

Before this, Besthorn *et al.* [168] applied PCA technique in the quantitative analysis of EEG signals to compress a group of predictor variables to a small set of factors or principle components. Later they applied linear discriminant classifier on these variables to discriminate AD patients from healthy subjects. Similarly, Peter *et al.* [169] applied PCA to remove the artifacts from EEG signals that were generated by eye-blink. To the best of our knowledge and the literature we have surveyed so far, we could not find the application of PCA to remove the redundant features from the data that can generate a biased result to compute the synchronization of brain areas.

6.7 Neural Synchrony Measurement Technique

In this section, we briefly review the synchrony measurement techniques that we have implemented on our datasets which include phase synchrony, cross correlation, and coherence.

6.7.1 Phase Synchrony (Hilbert Transform)

Synchronization of the two periodic non-identical oscillators refers to the adjustment of their rhythmicity i.e. the phase locking between the two signals [170, 171]. It refers to the interdependence between the instantaneous phases $\varphi_1(t)$ and $\varphi_2(t)$ of the two signals $x_1(t)$ and $x_2(t)$ respectively. It is usually written as:

$$\varphi_{n,m} = n\varphi_1(t) - n\varphi_2(t) = constant \tag{6.1}$$

Where n and m are integers indicating the ratio of possible frequency locking, and $\varphi_{n,m}$ is their relative phase or phase difference. To compute the phase synchronization, the instantaneous phase of the two signals should be known. This can be detected using analytical signals based on Hilbert Transform [143].

$$z(t) = x(t) + i\tilde{x}(t) \tag{6.2}$$

Here *z(t)* is a complex value with *x(t)* is a real time series and $\tilde{x}$*(t)* is its Hilbert transform.

6.7.2 Cross Correlation

Cross correlation is a mathematical operation used to measure the extent of similarity between two signals. If a signal is correlated to itself it is called auto-correlated. If we suppose that *x(n)* and *y(n)* are two time series then the correlation between them can be calculated as [172]:

$$\hat{R}_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_{n+m} y_n & m \geq 0 \\ \hat{R}_{yx}(-m) & m < 0 \end{cases} \tag{6.3}$$

Cross correlation returns a sequence of length 2*M−1 vector, where *x* and *y* are of length N vectors (N>1). If *x* and *y* are not of the same length then the shorter vector is zero-padded. Cross correlation returns value between −1 and +1. If both signals are identical to each other the value will be 1, otherwise it would be zero [132]. A high negative correlation indicates a high correlation but of the inverse of one of the series.

6.7.3 Magnitude Squared Coherence

The coherence functions estimates the linear correlation of signals in frequency domain [132]. The magnitude squared coherence is defined as the square of the modulus of the mean cross power spectral density (PSD) normalized to the product of the mean auto PSDs [173]. The coherence $C_{xy}(f)$ between two channel time series is computed as:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)} \tag{6.4}$$

$P_{xy}(f)$ is the cross PSD estimate of *x* and *y*. $P_{xx}(f)$ and $P_{yy}(f)$ are the PSD estimates of *x* and *y* respectively. For computation, each signal is divided into a section of 650ms and default value of 50% is used. Coherence returns the values between 0 and 1, showing how well the input *x* corresponds to the output *y* at each frequency.

6.8 Data Description

The datasets, we are analysing, have been recorded from three different countries of European Union. Specialist at memory clinic referred all patients to the EEG department of the hospital. All patients passed through a number of recommended tests; Mini Mental State Examination (MMSE) [174], The Rey Auditory Verbal Learning Test [175], Benton Visual Retention test [176] and memory recall tests [177]. The results are scored and interpreted by psychologists and a multidisciplinary team in the clinic. After that, each patient is referred to hospital for EEG assessment to diagnose the symptoms of AD. Patients were advised to be in a resting state with their eyes closed. The sampling frequency and number of electrodes for three datasets are all different. Detailed information is as follows:

6.8.1 Database A

The EEG dataset A contains 17 MiAD patients (10 males; aged 69.4 ± 11.5 years) while 24 healthy subjects (9 males; aged 77.6 ± 10 years). They all are of British nationality. This data is obtained using a strict protocol from Derriford Hospital, Plymouth, U.K. and had been collected using normal hospital practices. EEG signals were obtained using the modified Maudsley system which is similar to the traditional 10-20 international system [131]. EEGs were recorded for 20 sec at a sampling frequency of 256 Hz (later on sampled down to 128 Hz) using 21 electrodes.

6.8.2 Database B

This EEG dataset composed of 5 MiAD patients (2 males; aged 78.8 ± 5.6 years) as well as 5 healthy subjects (3 males; aged 76.6 ± 10.0 years). They all are of Italian nationality. Several tests, for instance; MMSE, the clinical dementia rating scale (CDRS) and the geriatric depression scale (GDS) were conducted to evaluate the cognitive state of the patients. The MMSE result for healthy subjects is (29.3 ± 0.7) while for MiAD patients is (22.3 ± 3.1). EEGs were recorded for 20 sec at a sampling frequency of 128 Hz using 19 electrodes at the University of Malta, Msida MSD06, Malta.

6.8.3 Database C

This dataset consists of 8 MiAD patients (6 males; aged 75 ± 3.4 years) and 3 healthy subjects (3 males; aged 73.5 ± 2.2 years). They all are of Romanian Nationality. The AD patients have been referred by a neurologist for EEG recordings. All subjects are diagnosed with AD by means of psychometric tests (MMSE, CDR, OTS), neuroimaging (CT) and clinical examination (gender, age, disease, duration, education and medication). The MMSE result for healthy subjects is (28-30) while for MiAD patients is (20-25). EEG data is recorded using a large equidistant 22-channel arrangement conforming to the international federation of clinical neurophysiology (IFCN) standards [178] for digital recording of clinical EEG from the Ecological University of Bucharest. The time series are recorded for 10 to 20 min at a sampling frequency of 512 Hz using 22 electrodes. The signals are notch filtered at 50 Hz.

6.9 Data filtering

The oscillatory, repetitive behaviour of the recorded EEG signals which represents the electrical activity of the cerebral cortex is called rhythm. This enormous diversity in the EEG signals usually depends on the mental state of the subject e.g. degree of attentiveness, walking and sleeping modes. Theses rhythms of the signals are attributed by relative amplitude and the frequency ranges.

Depending on the frequency ranges, the EEG signals are divided into five (5) narrow frequency bands. These frequency ranges are delta (δ), theta (θ), alpha (α), beta (β) and gamma (γ). Alpha and beta waves were introduced by Berger in 1929. Later in 1936 Walter introduced the delta waves and also informed about the theta waves that range between 4-8 HZ. Two years later, in 1938 Jasper and Andrew came to know about the waves above 25 HZ and called them gamma waves [179].

Each frequency band has its own physiological significance [140] [51]:

- Delta (δ: $1 \leq f \leq 4$ Hz): these are characterized for deep sleep and are correlated with different pathologies. They usually have high amplitude. They do not encounter in the awake and normal adults but are considered as an indication of cerebral damage or brain damage.
- Theta (θ: $4 \leq f \leq 8$ Hz): they play important role during childhood. They might have got their name because of their origin from thalamic region. High theta activities in adults are considered abnormal and associated with brain disorders. The theta range appears in drowsiness and in certain stages of sleep. They are somewhat related in unconsciousness, inspirations and deep mediations.
- Alpha (α: $8 \leq f \leq 12$ Hz): they usually appear during mental inactive conditions and under relaxation. They are best seen during eye closed and mostly pronounced in occipital location. They are over the occipital region of the brain and appear in the posterior half of the head. They are of sinusoidal shaped signal but could be found in a sharp shape.
- Beta (β: $12 \leq f \leq 25$ Hz): they are visible in central and frontal locations. Their amplitude is less than alpha waves and they mostly enhance during tension. They are associated with an activated cortex and are observed during certain sleep stages.

- Gamma (γ: $25 \leq f \leq 30$ Hz): sometimes they are also called fast beta waves. They are related to a state of active information processing of the cortex. They are best characterized for cognitive and motor functions. Figure Figure 6-6, represents the EEG signals in narrow frequency bands.

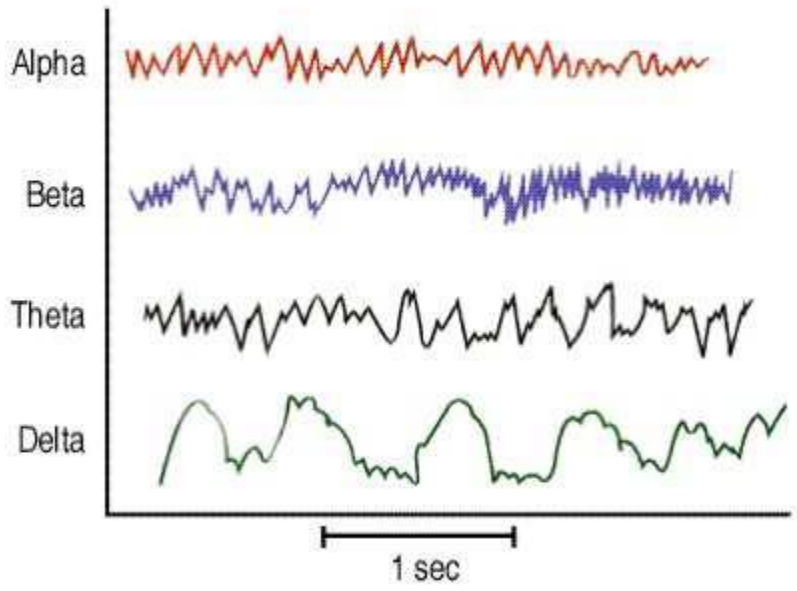


Figure 6-6: Typical normal brain waves in the EEG

A bandpass filter is applied to each EEG channel to extract the EEG data in specific frequency band [F:(F+W)] Hz. Butterworth filters of 2^{nd} order were used as they offer good transition band characteristics at low coefficient orders; thus, they can be implemented efficiently [180].

6.10 Different Approaches to Compute EEG Synchrony

Different approaches have already been implemented to measure the synchrony between different parts of the brain for Alzheimer's patients, MCI patients and healthy subjects. Dauwels *et al.* in their paper [132] have proposed two unique methods to compute synchrony which they named *Local* and *Global* synchrony measures. In *Local* synchrony they have computed the synchrony of different regions (left and right temporal, frontal, central, and occipital) separately and then compare the results of one region with the other. While in *Global* approach, synchrony measures are applied to all 21 channels simultaneously. They

named this computation ‘large-scale synchrony measure’ since each region spans several tens of millimetre.

Taking inspirations from these concepts, we have presented advance and novel approaches to compute EEG synchrony for Alzheimer’s patients, for all parts of the brain in optimized and narrow frequency bands. In this research work we are presenting Average and PCA based EEG synchrony measure for Alzheimer’s and healthy subjects. A detail description of these methods is provided in the next sections.

6.10.1 Average Synchrony Measure

Average EEG synchrony takes its name because the likelihood of synchronization between two parts of the brain is calculated by computing average of synchrony measures for all channel pairs between two respective parts. This means that, first we apply neural synchrony measurement technique on each channel pair (time series of two channels) of two different regions for all frequency bands and then we take the average of those results.

For instance, we apply phase synchrony measure on each channel pair of left and right temporal ((F_7-F_8), (F_7-T_4), (F_7-T_6), (T_3-F_8), (T_3-T_4), (T_3-T_6), (T_5-F_8), (T_5-T_4), (T_5-T_6)) and then we take the average result of right temporal-left temporal. Similarly, we compare the left temporal with frontal ((F_7- FP_1), (F_7- FP_2), (F_7- FP_z), (F_7- F_3), (F_7- F_4),(T_3- FP_1), (T_3- FP_2), (T_3- FP_z), (T_3- F_3), (T_3- F_4), (T_5- FP_1), (T_5- FP_2), (T_5- FP_z), (T_5- F_3), (T_5- F_4)), left temporal-central ((F_7- F_z), (F_7- C_3), (F_7- C_z), (F_7- P_z), (T_3- F_z), (T_3- C_3), (T_3- C_z), (T_3- C_4), (T_3- P_z), (T_5- F_z), (T_5- C_3), (T_5- C_z), (T_5- C_4), (T_5- P_z)), and left temporal-occipital ((F_7- P_3), (F_7- P_4), (F_7- O_1), (F_7- O_2), (F_7- O_z),(T_3- P_3), (T_3- P_4), (T_3- O_1), (T_3- O_2), (T_3- O_z), (T_5- P_3), (T_5- P_4), (T_5- O_1), (T_5- O_2), (T_5- O_z)).

Working on the same line, we compare the right temporal (F_8, T_4, T_6) to rest of the brain area. For instance, we apply phase synchrony measure on each channel pair of right and left temporal ((F_8-F_7), (T_4-F_7), (T_6-F_7), (F_8-T_3), (T_4-T_3), (T_6-T_3), (F_8-T_5), (T_4-T_5), (T_6-T_5)) and then we take the average result of right temporal-left temporal. Similarly, we compare the right temporal with frontal ((F_8- FP_1), (F_8- FP_2), (F_8- FP_z), (F_8- F_3), (F_8- F_4),(T_4- FP_1), (T_4- FP_2), (T_4- FP_z), (T_4- F_3), (T_4- F_4), (T_6- FP_1), (T_6- FP_2), (T_6- FP_z), (T_6- F_3), (T_6- F_4)), right

temporal-central ((F_8- F_z), (F_8- C_3), (F_8- C_z), (F_8- P_z), (T_3- F_z), (T_4- C_3), (T_4- C_z), (T_4- C_4), (T_4- P_z), (T_6- F_z), (T_6- C_3), (T_6- C_z), (T_6- C_4), (T_6- P_z)), and right temporal-occipital ((F_8- P_3), (F_8- P_4), (F_8- O_1), (F_8- O_2), (F_8- O_z),(T_4- P_3), (T_4- P_4), (T_4- O_1), (T_4- O_2), (T_4- O_z), (T_6- P_3), (T_6- P_4), (T_6- O_1), (T_6- O_2), (T_6- O_z)).

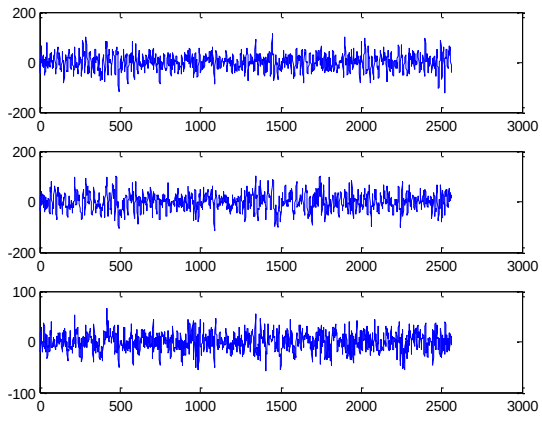
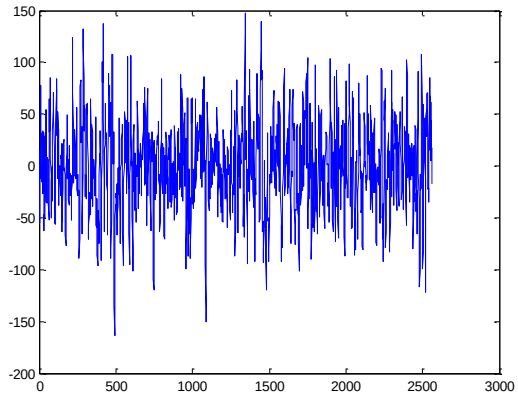
The same procedure has been repeated for rest of the synchrony measures i.e. cross correlation and coherence. After getting the results, we compare the neural synchronization of AD patients and healthy subjects, for all three measurement techniques (phase synchronization, cross correlation and coherence), by Mann-Whitney U test.

6.10.2 PCA Based Synchrony Measure

The basic purpose of Principal Component Analysis (PCA) is to reduce the dimensionality of a dataset to convert it to uncorrelated variables providing maximum information about a data eliminating interrelated variables. In other words it transforms highly dimensional dataset (of m dimensions) into low dimensional orthogonal features (of n dimension) where n<m [181].

In our case we apply PCA on all channels in one particular region, for instance, application of PCA for the left temporal is shown in the Table 6-1, where the signals from channel (F_7, T_3, T_5) are converted into a single variable. It contains almost all information from the left temporal, eliminating any redundant information.

Table 6-1: Application of PCA on left temporal channels

Left Temporal Signals (Channel (F_7, T_3, T_5))	Left Temporal Signal (After Applying PCA)
	

In this method, instead of applying synchrony measurement techniques directly on the filtered data, first we apply Principal Component Analysis (PCA) technique on all channels of one region. This eliminates any redundant information that a region could provide. For instance, we apply PCA on all three channels of left temporal (F_7, T_3, T_5) and consequently it provides a single signal without any redundant information. It is noteworthy here that still we have a signal into five narrow frequency bands (δ, θ, α, β, γ) for each part. This means that for left temporal after the application of PCA on three channels (F_7, T_3, T_5), we have a single signal (say LT) for all these frequency bands (LT_δ, LT_θ, LT_α, LT_β, LT_γ). Similarly, after applying PCA to right temporal (RT) we have the following signals (RT_δ, RT_θ, RT_α, RT_β, RT_γ). For frontal, central and occipital the signals are as (F_δ, F_θ, F_α, F_β, F_γ), (C_δ, C_θ, C_α, C_β, C_γ), (O_δ, O_θ, O_α, O_β, O_γ) respectively.

After the application of PCA, now we have a single comprehensive signal in five frequency bands in each part. Proceeding towards the findings of neural synchronization, we apply neural synchrony measure, say phase synchrony, on EEG time series of two regions. We calculated phase synchrony between left and right temporal ((LT_δ-RT_δ), (LT_θ-RT_θ)., (LT_α-RT_α), (LT_β-RT_β), (LT_γ-RT_γ)), left temporal-frontal ((LT_δ-F_δ), (LT_θ-F_θ)., (LT_α-F_α), (LT_β-F_β), (LT_γ-F_γ)), left temporal-central ((LT_δ-C_δ), (LT_θ-C_θ)., (LT_α-C_α), (LT_β-C_β), (LT_γ-C_γ)), and left temporal-occipital ((LT_δ-O_δ), (LT_θ-O_θ)., (LT_α-O_α), (LT_β-O_β), (LT_γ-O_γ)).

Similarly, we have compared right temporal with the rest of the brain areas; right temporal-left temporal ((RT_δ- LT_δ), (RT_θ-LT_θ)., (RT_α-LT_α), (RT_β-LT_β), (RT_γ-LT_γ)), right temporal-frontal ((RT_δ- F_δ), (RT_θ-F_θ)., (RT_α-F_α), (RT_β-F_β), (RT_γ-F_γ)), Right temporal-central ((RT_δ-C_δ), (RT_θ-C_θ)., (RT_α-C_α), (RT_β-C_β), (RT_γ-C_γ)), and right temporal-occipital ((RT_δ- O_δ), (RT_θ-O_θ)., (RT_α-O_α), (RT_β-O_β), (RT_γ-O_γ)).

The same procedure has been repeated for rest of the synchrony measures i.e. cross correlation and coherence. After getting the results, we compare the neural synchronization of MiAD patients and healthy subjects, for all three measurement techniques (phase synchronization, cross correlation and coherence) by Mann-Whitney U test.

Figure 6-7, shows the entire, step wise, procedure of *Average* as well as *PCA* based synchrony measurement to calculated neural synchronization for MiAD patients and its comparison with healthy subjects.

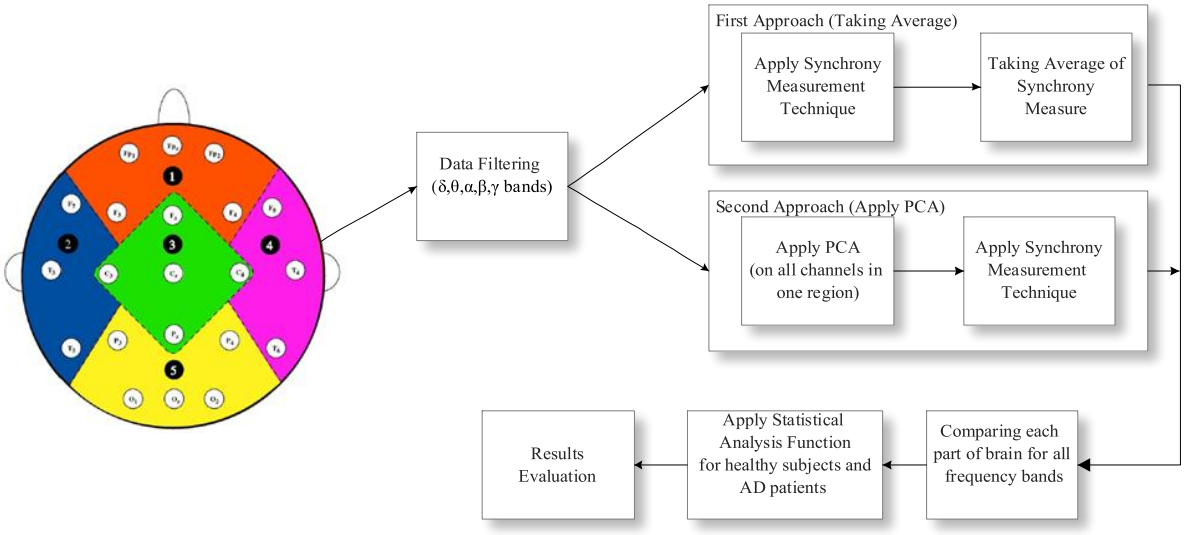


Figure 6-7: Average vs. PCA based Synchrony Measure

6.11 Statistical Analysis

To investigate whether there is a significant difference between the EEG signals of MiAD patients and control subject and also to prove the probable significance of our proposed methodology, we apply Wilcoxon ranksum (Mann-Whitney) test on our datasets. Ranksum function is a non-parametric test which allows us to check whether the statistics at hand, in our case synchrony results, take different values from two different populations. Lower p-values indicate higher significance in term of large difference in medians of two populations [132].

Since we are applying three different synchrony measures on three different sets of data, first we consider our one method (Taking average of synchrony values) to compute the synchrony measure. We apply all three measures for all 7 different comparisons of brain for all frequency bands and compute the results by Mann-Whitney test. Then we apply the same techniques on all, above mentioned, three datasets for 2nd method (PCA based synchrony measures). Now we are able to compare our results in two different perspectives:

i. Investigating three different synchrony measures at a time will help us to compare which measure works better for EEG signals.
ii. Secondly, we are able to compare two different methods for three synchrony measures and for three different datasets.

In addition to evaluate the statistical significance of our proposed methods, this will also help us to differentiate the MiAD patients from healthy subjects.

6.12 Results and Discussions

The aim of the present study is to find the relationship of EEG synchronization with AD, thus to explore further dimensions in disconnection theorem of cognitive dysfunction in AD. And also, to propose a better method to detect any change in EEG synchrony that can be considered as a biomarker for the early detection of AD. Here we investigate and discuss results in two different angles. First, we are discussing the role of synchrony measures to examine a change in EEG synchrony in MiAD patients and later we are conferring the significance of applying PCA before synchrony measures.

6.12.1 Functional disconnection of brain regions due to lower synchronization

We have observed that all of the synchrony measures, tested in this paper, show a decrease in EEG synchrony for MiAD patients as compare to healthy subjects. However, cross correlation shows a higher number of significant results at p=0.01 level as compare to phase synchrony and coherence. We have examined that mostly the areas that have shown less functional connectivity for all three synchrony measures are right temporal- central (RT-C) for delta, theta and alpha bands and also left temporal-occipital (LT-O) for delta and alpha bands. Rest of the paper discusses these two regions where we find highly significant results as compare to the rest of the regions.

First we discuss *Dataset A* for all three synchrony measures with PCA based method. The p-values for cross correlation in RT-C region are $2.47x10^{-4}$, $1.46x10^{-4}$, 0.009 for delta, theta and alpha bands respectively. In LT-O region the smallest p-vales for delta and theta bands are $8.50x10^{-5}$ and $6.8x10^{-5}$ respectively. The 2nd best measure which has given us remarkable

results is phase synchrony, where we get 0.0067, 0.0403, and 0.0585 p-values for delta, theta and alpha bands respectively in RT-C region. We get 0.0041 and 0.0271p-values for delta and alpha bands in LT-O region. Lastly, the coherence function shows significant results in RT-C region for delta band, p-vale=0.0378 and in LT-O region with p-values $9.8x10^{-4}$ and 0.05 for delta and alpha bands respectively. Coherence function does not provide significant results and hence contradicts *Bahar* theory [182] where control group showed higher values of evoked coherence in delta, theta and alpha bands in the left fronto-parietal electrode pairs as compare to AD patients.

Lower p values at delta and alpha bands are shown by *Babiloni et al.* [136] at fronto-parietal coupling of electrodes which indicates a lower synchronization in MCI and AD subjects. Further to the previous findings, our results show higher difference of synchronization for temporal, occipital and central areas in MiAD patients at delta, theta and alpha level. They show the lower magnitude of delta, theta and alpha bands in temporal, central and occipital areas in MiAD patients as compare to healthy subjects. Temporal regions are characterized for short term and long term memory and any neuronal change in these regions is a clear indication of progression of AD.

Interestingly, we find a decrease in alpha band synchronization for all three synchrony measures in almost all regions. For instance, for cross correlation p-value<0.01 in almost all parts of the brain, for phase synchrony the p-values are 0.058, 0.0038, 0.011, and 0.027 in RT-C, RT-O, RT-F and LT-O respectively. This shows the importance of alpha rhythm for the early detection of AD which is in accordance with the phenomena that alpha rhythms are mainly modulated by thalamo-cortical and cortio-cortical systems [183, 184]. Alpha band is mainly related to a subject's global attentional readiness and engagement of specific neural channels for the elaboration of sensorimotor or semantic information [136].

As aforementioned, mostly the areas that show lower dysfunctional connectivity are right temporal-central and left temporal-occipital. A lower synchronization in these connections, especially in RT-C region, for alpha band indicates a disturbance in the perception and integration of somatosensory information, visuospatial processing, and cognitive disorder. This information is in line with clinical findings presented in [185] for increasing visual and

spatial deficits in MCI and MiAD patients. Table 6-2, shows the significant p-values in different parts of the brain in different frequency bands for *Dataset A.*

Table 6-2: P-values for dataset A, different frequency bands in different brain connections

Synchrony Measure	Brain-Connections	Frequency regions	P-values
Cross Correlation	RT-C	Delta (δ)	$2.47x10^{-4}$
		Theta(θ)	$1.46x10^{-4}$
		Alpha(α)	0.009
	RT-O	Delta (δ)	$6.9\ x10^{-5}$
		Theta(θ)	$2.7\ x10^{-5}$
		Alpha(α)	0.0029
	RT-F	Delta (δ)	$5.01x10^{-4}$
		Theta(θ)	$6.8\ x10^{-5}$
		Alpha(α)	0.0062
	LT-C	Delta (δ)	$4.3\ x10^{-5}$
		Theta(θ)	$3.8\ x10^{-5}$
		Alpha(α)	0.0192
	LT-O	Delta (δ)	$8.5\ x10^{-5}$
		Theta(θ)	$6.8\ x10^{-5}$
		Alpha(α)	0.0052
	LT-F	Delta (δ)	$2.2\ x10^{-4}$
		Theta(θ)	$5.4\ x10^{-5}$
		Alpha(α)	0.0091
	LT-RT	Delta (δ)	$3.3\ x10^{-4}$
		Theta(θ)	$6\ x10^{-5}$
		Alpha(α)	0.0253
Phase Synchrony	RT-C	Delta (δ)	0.0067
		Theta(θ)	0.0403
		Alpha(α)	0.05
	RT-O	Delta (δ)	0.0041

		Alpha(α)	0.0271
Coherence	RT-C	Delta (δ)	0.0378
	RT-O	Delta (δ)	0.0378
		Alpha(α)	0.0192

Similarly, for *Dataset B* and *Dataset C* we found low p-values in the same regions for same frequency bands but not as much significant as for the *Dataset A*. Nonetheless, one thing is common in all three datasets that they show lower p-values in alpha frequency bands in RT-C region.

6.12.2 Significance of PCA approach over Average approach

Our second hypothesis was to show the significance of using PCA method where we apply PCA algorithm on EEG signals before applying synchrony measures to eliminate the redundant information from the data to avoid biased results. As expected, we found a big difference in results with and without PCA method. We have found that more than 90% of the values are better in case of *PCA* method as compare to *Average* method for all of three datasets.

For instance, for *Dataset A*, in case of PCA method, we have found 8 significant values below 0.01 ($p<0.01$) and 11 significant values below 0.05 ($p<0.05$) while only 2 values below 0.01 ($p<0.01$) and 8 values below 0.05 ($p<0.05$) in case of Average method for phase synchrony measure. Similarly, for cross correlation measure, although the difference is not very high yet PCA method has shown more significant values. For example, the number of p-values below 0.01($P<0.01$) are 26 while almost all 35 values below 0.05 ($p<0.05$) while for Average method 22 values are below 0.01 while 30 values below 0.05 ($p<0.05$). As aforementioned, coherence function doesn't perform better as compare to other two synchrony measures but again we found more significant results in case of *PCA* method (2 values below 0.01 and 7 values below 0.05) as compare to *Average* method where we found only one significant value below 0.01 and 7 significant values below 0.05.

Table 6-3: Total number of Significant Values in case of PCA and Average method

Synchrony Measure	Method	P<0.01 (Total Values)	P<0.05 (Total Values)
Cross correlation	PCA	26	35
	Average	22	30
Phase Synchrony	PCA	8	11
	Average	2	8
MS Coherence	PCA	2	7
	Average	1	4

Figure 6-8 and Figure 6-9, shows the comparison of results for phase synchrony measure for *Dataset A.*

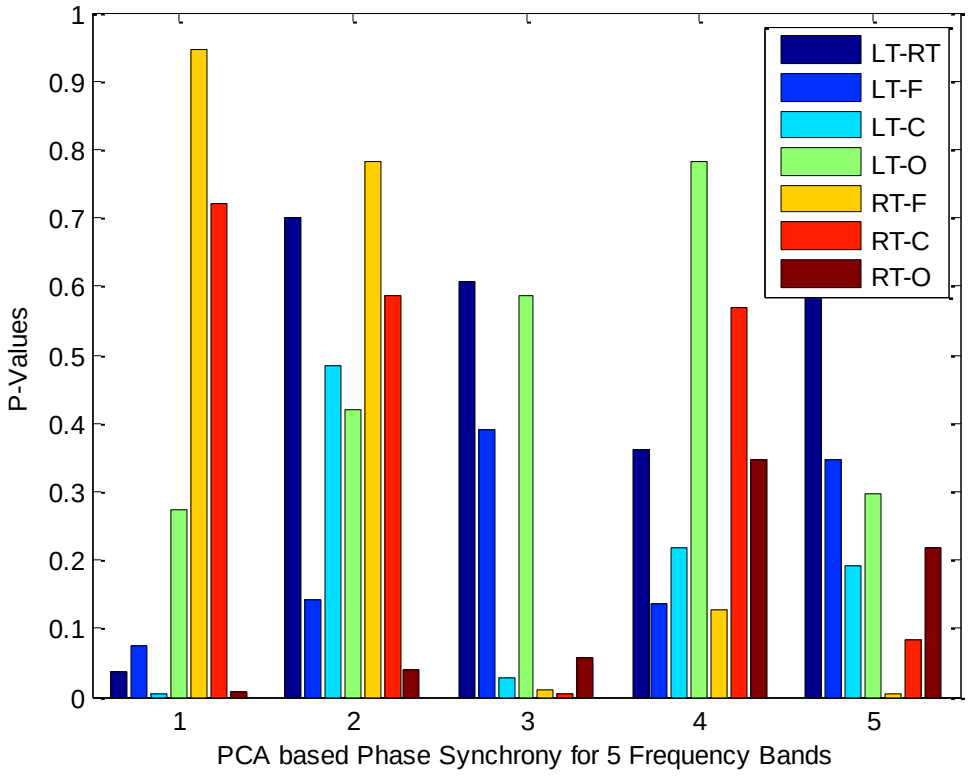


Figure 6-8: PCA based Phase Synchrony for Dataset A

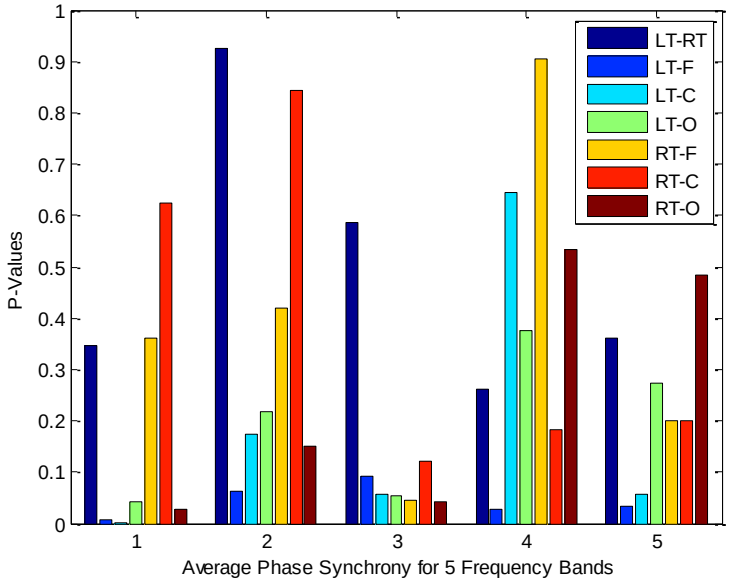


Figure 6-9: Average Phase Synchrony for Dataset A

The x-axis represents the seven (7) comparisons of brain region (LT-RT, LT-F, LT-C, LT-O, RT-F, RT-C, and RT-O) into five (5) different frequency bands (δ, θ, α, β, γ) while y-axis represents the p-values. Similarly, Figure 6-10 and Figure 6-11, represents the results of MS coherence for *Dataset A.*

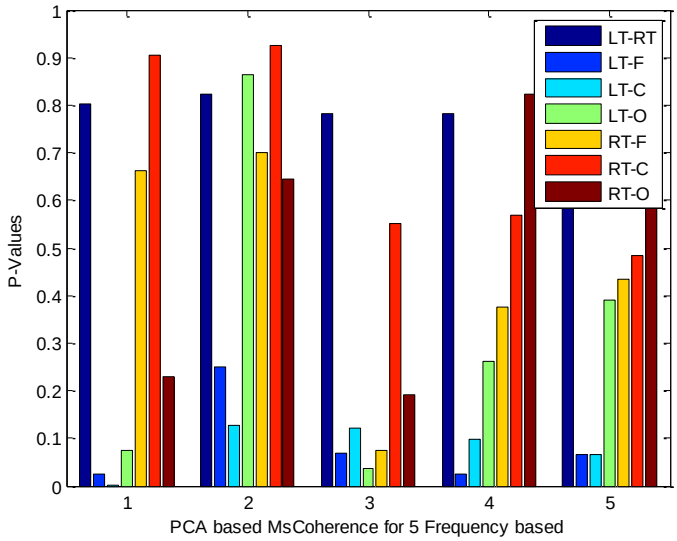


Figure 6-10: PCA based Ms Coherence for Dataset A

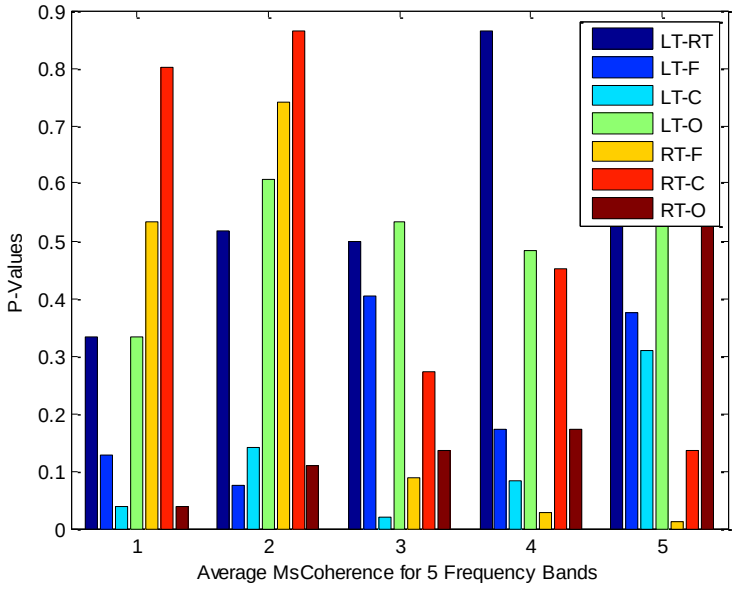


Figure 6-11: Average Ms Coherence for Dataset A

The bar chart results for cross correlation are demonstrated in Figure 6-12 and Figure 6-13.

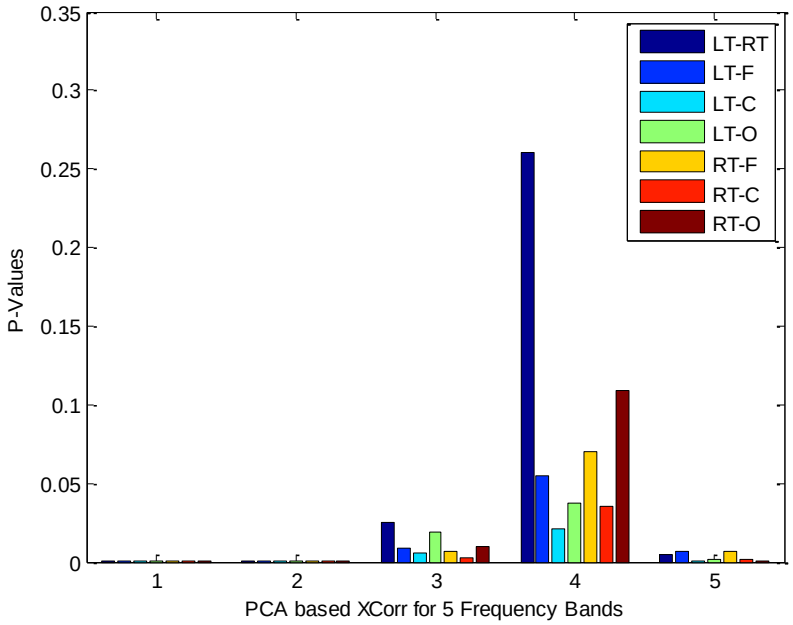


Figure 6-12: PCA based Cross Correlation for Dataset A

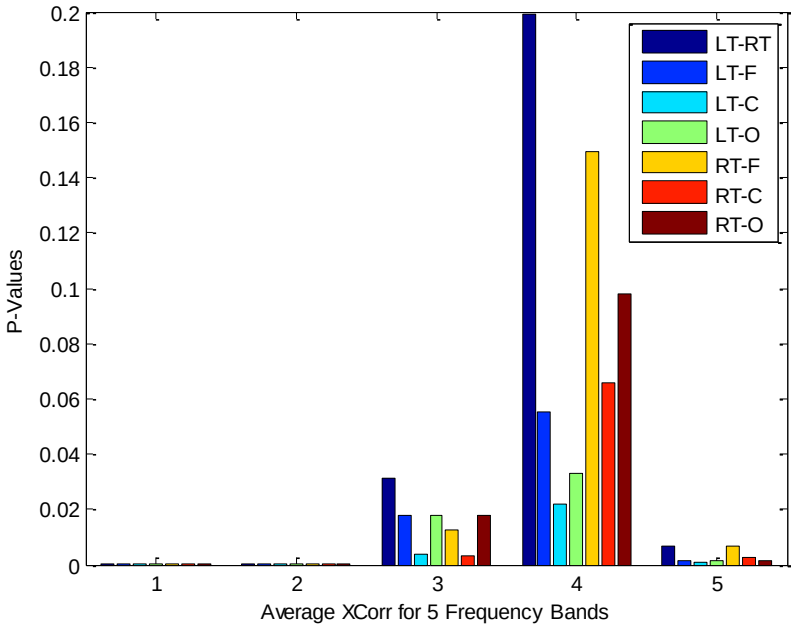


Figure 6-13: Average Cross Correlation for Dataset A

The results are also shown by boxplot in Figure 6-14 that demonstrate the difference of p-values for all three synchrony measures in all seven (7) brain comparison for dataset A. They compare the results of synchrony measures for PCA and Average methods.

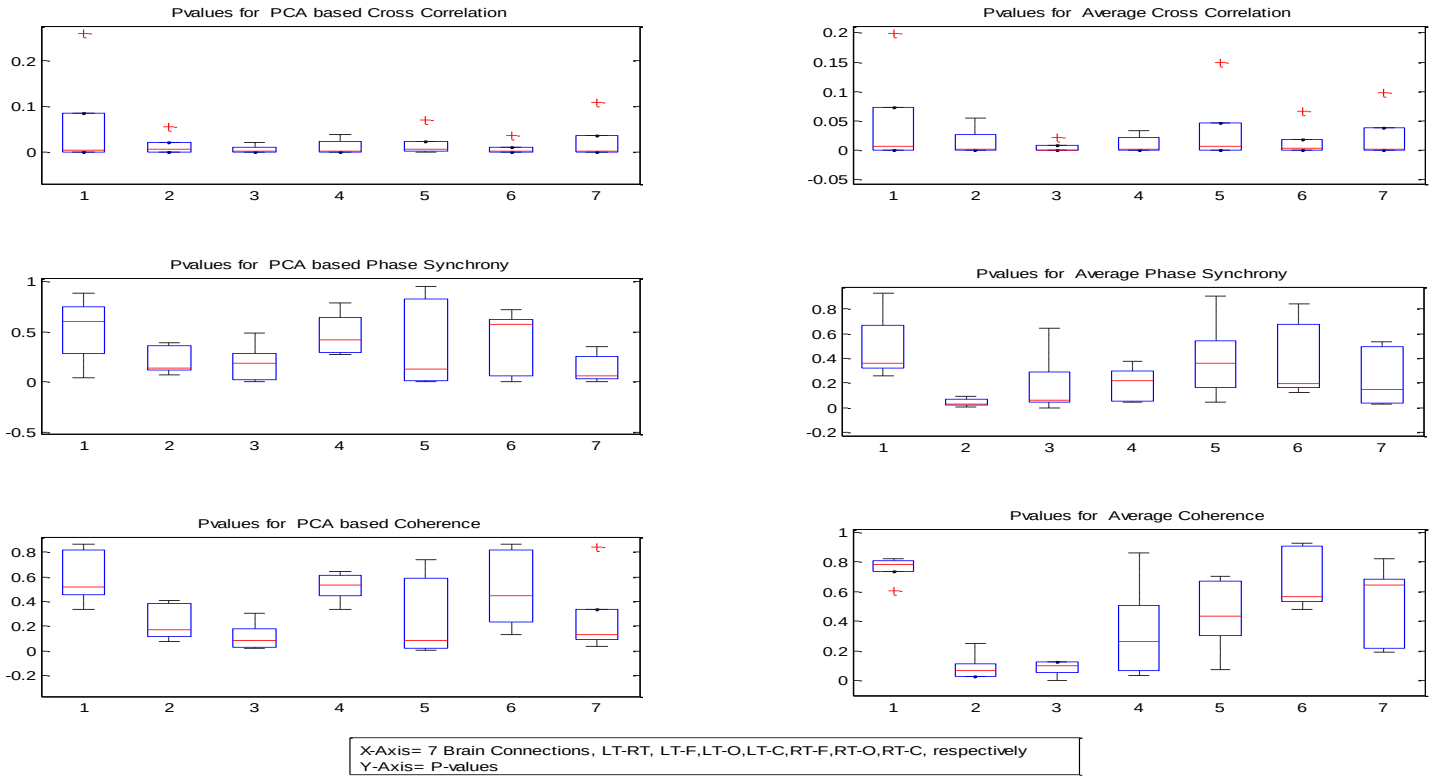


Figure 6-14: Boxplots show the results of three synchrony measures for PCA and Average methods

6.12.3 Challenges with Large Number of Features

Further to the previous section, after the completion of signal processing part, the next challenging step is the classification of the features. As we have seen that, after the implementation of a single synchrony measure, we receive 35 features. For instance, as a result of successful implementation of phase synchrony measure we obtain 35 significant features that can help us to distinguish between MiAD and healthy subjects. Going into more detail, there were total seven (7) comparisons of brain regions (LT-RT, LT-F, LT-C, LT-O, RT-F, RT-C, and RT-O) into five (5) different frequency bands (δ, θ, α, β, γ) which give us 7x5=35 number of features for a single measurement. Similarly, as a result of MS coherence and cross correlation, we receive 35 features each. Hence, total number of features we receive at the end are (35+35+35) =105.

The problem here is if we classify our data with the same set of features (105 features) then the possibility of getting biased results is very high. This would ultimately raise the question about the reliability of the results. The ambiguity of the results might be due to these reasons; 1) some of the features might be providing the same information as others 2) some of the feature included might not have any relation with the early diagnosis of Alzheimer's. 3) Obviously, the computation time of the classifiers will be very high if we process all 105 features at once and also it will occupy more memory space.

To incorporate these issues, we consider applying Gram Schmidt Orthogonalization algorithm. This algorithm works by ranking the most relevant features to the top from a huge set of features that best describe the output— MiAD and healthy subjects. Later, we apply "n-probe" function to select the most relevant features. A detail description is provided in the next section.

6.13 Gram-Schmidt Orthogonalization

Gram Schmidt orthogonalization is first implemented to rank the features according to their significance in term of the provided output—in our case output is Alzheimer's patients and healthy subjects. This means that all the 105 features are ranked according to the most relevant features at the top of the list. After that, the next step is setting the boundary between

the “top” and the “bottom” features—those which should be selected and those which should be discarded. Ranking of the features is demonstrated in a screen short, Figure 6-15.

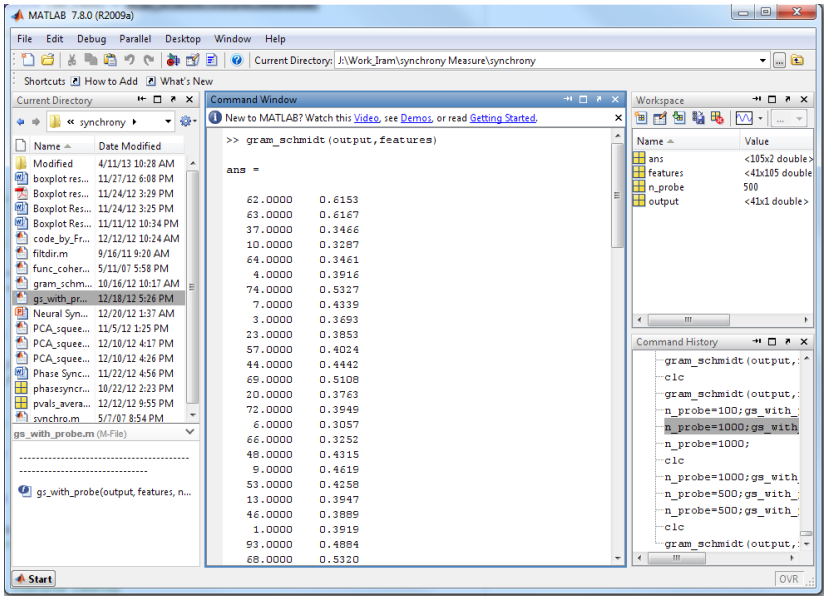


Figure 6-15 : Feature Ranking by Gram-Schmidt Orthogonalization

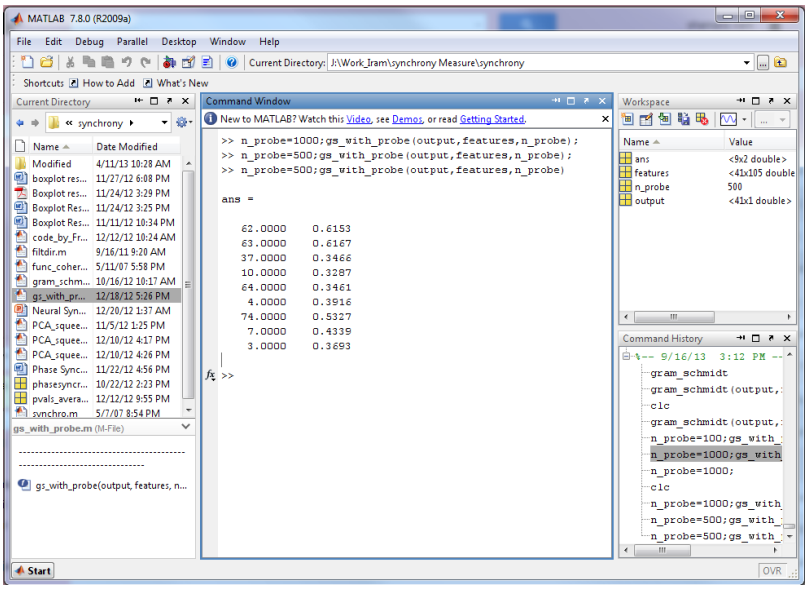


Figure 6-16: Feature Selection by n-Probe

Once all the features are ranked in order of decreasing relevance of the output, an "n-Probe" function is used to select the most relevant features, as shown in Figure 6-16. Since the amount of available data is finite, the probe feature will appear somewhere in the ranked feature list; all features that are ranked below the probe should be discarded. Since the probe is a random variable, its rank in the list is a random variable too. Therefore decision of keeping or discarding a given feature is based on the probability that this feature be ranked higher or lower than the probe. In our case the value of the probe is varied from 500 to 1000 to check if there is any variation in the results.

Table 6-4: Set of Resulting Features

Synchrony Measure	Regions	Frequency Band	**P-Values**
XCorr	Right Temporal-Occipital	Theta Range	2.7×10^{-5}
XCorr	Right Temporal-Occipital	Alpha Range	0.0029
XCorr	Left Temporal-Right Temporal	Theta Range	6×10^{-5}
Phase Synchrony	Left Temporal- Frontal	Gamma Range	0.00741
XCorr	Right Temporal-Occipital	Beta Range	6.9×10^{-5}
Phase Synchrony	Left Temporal-Right Temporal	Beta Range	0.0115

Table 6-4, provides a set of resulting features after the application of Gram-Schmidt Orthogonalization and "n-Probe" function. According to these results, the best synchrony

measure that can differentiate between healthy and MiAD subjects is cross correlation and the 2nd best is phase synchrony. Moreover, these results reveal that cross correlation (xcorr) synchrony in Right temporal-Occipital (RT-O) for theta (θ), alpha (α) and beta (β) ranges provides an optimal information for the early diagnosis of Alzheimer's patients. Also, phase synchrony measure in Left Temporal-Frontal (LT-F) and Left Temporal-Right Temporal (LT-RT) regions provides significant information for theta (θ) and beta (β) ranges which can help the clinicians to early diagnose the Alzheimer's patients. As we can see from the results that mostly the temporal zones are involved to provide us significant information about neural synchronization in MiAD. This is exactly in accordance with the clinical findings that the temporal zones are responsible for the visual processing of objects and pattern recognition. Also, the medial and anterior parts of the temporal zones are involved in high level memory. So, any disturbance in these regions could have a direct link with AD [185].

6.14 Summary

In this chapter we have shown the importance of neural synchronization for the early detection of Alzheimer's disease. It has been shown from previous findings that neural synchronization is considered one of the important biomarkers for the early detection of neurological diseases. EEG has been used to detect the neural synchrony in different parts of MiAD and healthy subjects. Three different neural synchrony measurement techniques named phase synchrony, cross correlation and magnitude squared coherence has been implemented on three different EEG datasets of real MiAD and healthy subjects.

Two novel methods are proposed to calculate the neural synchronization of MiAD and healthy subjects with the above mentioned synchrony techniques—Average Synchrony Measure and PCA based Synchrony Measure. Results are compared using Wilcoxon ranksum (Mann-Whitney) test which reveals that PCA based synchrony measure outperforms the Average synchrony measure. Which means that we have obtained more significant p-values with PCA based method. Moreover, results are more reliable because PCA based synchrony methods works by eliminating any redundant information from the datasets.

Finally, we have implemented Gram Schmidt orthogonalization algorithm with "n-probe" function to get the most important features that can help for the classification and early

diagnosis of AD. Results revealed that cross correlation (xcorr) synchrony in Right temporal-Occipital (RT-O) for theta (θ), alpha (α) and beta (β) ranges provides an optimal information for the early diagnosis of Alzheimer's patients. Also, phase synchrony measure in Left Temporal-Frontal (LT-F) and Left Temporal-Right Temporal (LT-RT) regions provides significant information for theta (θ) and beta (β) ranges. The provided results have given a support to our hypothesis where we have claimed that a decrease in synchronization between temporal regions have a direct link with the progression of Alzheimer's disease. These findings will help clinicians for the early diagnosis of AD patients.

Chapter 7 Conclusions and Future Work

7.1 Introduction

This chapter presents the summary of the whole thesis along with the conclusions that have been derived from the results. It provides the novel contributions of the project from IT and clinical perspectives. This chapter also highlights the importance of the results from clinical point of view and their practical implementations in the hospitals with the help of computer engineers. Limitations and the possible future research directions are also provided in this chapter to get the maximum benefits from this research work.

7.2 Thesis Summary

This thesis has presented a framework for the early detection of neurodegenerative diseases using signal processing and signal classification techniques. As we have previously discussed that a NDD starts with the deterioration of short and long term memory of the subjects due to abnormal brain functionality. This kind of perturbation in the brain is considered an initial symptom for the progression of a neurodegenerative disease. Moreover, a healthy gait pattern requires a direct input from neurological system of the brain. Any disturbance in the brain has direct impact on the gait patterns of a person. So an abnormal gait pattern in these patients is considered a final symptom of any neurological disease.

To keep things in a normal sequence and to make them easily understandable, we have divided our project into two main parts—classification of gait signals to discriminate between different NDDs (Parkinson's disease, Huntington's disease, and Amyotrophic Lateral Sclerosis) and analysis of EEG signals to compute neural synchronization of MiAD patients. The first part of the project is completed using an online database repository called "Physionet" while the processing and analysis of EEG of MiAD and healthy subjects has been done in France at "SIGMA Laboratory".

Machine learning approach has been selected to complete this project. A set of eleven classification algorithms is implemented and evaluated using PRTools in Matlab for the gait pattern recognition. A comparison of various evaluation techniques is provided, based on visualization and statistical analysis, which helps us to understand the difference and importance of different performance evaluation techniques. In the second half of the thesis we have presented a novel idea of combining base-level classifiers to increase the percentage of classification accuracy. Lastly, we have presented the computation of neural synchronization of the EEG signals for MiAD and control subjects to determine the significant features that can help the clinicians to diagnose Alzheimer's at its earlier stage.

The chapter wise summaries of the whole thesis with their derived conclusions are provided below:

Chapter 1 outlined the potential issues of NDDs along with the challenges related to machine learning. We argued here, the potential of machine learning approaches to provide significant improvements in the early diagnosis of NDDs. It provided a brief introduction of the methods we have proposed in this research work. Finally, it outlined research aims and novel contributions of the thesis.

Chapter 2 presented detail information about neurodegenerative diseases and a description about their development stages with their probable symptoms. It provided background and preliminary information about signal processing and signal classification. Matters like feature extraction, feature selection and feature classification are discussed in this chapter. In addition to NDDs and signal processing, we have discussed supervised machine learning approach in detail. We have discussed logic based classifiers, artificial neural networks, and statistical learning algorithms.

Chapter 3 presented our proposed strategic framework for the early detection of neurodegenerative diseases and discussed each of its components, *i.e.* the data collection, data preprocessing and data classification and decision making. It provided information about the tools (PRTools, Statistical, Communication and Signal Processing tool) that are selected for this particular project and the reasons behind their selection.

In **Chapter 4**, we demonstrated the assessment of gait signals. In this chapter we have discussed the problems with imbalanced datasets, missing data entries, multiclass pattern recognition, and discrimination among similar diseases along with their possible solutions. A set of eleven statistical learning algorithms has been selected to process the gait signals of 16 CO, 15 PD, 20 HD, and 13 ALS subjects. They belonged to normal density based classifiers, linear and nonlinear classifiers. At first, the classification accuracy results are presented using *confusion matrix*. Later, the results are also presented using other visualization and statistical analysis techniques. Two Bayes classifiers (*LDC and UDC*) and one linear classifier (*Parzen*) have outperformed other.

Chapter 5 highlighted our novel idea for combining the *base-level* classifiers to check if we can obtain higher classification accuracy. Three base-level classifiers (*LDC, UDC, Parzen*) are combined together by six fixed combining rules. We observed that total mean error rate in case of combined classifier is less than base-level classifiers. Moreover, it has also been noted that voting combing rule has provided the highest accuracy rate as compare to other combining rules.

Chapter 6 presented the second half of our project that we have completed in France in SIGMA laboratory. Here we have analysed three different sets of MiAD and healthy subjects to compute the neural synchronization of EEG signals. Two novel methods are proposed to apply three neural synchrony measure techniques (phase synchrony, cross correlation and MS coherence) on three datasets. One of the methods is named *PCA* based synchrony while the second method is called *Average* synchrony. Results revealed that *PCA* based synchrony has given us more significant results that can help us to diagnose Alzheimer's at its earlier stage than *Average* synchrony. Moreover, cross correlation measure has proved to be the best one among others to provide better results. Results have been compared using Wilcoxon ranksum (Mann-Whitney) test. Later, Gram Schmidt orthogonalization algorithm is applied with "n-

probe" function to get the most important features that can help clinicians for the classification and early diagnosis of AD.

7.3 Research Contributions

The importance and the novel contributions of this research work can be assessed from two different perspectives—medical and IT fields. It has not only provided a very narrow research work on the causes and symptoms of the NDDs but has also highlighted those crucial moments where IT could play its significant role to diagnose the progression of these diseases from their onset to acute stages. Furthermore, it has explored more directions and innovations in the field of machine learning, signal processing and signal classification to get maximum benefits from this filed. Adding to the solutions, it has also discovered those hidden significant features that can be used as biomarkers from the early diagnosis of these life threatening diseases.

Following are the main contributions of the thesis:

- **Solutions for Skewed Datasets:** In Chapter 4, we have highlighted the limitations of imbalanced datasets in term of getting biased results for majority class patterns. Possible solutions in term of *under sampling* and *over-sampling* are also provided. Later, *over-sampling* method has been demonstrated using gait dynamics of different neurodegenerative diseases and control subjects.
- **Extended set of gait features:** Previous findings are based on a very limited set of features that are used to classify the neurodegenerative diseases. In our research work, we have selected all significant features that have direct or indirect impact on the progression of NDDs. For instance, age and gender have significant relation with NDDs. The likelihood of developing Alzheimer's increases in advance age. Also, we have calculated BMI of each person to notice if weight of the person has any link with the progression of neurological diseases. We have observed and demonstrated that all these factors are equally important in the early diagnosis of NDDs.
- **Different sets of classifiers:** As we already have mentioned that the classifier space that is considered for pattern recognition does not always contain the optimal classifier. For instance, a set of linear classifiers is chosen for a dataset that can best

recognize by nonlinear classifier can never help us to find an optimal classifier. Instead of focusing on a single classifier we have selected a number of different classifiers that belong to different categories—linear, nonlinear and Bayes classifiers.

- **Performance Evaluation Matrices:** It has already been demonstrated through literature survey that one particular performance measure may evaluate a classifier on a single perspective while fails to measure on another [112]. Although researchers have been evaluating classification algorithms by various techniques, yet there is no single authorized criterion that outperforms others. To overcome this complication, in our research work (Chapter 4), we have presented the results both by statistical and visualizing techniques. This has helped us to compare the results from different perspectives and eventually to select one of the classifiers that best suits to our dataset.
- **Classifiers Fusion Strategy:** In Chapter 5, we have proposed a novel idea to combine all those *base-level* classifiers that has given us comparatively better results. By combining different classifiers together, an opportunity is provided to the designer to have an access of different classifiers that belong to different context and are developed for entirely different representation. Moreover, an ensemble classifier can also handle the multivariate training sets that are collected at different times and also in different environment. The training set may also have different features. Results revealed that combining classifiers is good idea especially in a case where data belongs to multidimensional (different NDDs such as AD, PD, HD and ALS) as well as multivariate (a large set of features) datasets.
- **Dividing EEG Signals into Narrow frequency bands:** In Chapter 6, for the early diagnosis of Alzheimer's disease, we have filtered EEG signals into five (5) different frequency bands. These frequency ranges are delta (δ), theta (θ), alpha (α), beta (β) and gamma (γ). These narrow frequency bands helped us to find all those hidden patterns that can be used as biomarkers for the early diagnosis of Alzheimer's.
- **Different sets of EEG data:** A high inter-subject variability has been seen in the EEG signals of AD patients, especially with different levels of severity and comorbidities. In this situation the findings are not more reliable on a single set of data. Most of the existing studies focus on a single synchrony measure with a single set of data. In this case it's hard to compare the results to conclude a single

hypothesis. To extract a general set of feature we have analysed three different databases, each from one hospital at a time. This has not only increased the validity of our research but has also provided more reliable findings that can later be used in clinical applications.

- **PCA technique:** Literature survey does not provide us any obvious solution for removing the redundant information or noise from EEG dataset before applying synchrony measurement techniques. Analysing results without removing the redundant information or without eliminating the noise leads to non-reliable results. To remove the ambiguity of biased results due to "features redundancy" we have applied PCA (Principal Component Analysis) technique before applying synchrony measurement techniques (Chapter 6). This helped us finding more reliable results that can be used for clinical practices.
- **Average Method vs. PCA Method**: In Chapter 6, we have compared our proposed method *PCA* based synchrony measure with another proposed method called *Average* synchrony measure. Results revealed that although *Average* method provides us some useful information for the early diagnosis of Alzheimer's yet *PCA* method provides us more significant and authentic information that can be used as a biomarker for the early diagnosis of Alzheimer's.
- **Gait Signals Vs. EEG Signals:** Two different kinds of signals are computed in this research work—Gait and EEG signals. Gait signals are computed to discriminate different NDDs for an accurate and exact diagnosis of a disease and also to provide in time treatment of the patient. On the other hand, EEG signals are computed to detect any perturbation in the brain. Any change in the synchronization of EEG signals is an indication of onset of an abnormality/pathology in the brain.

7.4 Future Work

Beside the main contributions that have been presented for gait and EEG signals analysis, we hereby, present several extensions and possible changes that might help to improve the shortcomings. This section also highlights the possible extensions of our research work to clinical applications.

- **Possible extension of feature set for gait data analysis**: In this research work only the left and right feet signals are used with other set of possible variables. However, future work may enhance this feature set to advance level. For instance, calculating correlation and synchrony between two feet signals can provide some more significant information in the classification and later in the analysis of a normal and abnormal gait pattern.
- **Classification of EEG features and further implementation on other NDDs:** During the EEG signal processing, this research work has concentrated on exploring the significant features that can later be used as biomarkers in the early detection or early diagnosis of neurodegenerative diseases (here Alzheimer's). Significant features are extracted using Wilcoxon ranksum (Mann-Whitney) test. Future work may involve this feature set to discriminate the healthier and Alzheimer's patients using all those classifiers that have been implemented for gait signal analysis.

  Furthermore, in this research work, the analysis of EEG signals is confined to the early diagnosis of Alzheimer's. Same techniques and procedures can further be applied in the early diagnosis of Parkinson's, Huntington's, ALS and other NDDs for clinical practices.

- **Extension of the work to Brain-Computer Interface (BCI):** A future augmentation of this research work is a complete BCI system which measures the brain activity and delivers a feedback after the processing of data.

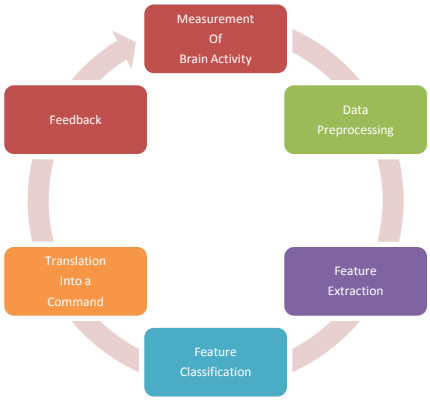


Figure 7-1: Activities of Brain Computer Interface

- **Integration of EEG with other techniques**: Combining EEG with other imaging modalities can provide us further remarkable results. For instance, Francois and his team have claimed in [186] that EEG with magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), Doppler technique, transactional magnetic stimulation (TMS), and single photon emission computed tomography (SPECT) can provide useful information about the anatomy of the brain that can help in the early diagnosing of Alzheimer's. Further adding to this argument they emphasise that multi-model approaches seem to have strong potential to diagnose all kinds of dementias. Theses interesting areas are yet to be explored and can add to this project to get maximum benefits from it.

  The principal drawback with EEG is its low spatial resolution due to its dependence on the number of electrodes that are used for extracting the EEG signals. However, fMRI resolves this issue by providing better spatial resolution on an order of millimetres. The integration of these techniques will help to implement this project in neuroscience studies and better solutions can be provided for other neurological diseases such as epilepsy, Seizure, depression and dementia.

7.5 Concluding Remarks

The problem with the NDDs is thet they are incurable, hard to detect at earlier stage due to non-obvious symptoms and also hard to discriminate at latter stage due to pattern similarities of different NDDs. Since, there is no single authentic remedy available for such diseases; scientists find a lot of interest in finding those hidden patterns that can help us in the early diagnosis of NDDs. This research work has highlighted the importance of machine learning and signal processing in the early diagnosis of life threatening diseases such as Alzheimer's Parkinson's, Huntington's and ALS. In this thesis, we have presented the issues with the early diagnosis of NDDs and also with their possible solutions. The analysis and classification of gait signals is presented using a set of well-known classifiers—linear, non-linear and Bayes classifiers. Results are presented and elaborated from various dimensions using more than one performance evaluation techniques. In addition, we have proposed and demonstrated a novel idea of combining classifiers to improve the classification accuracy.

The latter half of the thesis presented the implementation of neural synchrony measurement techniques using EEG signals to calculate synchronization in different parts of the brain for Alzheimer's and non-Alzheimer's. The research work has presented novel and significant findings that can be used in clinical practices for the early diagnosis of ND.

Appendix A Ranking of Features

The ranking of all 105 features by Gram-Schmidt Orthogonalization is presented in this table. Here the value in the first column “62” represents the index number of the feature with value “0.61532” that has highest importance in the early diagnosis of Alzheimer’s. Similarly, the other features are ranked accordingly.

62	0.61532	25	0.671679	17	0.466411	67	0.409222
63	0.616709	73	0.732044	45	0.46705	56	0.44928
37	0.346602	77	0.781445	105	0.437859	94	0.380699
10	0.328678	88	0.883991	97	0.355384	78	0.554023
64	0.346054	58	0.760874	102	0.383395	39	0.667188
4	0.391557	92	0.966538	32	0.385404	71	0.408724
74	0.532659	26	0.998456	101	0.482414	49	0.42052
7	0.433891	8	0.945611	91	0.436949	2	0.26958
3	0.369268	34	0.999767	19	0.459629	60	0.273432
23	0.385343	42	1	15	0.542015	21	0.381708
57	0.402352	16	0.837752	79	0.459497	103	0.359936
44	0.44419	27	0.672242	35	0.449874	84	0.238438
69	0.510834	61	0.55518	65	0.590408	11	0.173543
20	0.37626	18	0.632625	99	0.464793	14	0.015892
72	0.394932	86	0.441174	70	0.398272	12	0.000888
6	0.30568	50	0.348897	89	0.689701		
66	0.325174	90	0.339989	33	0.82886		
48	0.431468	82	0.539476	80	0.809667		
9	0.461903	22	0.575493	5	0.887011		
53	0.425815	24	0.479011	29	0.99994		
13	0.394697	96	0.519094	30	1		
46	0.388942	40	0.362037	75	0.940345		
1	0.39193	59	0.403948	98	0.745889		
93	0.488403	100	0.44972	41	0.595915		
68	0.531995	87	0.337776	31	0.605216		
76	0.580376	95	0.439673	47	0.655826		
51	0.593223	43	0.36418	28	0.624386		
55	0.555383	52	0.385324	54	0.29426		
85	0.475436	83	0.389958	36	0.435756		
104	0.658481	81	0.411989	38	0.445491		

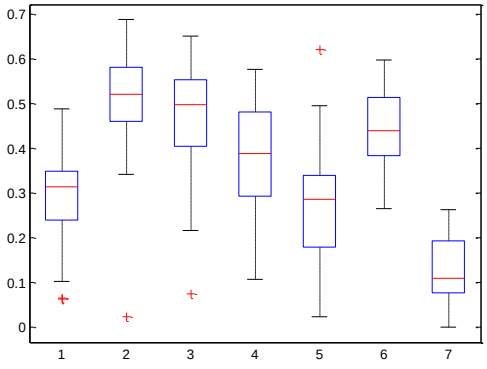
Appendix B Plotbox results

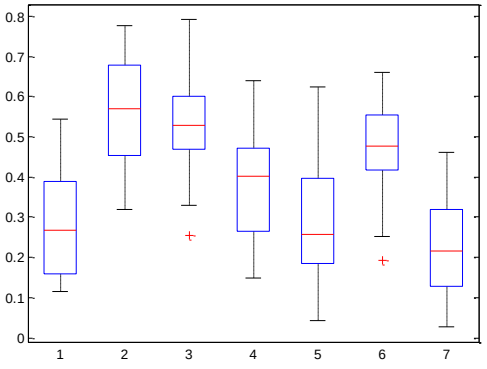
Previous Results without applying filter to the data

(Sampling frequency 128 Hz)

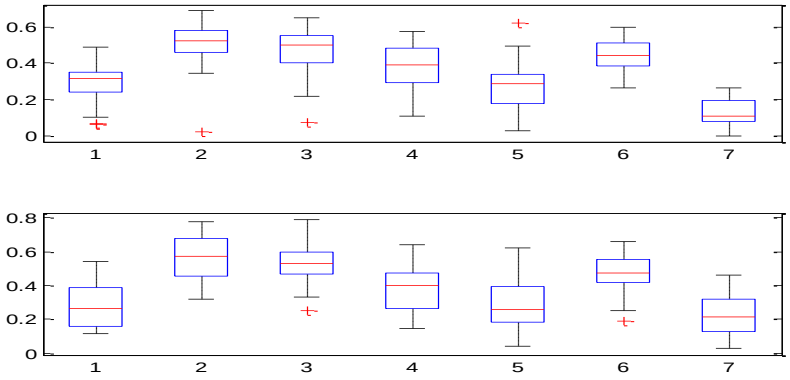
The 7 digits on X-axis represents the comparison of left temporal-right temporal, , left temporal and frontal, left temporal and occipital , left temporal and central, right temporal and frontal, right temporal and occipital and finally right temporal and central respectively . while Y-axis represents the median of phase synchrony (0-1).

For Alzheimer's Patients	For Control Subjects
AssA_AD_Boxplot (Dataset A)	AssA_Ctr_Ctr (Dataset A)

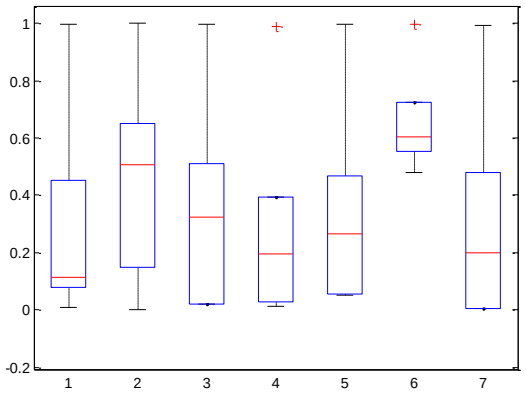




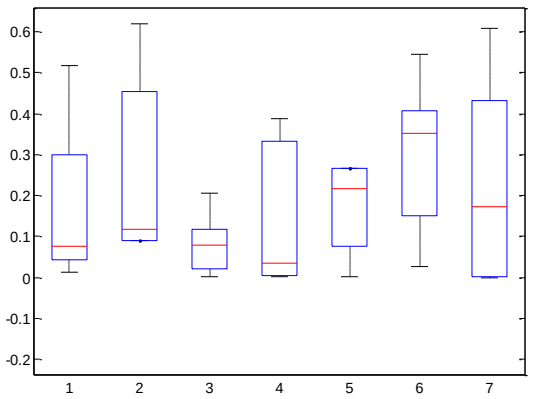
Comparison of Alzheimer's and Control (Dataset A)



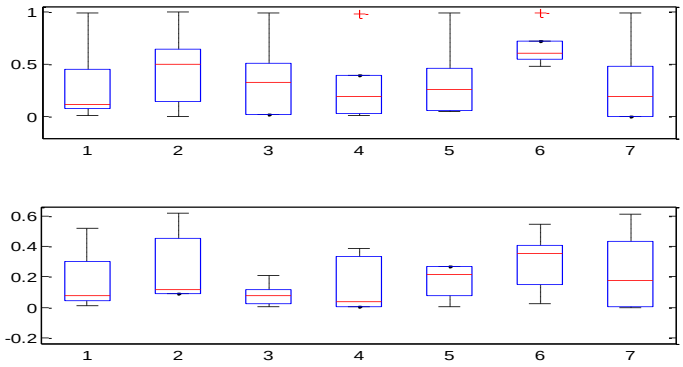
AssB_AD_boxplot (Dataset B)



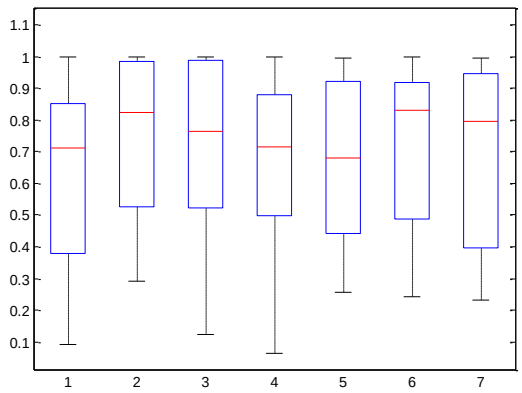
AssB_Ctr_boxplot (Dataset B)



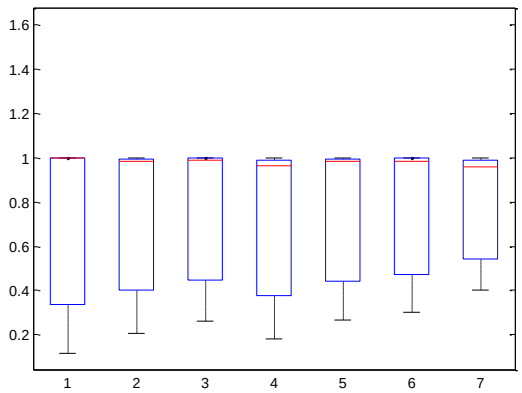
Comparison of Alzheimer's and Control (Dataset B)



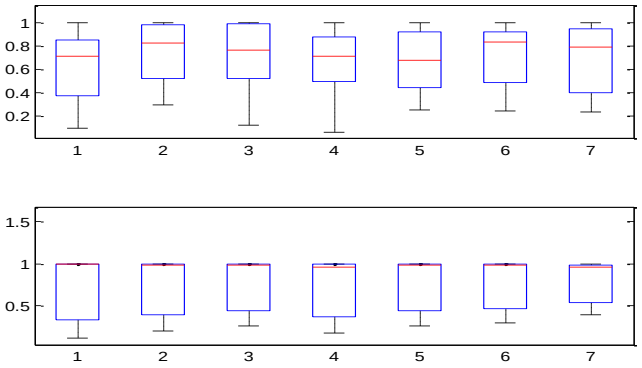
AssC_AD_boxplot (Dataset C)



AssC_Ctr_boxplot (Dataset C)



Comparison of Alzheimer's and Control (Dataset c)



Results of P-Values by using Ranksum function for Dataset B:

The P-Values of *Dataset B* and *Dataset C* are not as significant as of *Dataset A*. However, we observed that PCA based synchrony measure provides us relatively more significant values as compare to Average synchrony measures. This can be visualized in the results of *Dataset B* provided below.

Average Phase Synchrony:

0.54332	0.84127	0.8763	0.690476	0.6372	0.222222	0.690476
0.84127	0.309524	0.420635	0.150794	0.690476	0.309524	0.309524
0.84127	0.54332	0.690476	0.54333	0.84127	0.309524	0.547619
0.547619	0.420635	0.420635	0.309524	0.690476	0.150794	0.309524
0.84127	0.309524	0.420635	0.309524	0.690476	0.420635	0.309524

PCA based Phase Synchrony

0.547619	0.690476	0.84127	0.84127	0.690476	0.309524	0.045

0.84127	0.069048	0.042063	0.420635	0.420635	0.309524	0.0532
0.7544	0.547619	0.054762	0.084127	0.690476	0.420635	0.344
0.6788	0.690476	0.4367	0.4522	0.309524	0.309524	0.690476
0.84127	0.150794	0.4566	0.309524	0.150794	0.309524	0.2342

Average Cross Correlation

0.055556	0.031746	0.095238	0.055556	0.095238	0.055556	0.055556
0.015873	0.007937	0.007937	0.007937	0.007937	0.015873	0.007937
0.222222	0.150794	0.222222	0.095238	0.095238	0.095238	0.150794
0.9867	0.8755	0.84127	0.8677	0.7653	0.84127	0.9873
0.095238	0.095238	0.095238	0.095238	0.015873	0.095238	0.095238

PCA based Cross Correlation

0.73224	0.7003	0.797479	0.04003	0.764651	0.73224	0.764651
0.7003	0.638031	0.7003	0.607796	0.668881	0.668881	0.638031
0.931835	0.864166	0.931835	0.864166	0.73224	0.5344	0.02355
0.05433	0.931835	0.965886	0.965886	0.06543	0.965886	0.04533

0.73224	0.764651	0.73224	0.83067	0.764651	0.73224	0.0668881

Average MS Coherence

0.63455	0.84127	0.84127	0.690476	0.7866	0.420635	0.84127
0.690476	0.84127	0.84127	0.420635	0.8677	0.309524	0.420635
0.8325	0.84127	0.420635	0.690476	0.9667	0.309524	0.690476
0.420635	0.309524	0.309524	0.309524	0.8433	0.547619	0.222222
0.420635	0.547619	0.690476	0.309524	0.84127	0.84127	0.3542

PCA based MS Coherence

0.690476	0.690476	0.690476	0.26788	0.3827	0.309524	0.84127
0.690476	0.690476	0.547619	0.309524	0.690476	0.420635	0.84127
0.53662	0.84127	0.690476	0.6756	0.547619	0.84127	0.5342
0.547619	0.690476	0.420635	0.690476	0.420635	0.095238	0.690476
0.420635	0.547619	0.4533	0.309524	0.309524	0.222222	0.84127

Bibliography

[1] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," ed, 2007.

[2] J. Rogers, S. Webster, L.-F. Lue, L. Brachova, W. Harold Civin, M. Emmerling, B. Shivers, D. Walker, and P. McGeer, "Inflammation and Alzheimer's disease pathogenesis," *Neurobiology of Aging,* vol. 17, pp. 681-686, 1996.

[3] H. Braak and E. Braak, "Neuropathological stageing of Alzheimer-related changes," *Acta Neuropathol,* vol. 82, pp. 239-59, 1991.

[4] L. E. Hebert, B. L. A., S. P. A., and D. A. Evans, "Annual Incidence of Alzheimer Disease in the United States Projected to the Years 2000 Through 2050," *Alzheimer Disease & Associated Disorders,* vol. 15, pp. 169-173, 2001.

[5] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C. K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease," *Applied Physiology,* vol. 82, pp. 262–269, 1997.

[6] W. Yunfeng and S. Krishnan, "Statistical Analysis of Gait Rhythm in Patients With Parkinson's Disease," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 18, pp. 150-158, 2010.

[7] M. Banaie, M. Pooyan, and M. Mikaili, "Introduction and application of an automatic gait recognition method to diagnose movement disorders that arose of similar causes," in *Expert Syst. Appl.* vol. 38, ed, 2011, pp. 7359-7363.

[8] H. Berger, "Über das Elektrenkephalogramm des Menschen," *Archiv für Psychiatrie und Nervenkrankheiten,* vol. 87, pp. 527-570, 1929/12/01 1929.

[9] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer, and M. Scazufca, "Global prevalence of dementia: a Delphi consensus study," *The Lancet,* vol. 366, pp. 2112-2117, 2005.

[10] V. S. Sierpina and M. J. Kreitzer, "Life-Long Learning in Integrative Healthcare," *EXPLORE: The Journal of Science and Healing,* vol. 8, pp. 210-212, 2012.

[11] J. Hou and B. Pakkenberg, "Age-related degeneration of corpus callosum in the 90+ years measured with stereology," *Neurobiology of Aging,* vol. 33, pp. 1009.e1-1009.e9, 2012.

[12] H. Hampel, D. Prvulovic, S. Teipel, F. Jessen, C. Luckhaus, L. Frölich, M. W. Riepe, R. Dodel, T. Leyhe, L. Bertram, W. Hoffmann, and F. Faltraco, "The future of Alzheimer's disease: The next 10 years," *Progress in Neurobiology,* vol. 95, pp. 718-728, 2011.

[13] A. Harvey, A. Brand, S. T. Holgate, L. V. Kristiansen, H. Lehrach, A. Palotie, and B. Prainsack, "The future of technologies for personalised medicine," *New Biotechnology*.

[14] G. Gillespie. (2000) Deploying an I.T. cure for chronic diseases. *US National Library of Medicine National Institutes of Health*. 68-70.

[15] D. Buchtela, Z. Anger, J. Peleska, A. Vesely, and J. Zvarova, "Presentation of medical guidelines on a computer," presented at the Studies In Health Technology And Informatics 2004.

[16] R. S. Turner, "Biomarkers of Alzheimer's disease and mild cognitive impairment: are we there yet?," *Experimental Neurology,* vol. 183, pp. 22-26, 2003.

[17] C. Plant, C. Sorg, V. Riedl, and A. Wohlschl, "Homogeneity-based feature extraction for classification of early-stage alzheimer's disease from functional magnetic resonance images," presented at the Proceedings of the 2011 workshop on Data mining for medicine and healthcare, San Diego, California, USA, 2011.

[18] F. Ganji, S. Abadeh, M. Hedayati, and N. Bakhtiari, "Fuzzy classifcation of imbalanced data sets for medical diagnosis," in *17th Iranian Conference of Biomedical Engineering (ICBME),* 2010, pp. 1-5.

[19] B. M. Marlin, "Missing data problems in machine learning," University of Toronto, 2008.

[20] X.-M. Zhao, X. Li, L. Chen, and K. Aihara, "Protein classification with imbalanced data," *Proteins: Structure, Function, and Bioinformatics,* vol. 70, pp. 1125-1132, 2008.

[21] Z. Zhi-Hua and L. Xu-Ying, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 18, pp. 63-77, 2006.

[22] A. Ani and M. Deriche, "A New Technique for Combining Multiple Classifiers using The Dempster-Shafer Theory of Evidence," *Journal of Artificial Intelligence Research,* vol. 17, pp. 333-361, 2002.

[23] P. Meriggi, P. Castiglioni, F. Rizzo, V. Gower, R. Andrich, M. Rabuffetti, M. Ferrarin, and M. Di Rienzo, "Potential role of wearable, ambulatory and home monitoring systems for patients with neurodegenerative diseases and their caregivers," in *5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)* 2011, pp. 316-319.

[24] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The Anatomy of a Context-Aware Application," *Wireless Networks,* vol. 8, pp. 187-97, 2002.

[25] A. Wimo, B. Winblad, and L. Jönsson, "An estimate of the total worldwide societal costs of dementia in 2005," *Alzheimer's & Dementia,* vol. 3, pp. 81-91, 2007.

[26] A. Wimo, B. Winblad, and L. Jönsson, "The worldwide societal costs of dementia: Estimates for 2009," *Alzheimer's & Dementia,* vol. 6, pp. 98-103, 2010.

[27] A. Abbott, "Dementia: A problem for our age," *Nature*, vol. 475, pp. S2-S4, 2011.

[28] R. Barker, *2030: The Future of Medicine: Avoiding a Medical Meltdown*. New York: Oxford university New York, 2011.

[29] R. A. Armstrong, N. J. Cairns, R. Patel, P. L. Lantos, and M. N. Rossor, "Relationships between β-amyloid (Aβ) deposits and blood vessels in patients with sporadic and familial Alzheimer's disease," *Neuroscience Letters*, vol. 207, pp. 171-174, 1996.

[30] J. Parkinson, "An Essay on the Shaking Palsy," *Neuropsychiatry and Clinical Neurosciences*, vol. 14, pp. 223-236, 2002.

[31] H. Allain, D. Bentué-Ferrer, and Y. Akwa, "Disease-modifying drugs and Parkinson's disease," *Progress in Neurobiology*, vol. 84, pp. 25-39, 2008.

[32] L. Findley, M. Aujla, P. G. Bain, M. Baker, C. Beech, C. Bowman, J. Holmes, W. K. Kingdom, D. G. MacMahon, V. Peto, and J. R. Playfer, "Direct economic impact of Parkinson's disease: a research survey in the United Kingdom.," *Mov Disord*, vol. 18, pp. 1139-45, 2003.

[33] J. Beck. (2012). *Parkinson's Disease Foundation*. Available: http://www.pdf.org/en/about_pd

[34] J.-G. G. Hou and E. C. Lai, "Non-motor Symptoms of Parkinson's Disease," *International Journal of Gerontology*, vol. 1, pp. 53-64, 2007.

[35] S. Dufrasne, M. Roy, M. Galvez, and D. S. Rosenblatt, "Experience over fifteen years with a protocol for predictive testing for Huntington disease," *Molecular Genetics and Metabolism*, vol. 102, pp. 494-504, 2011.

[36] D. Langbehn, R. Brinkman, D. Falush, J. Paulsen, and M. Hayden, "A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length," *Clinical Genetics*, vol. 65, pp. 267–277, 2004.

[37] D. Neela and K. Rangarajan, "Hybrid Workflow Net Based Architecture for Modeling Huntington's Disease," in *Bio-Inspired Computing: Theories and Applications (BIC-TA), 2011 Sixth International Conference on*, 2011, pp. 319-323.

[38] Y. Wang, Q. Chen, and W. Hu, "Behavior Selection Mechanism of Two Typical Brain Movement Disorders: Comparative Study Using Robot," in *Digital Manufacturing and Automation (ICDMA), 2010 International Conference on*, 2010, pp. 319-323.

[39] P. K. Kasi, L. S. Krivickas, M. Meister, E. Chew, P. Bonato, M. Schmid, G. Kamen, L. Pu, and E. A. Clancy, "Characterization of motor unit behavior in patients with amyotrophic lateral sclerosis," in *4th International IEEE/EMBS Conference on Neural Engineering, 2009. NER '09.* , 2009, pp. 10-13.

[40] F. Gros-Louis, C. Gaspar, and G. A. Rouleau, "Genetics of familial and sporadic amyotrophic lateral sclerosis," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1762, pp. 956-972, 2006.

[41] J. Matías-Guiu, L. Galán, R. García-Ramos, J. A. Barcia, and A. Guerrero, "Cerebrospinal fluid cytotoxicity in lateral amyotrophic sclerosis," *Neurología (English Edition)*, vol. 25, pp. 364-373, 2010.

[42] D. C. Dugdale. (2010). *Amyotrophic lateral sclerosis and other motor neuron diseases*. Available: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001708/

[43] T. Madarame, H. Tanaka, T. Inoue, M. Kamata, and M. Shino, "The development of a brain computer interface device for amyotrophic lateral sclerosis patients," in *IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008.*, 2008, pp. 2401-2406.

[44] A. s. Association. (2007, Stages of Alzheimer's disease. 1-3.

[45] E. Scherder, L. Eggermont, C. Visscher, P. Scheltens, and D. Swaab, "Understanding higher level gait disturbances in mild dementia in order to improve rehabilitation: 'Last in–first out'," *Neuroscience and Biobehavioral Reviews*, vol. 35, pp. 699-714, 2011.

[46] Y. F. Zhu and J. L. Henry, "Excitability of Aβ sensory neurons is altered in an animal model of peripheral neuropathy," *BMC Neuroscience*, vol. 13, pp. 1-15, 2012.

[47] H. J. Woodford and J. George, "Neurological and cognitive impairments detected in older people without a diagnosis of neurological or cognitive disease," *Postgraduate Medical Journal*, vol. 87, pp. 199-206, March 1, 2011 2011.

[48] S. V. Vaseghi. (2006). *Advanced Digital Signal Processing and Noise Reduction*.

[49] J. Han, H. S. Jeon, B. S. Jeon, and K. S. Park, "Gait detection from three dimensional acceleration signals of ankles for patients with Parkinson's disease," in *Proceedings of the international conference on technology and applications in biomedicine,*, 2008, pp. 349–352.

[50] M Hausdorff , A Lertratanakul , E Cudkowicz , L Peterson , D Kaliton, and L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J Applied Physiology*, vol. 88, pp. 2045-2053, 2000.

[51] R. Q. Quiroga, "Quantitative analysis of EEG signals: Time-frequency methods and Chaos theory," Ph.D, Institute of Signal Processing, - Medical University Lubeck, 1998.

[52] S. Dalal and L. Malik, "A Survey of Methods and Strategies for Feature Extraction in Handwritten Script Identification," presented at the First International Conference on Emerging Trends in Engineering and Technology, 2008.

[53] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. NewYork: John Wiley & Sons, 2004.

[54] C. Guerrero-Mosquera, M. Verleysen, and A. Navia Vazquez, "Dimensionality reduction for EEG classification using Mutual Information and SVM," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2011, pp. 1-6.

[55] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205-1224, 2004.

[56] Z. Zheng, "Constructing conjunctions using systematic search on decision trees," *Knowledge-Based Systems*, vol. 10, pp. 421-430, 1998.

[57] J. R. Quinlan, *C4.5: programs for machine learning*: Morgan Kaufmann Publishers Inc., 1993.

[58] J. Fürnkranz, "Pruning Algorithms for Rule Learning," *Machine Learning*, vol. 27, pp. 139-172, 1997/05/01 1997.

[59] N. Lavrac, P. A. Flach, and B. Zupan, "Rule Evaluation Measures: A Unifying View," presented at the Proceedings of the 9th International Workshop on Inductive Logic Programming, 1999.

[60] P. Jeatrakul and K. W. Wong, "Comparing the performance of different neural networks for binary classification problems," in *Eighth International Symposium on Natural Language Processing, 2009.*, 2009, pp. 111-115.

[61] R. Ball and P. Tissot, "Demonstration of Artificial Neural Network in Matlab," in *Division of Nearhsore Research*, ed: Texas A&M University, 2009.

[62] D. Anderson and G. McNeill, "Artificial Neural Network Technology, A DACS State-of-the-Art Report," *Kaman Sciences Corporation*, 1992.

[63] P. Jeatrakul and K. W. Wong, "Comparing the performance of different neural networks for binary classification problems," in *Eighth International Symposium on Natural Language Processing, 2009. SNLP '09.* , 2009, pp. 111-115.

[64] W. Yao and J. Vassileva, "Bayesian network-based trust model," in *IEEE/WIC International Conference on Web Intelligence*, 2003, pp. 372-378.

[65] E. Charniak, "Bayesian networks without tears," *AI Mag.*, vol. 12, pp. 50-63, 1991.

[66] F. Lotte, "Study of Electroencephalographic Signal Processing and Classification Techniques towards the use of Brain-Computer Interfaces in Virtual Reality Applications " PhD, National Institute of Applied Sciences in Rennes, Rennes, 2009.

[67] T. Lissack and F. King-Sun, "Error estimation in pattern recognition via L-distance between posterior density functions," *IEEE Transactions on Information Theory*, vol. 22, pp. 34-45, 1976.

[68] W. Duin, P. Paclik, and D. Tax, "Exercises with PRTools," ed Netherland: Information and Communication Theory Group, Delft University of Technology, 2009, pp. 1-54.

[69] A. R. Webb, *Statistical Pattern Recognition*: John Wiley & Sons, 2002.

[70] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* vol. 2: John Wiley and Sons, Inc., New York, 2001.

[71] C. L. Liu and H. Sako, "Class-specific feature polynomial classifier for pattern classification and its application to handwritten numeral recognition," *The Journal of Pattern Recognition Society,* vol. 39, pp. 669 – 681, 2006.

[72] P. Argentiero, R. Chin, and P. Beaudet, "An Automated Approach to the Design of Decision Tree Classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 4, pp. 51-57, 1982.

[73] C. w. Hsu, C. c. Chang, and C. j. Lin, "A Practical Guide to Support Vector Classification," *Bioinformatics,* vol. 1, pp. 1-6, 2010.

[74] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*: Diane Cerra, 2000.

[75] L. Bouarfa, A. Schneider, H. Feussner, N. Navab, H. U. Lemke, P. P. Jonker, and J. Dankelman, "Prediction of intraoperative complexity from preoperative patient data for laparoscopic cholecystectomy," *Artificial Intelligence in Medicine,* vol. 52, pp. 169-176, 2011.

[76] R.-H. Lin, "An intelligent model for liver disease diagnosis," *Artificial Intelligence in Medicine,* vol. 47, pp. 53-62, 2009.

[77] M. C. Lee, L. Boroczky, K. Sungur-Stasik, A. D. Cann, A. C. Borczuk, S. M. Kawut, and C. A. Powell, "Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction," *Artificial Intelligence in Medicine,* vol. 50, pp. 43-53, 2010.

[78] S.-Q. Shi, W. L. Maner, L. B. Mackay, and R. E. Garfield, "Identification of term and preterm labor in rats using artificial neural networks on uterine electromyography signals," *American Journal of Obstetrics and Gynecology,* vol. 198, pp. 235.e1-235.e4, 2008.

[79] D. Long, J. Wang, M. Xuan, Q. Gu, X. Xu, D. Kong, and M. Zhang, "Automatic Classification of Early Parkinson's Disease with Multi-Modal MR Imaging," *PLOS ONE,* vol. 7, p. e47714, 2012.

[80] C. A. Miller and M. K. Hinders, "Classification of flaw severity using pattern recognition for guided wave-based structural health monitoring," *Ultrasonics,* 2013.

[81] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. d. Ridder, D. Tax, and S. Verzakov, "Pr-tools 4.1, a matlab toolbox for pattern recognition," 2007.

[82] M. Budka and B. Gabrys, "Ridge regression ensemble for toxicity prediction," *Procedia Computer Science,* vol. 1, pp. 193-201, 2010.

[83] M. L. Grimes, W. J. Lee, L. van der Maaten, and P. Shannon, "Wrangling phosphoproteomic data to elucidate cancer signaling pathways," *PLOS ONE,* vol. 8, p. e52884, 2013.

[84] L. Fu, H. Li, H. Wang, B. Xu, Y. Fan, and J. Tian, "SUVmax/THKmax as a biomarker for distinguishing advanced gastric carcinoma from primary gastric lymphoma," *PLOS ONE,* vol. 7, p. e50914, 2012.

[85] S. Iram, D. Al-jumeily, P. Fergus, M. Randles, and A. Hussain, "Computational Data Analysis for Movement Signals Based on Statistical Pattern Recognition Techniques for Neurodegenerative Diseases," in *13th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting*, 2012, pp. 1-6.

[86] W. Hurst, M. Merabti, S. Iram, and P. Fergus, "Protecting critical infrastructures through behavioural observation," *International Journal of Critical Infrastructures*, 2013.

[87] P. Fergus, P. Cheung, A. Hussain, S. Iram, C. Dobbins, and D. Al-Jumeily, "Prediction of Preterm Deliveries from EHG Signals using Machine Learning," *PLOS ONE*, 2013.

[88] K. Revett, F. Gorunescu, and A.-B. M. Salem, "Feature selection in Parkinson's disease: A rough sets approach," in *International Multiconference on Computer Science and Information Technology, 2009. IMCSIT '09.* , pp. 425 – 428.

[89] C. Kennard, "Extrapyramidal symptoms (EPS) and Alzheimer's disease," in *Health's Disease and Condition* vol. 2012, M. R. Board, Ed., ed, 2006.

[90] J. P. Taylor, J. Hardy, and K. H. Fischbeck, "Toxic Proteins in Neurodegenerative Disease," *Science*, vol. 296, pp. 1991-1995 2002.

[91] W. L. Woon, A. Cichocki, F. Viallate, and T. Musha, "Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings," presented at the IOP Publishing 2007.

[92] S. Iram, D. Al-Jumeily, P. Fergus, M. Randles, and M. Davies, "E-Health: The Potential of Linked Data and Stream Reasoning for Personalised Healthcare," in *Developments in E-systems Engineering (DeSE), 2011*, 2011, pp. 46-49.

[93] K. van der Hiele, A. A. Vein, R. H. A. M. Reijntjes, R. G. J. Westendorp, E. L. E. M. Bollen, M. A. van Buchem, J. G. van Dijk, and H. A. M. Middelkoop, "EEG correlates in the spectrum of cognitive decline," *Clinical Neurophysiology*, vol. 118, pp. 1931-1939, 2007.

[94] Z. J. Koles, J. C. Lind, and P. Flor-Henry, "Spatial patterns in the background EEG underlying mental disease in man," *Electroencephalography and Clinical Neurophysiology*, vol. 91, pp. 319-328, 1994.

[95] D. H. Mantzaris, G. C. Anastassopoulos, and D. K. Lymberopoulos, "Medical disease prediction using Artificial Neural Networks," in *8th IEEE International Conference on BioInformatics and BioEngineering*, 2008, pp. 1-6.

[96] E. Parzen, "On Estimation of a Probability Density Function and Mode," 1962.

[97] C.-F. V. Latchoumane, F.-B. Vialatte, J. Solé-Casals, M. Maurice, S. R. Wimalaratna, N. Hudson, J. Jeong, and A. Cichocki, "Multiway array decomposition analysis of EEGs in Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 207, pp. 41-50, 2012.

[98] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, B\, \#252, and l. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics,* vol. 23, pp. i10-i18, 2007.

[99] C. A. Andersson and R. Bro, "The N-way Toolbox for MATLAB," *Chemometrics and Intelligent Laboratory Systems,* vol. 52, pp. 1-4, 2000.

[100] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," in *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on*, 2008, pp. 108-115.

[101] S. Joshi, V. Simha, D. Shenoy, V. K, and P. M, "Classification and treatment of different stages of alzheimer's disease using various machine learning methods," *International Journal of Bioinformatics Research,* vol. 2, pp. 44-52, 2010.

[102] M. Orozco, M. E. García, R. P. W. Duin, and C. G. Castellanos, "Dissimilarity Based Classification of Seismic Signals at Nevado Del Ruiz Volcano," *Earth Science Research* vol. 10, pp. 57-65, 2006.

[103] U. Knauer and B. Meffert, "Evaluation based combining of classifiers for monitoring honeybees," presented at the Workshop on Applications of Computer Vision (WACV), Germany, 2009.

[104] A. S. Ghanem, S. Venkatesh, and G. West, "Multi-class Pattern Classification in Imbalanced Data," presented at the Proceedings of the 20th International Conference on Pattern Recognition, 2010.

[105] D. Xiong, F. Xiao, L. Liu, K. Hu, Y. Tan, S. He, and X. Gao, "Towards a better detection of horizontally transferred genes by combining unusual properties effectively," *PLOS ONE,* vol. 7, p. e43126, 2012.

[106] T. Sun, R. Zhang, J. Wang, X. Li, and X. Guo, "Computer-Aided Diagnosis for Early-Stage Lung Cancer Based on Longitudinal and Balanced Data," *PLoS One* vol. 8, p. e63559, 2013.

[107] M. Beigi and A. Zell, "SPSO: Synthetic Protein Sequence Oversampling for Imbalanced Protein Data and Remote Homology Detection," in *Biological and Medical Data Analysis*. vol. 4345, N. Maglaveras, I. Chouvarda, V. Koutkias, and R. Brause, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 104-115.

[108] C.-Y. Chang, M.-T. Hsu, E. X. Esposito, and Y. J. Tseng, "Oversampling to Overcome Overfitting: Exploring the Relationship between Data Set Composition, Molecular Descriptors, and Predictive Modeling Methods," *Journal of Chemical Information and Modeling,* vol. 53, pp. 958-971, 2013/04/22 2013.

[109] X. T. Dang, O. Hirose, D. H. Bui, T. Saethang, V. A. Tran, L. A. T. Nguyen, T. K. T. Le, M. Kubo, Y. Yamada, and K. Satou, "A Novel Over-Sampling Method and its Application to Cancer Classification from Gene Expression Data," *Chem-Bio Informatics,* vol. 13, pp. 19-29, 2013.

[110] A. Fernández, S. García, M. J. d. Jesus, and F. Herrera, "A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets," *Fuzzy Sets Syst.*, vol. 159, pp. 2378-2398, 2008.

[111] N. Seliya, T. M. Khoshgoftaar, and J. Van Hulse, "Aggregating performance metrics for classifier evaluation," in *IEEE International Conference on Information Reuse & Integration*, 2009, pp. 35-40.

[112] N. Japkowicz, "Classifier evaluation: A need for better education and restructuring," in *Proceedings of the 3rd Workshop on Evaluation Methods for Machine Learning, ICML 2008*, Helsinki, Finland, 2008.

[113] N. D. Marom, L. Rokach, and A. Shmilovici, "Using the confusion matrix for improving ensemble classifiers," in *26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, 2010, pp. 000555-000559.

[114] C. G. Weng and J. Poon, "A New Evaluation Measure for Imbalanced Datasets," in *Proc. Seventh Australasian Data Mining Conference (AusDM 2008)*, South Australia, 2008, pp. 27-32.

[115] T. C. W. Landgrebe and R. P. W. Duin, "Approximating the multiclass ROC by pairwise analysis," *Pattern Recogn. Lett.*, vol. 28, pp. 1747-1758, 2007.

[116] P. Simeone, C. Marrocco, and F. Tortorella, "Shaping the error-reject curve of error correcting output coding systems," presented at the Proceedings of the 16th international conference on Image analysis and processing, Ravenna, Italy, 2011.

[117] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226-239, 1998.

[118] T. G. Dietterich, "Ensemble Methods in Machine Learning," presented at the Proceedings of the First International Workshop on Multiple Classifier Systems, 2000.

[119] A. K. Jain, R. P. W. Duin, and M. Jianchang, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 22, pp. 4-37, 2000.

[120] L. A. Alexandre, A. C. Campilho, and M. Kamel, "On combining classifiers using sum and product rules," *Pattern Recognition Letters*, vol. 22, pp. 1283-1289, 2001.

[121] T. Woloszynski and M. Kurzynski, "Combining Classifiers in a Tree Structure," in *Computational Intelligence for Modelling Control & Automation, 2008 International Conference on*, 2008, pp. 785-790.

[122] M. Mejdoubi, D. Aboutajdine, M. A. Kerroum, and A. Hammouch, "Combining classifiers using Dempster-Shafer evidence theory to improve remote sensing images classification," in *International Conference on Multimedia Computing and Systems (ICMCS)* 2011, pp. 1-4.

[123] M. P. Ponti and J. P. Papa, "Improving accuracy and speed of optimum-path forest classifier using combination of disjoint training subsets," presented at the Proceedings of the 10th international conference on Multiple classifier systems, Naples, Italy, 2011.

[124] V. Tresp and M. Taniguchi, "Combining Estimators Using Non-Constant Weighting Functions," presented at the Neural Information Processing Systems, 1994.

[125] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.*, vol. 3, pp. 79-87, 1991.

[126] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics, ,* vol. 22, pp. 418-435, 1992.

[127] R. P. W. Duin and D. M. J. Tax, "Experiments with Classifier Combining Rules," presented at the Proceedings of the First International Workshop on Multiple Classifier Systems, 2000.

[128] R. P. W. Duin, "The combining classifier: to train or not to train?," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, pp. 765-770 vol.2.

[129] L. Todorovski and S. Dzeroski, "Combining Classifiers with Meta Decision Trees," *Mach. Learn.*, vol. 50, pp. 223-249, 2003.

[130] C. A. e. G. Mosquera, "Artifact EOG Removal by ICA-RLS Scheme and Tracks Extraction Method," Ph.D, Signal Processing, Universidad Carlos III de Madrid, 2011.

[131] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and Clinical Neurophysiology*, vol. 66, pp. 376-382, 1987.

[132] J. Dauwels, F. Vialatte, and A. Cichocki, "A Comparative Study of Synchrony Measures for the Early Detection of Alzheimer's Disease Based on EEG," in *Neural Information Processing*. vol. 4984, ed: Springer Berlin Heidelberg, 2008, pp. 112-125.

[133] K. Yun, K. Watanabe, and S. Shimojo, "Interpersonal body and neural synchronization as a marker of implicit social interaction," *Scientific Reports*, vol. 2, 2012.

[134] L. M. Ward, "Synchronous neural oscillations and cognitive processes," *Trends in Cognitive Sciences*, vol. 7, pp. 553-559, 2003.

[135] L. M. W. mail, S. E. MacLean, and A. Kirschner, "Stochastic resonance modulates neural synchronization within and between cortical sources," *PLoS One.*, vol. 5, p. e14371, 2010.

[136] C. Babiloni, R. Ferri, G. Binetti, A. Cassarino, G. D. Forno, M. Ercolani, F. Ferreri, G. B. Frisoni, B. Lanuzza, C. Miniussi, F. Nobili, G. Rodriguez, F. Rundo, C. J. Stam,

T. Musha, F. Vecchio, and P. M. Rossini, "Fronto-parietal coupling of brain rhythms in mild cognitive impairment: A multicentric EEG study," *Brain Research Bulletin,* vol. 69, pp. 63-73, 2006.

[137] C. Babiloni, G. B. Frisoni, M. Pievani, F. Vecchio, R. Lizio, M. Buttiglione, C. Geroldi, C. Fracassi, F. Eusebi, R. Ferri, and P. M. Rossini, "Hippocampal volume and cortical sources of EEG alpha rhythms in mild cognitive impairment and Alzheimer disease," *NeuroImage,* vol. 44, pp. 123-135, 2009.

[138] G. B. Frisoni, A. Padovani, and L. O. Wahlund, "The predementia diagnosis of Alzheimer disease," *Alzheimer Dis Assoc Disord,* vol. 18, pp. 51-53, 2004.

[139] B. Jelles, P. Scheltens, W. M. van der Flier, E. J. Jonkman, F. H. L. da Silva, and C. J. Stam, "Global dynamical analysis of the EEG in Alzheimer's disease: Frequency-specific changes of functional interactions," *Clinical Neurophysiology,* vol. 119, pp. 837-841, 2008.

[140] M. C. Faustino, R. P. Serquiña, P. E. Rapp, and A. M. Albano, "Phase synchronization of electroencephalographic signals in the different frequency bands," *Philippine Science Letters,* vol. 5, pp. 131-137, 2012

[141] A. A. Fingelkurts, A. A. Fingelkurts, and S. Kähkönen, "Functional connectivity in the brain—is it an elusive concept?," *Neuroscience & Biobehavioral Reviews,* vol. 28, pp. 827-836, 2005.

[142] X. Delbeuck, M. Van der Linden, and F. Collette, "Alzheimer' Disease as a Disconnection Syndrome?," *Neuropsychology Review,* vol. 13, pp. 79-92, 2003.

[143] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human brain mapping,* vol. 28, pp. 1178-1193, 2007.

[144] W. Singer, "Neuronal synchrony: a versatile code for the definition of relations?," *Neuron,* vol. 24, 1999.

[145] P. Fries, "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence," *Trends in Cognitive Sciences,* vol. 9, pp. 474-480, 2005.

[146] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: phase synchronization and large-scale integration," *Nature reviews. Neuroscience,* vol. 2, pp. 229-239, 2001.

[147] E. Gallego-Jutgla, M. Elgendi, F. Vialatte, J. Sole-Casals, A. Cichocki, C. Latchoumane, J. Jaesung, and J. Dauwels, "Diagnosis of Alzheimer's disease from EEG by means of synchrony measures in optimized frequency bands," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2012*, 2012, pp. 4266-4270.

[148] M. Breakspear, "Nonlinear phase desynchronization in human electroencephalographic data," *Hum Brain Mapp,* vol. 15, pp. 175-198, 2002.

[149] A. Bruns, "Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches?," *Journal of Neuroscience Methods,* vol. 137, pp. 321-332, 2004.

[150] R. Anghinah, P. A. Kanda, M. S. Jorge, E. E. Lima, L. Pascuzzi, and A. C. Melo, "Alpha band coherence analysis of EEG in healthy adult's and Alzheimer's type dementia patients," *Arq Neuropsiquiatr,* vol. 58, pp. 272-275, 2000.

[151] T. Locatelli, M. Cursi, D. Liberati, M. Franceschi, and G. Comi, "EEG coherence in Alzheimer's disease," *Electroencephalography and Clinical Neurophysiology,* vol. 106, pp. 229-237, 1998.

[152] J. J. Claus, R. L. M. Strijers, E. J. Jonkman, B. W. Ongerboer de Visser, C. Jonker, G. J. M. Walstra, P. Scheltens, and W. A. van Gool, "The diagnostic value of electroencephalography in mild senile Alzheimer's disease," *Clinical Neurophysiology,* vol. 110, pp. 825-832, 1999.

[153] G. Jóhannesson, A. Brun, I. Gustafson, and D. H. Ingvar, "EEG in presenile dementia related to cerebral blood flow and autopsy findings," *Acta Neurologica Scandinavica,* vol. 56, pp. 89-103, 1977.

[154] B. Szelies, M. Grond, K. Herholz, J. Kessler, T. Wullen, and W. D. Heiss, "Quantitative EEG mapping and PET in Alzheimer's disease," *Journal of the Neurological Sciences,* vol. 110, pp. 46-56, 1992.

[155] C. Huang, L. O. Wahlund, T. Dierks, P. Julin, B. Winblad, and V. Jelic, "Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study," *Clinical Neurophysiology,* vol. 111, pp. 1961-1967, 2000.

[156] M. Grunwald, F. Busse, A. Hensel, F. Kruggel, S. Riedel-Heller, H. Wolf, T. Arendt, and H. J. Gertz, "Correlation between cortical theta activity and hippocampal volumes in health,mild cognitive impairment, and mild dementia," *J Clin Neurophysiol* vol. 18, pp. 178-184 2001.

[157] L. C. Fonseca, G. M. Tedrus, L. R. Prandi, and A. C. Andrade, "Quantitative electroencephalography power and coherence measurements in the diagnosis of mild and moderate Alzheimer's disease," *Arq Neuropsiquiatr,* vol. 69, pp. 297-303, 2011.

[158] C. J. Stam, B. F. Jones, I. Manshanden, A. M. van Cappellen van Walsum, T. Montez, J. P. A. Verbunt, J. C. de Munck, B. W. van Dijk, H. W. Berendse, and P. Scheltens, "Magnetoencephalographic evaluation of resting-state functional connectivity in Alzheimer's disease," *NeuroImage,* vol. 32, pp. 1335-1344, 2006.

[159] M. J. Hogan, G. R. J. Swanwick, J. Kaiser, M. Rowan, and B. Lawlor, "Memory-related EEG power and coherence reductions in mild Alzheimer's disease," *International Journal of Psychophysiology,* vol. 49, pp. 147-163, 2003.

[160] T. Dierks, R. Ihl, L. Frölich, and K. Maurer, "Dementia of the alzheimer type: Effects on the spontaneous EEG described by dipole sources," *Psychiatry Research: Neuroimaging,* vol. 50, pp. 151-162, 1993.

[161] M. Mesulam, "The cholinergic lesion of Alzheimer's disease: pivotal factor or side show? ," *Learn. Mem.,* vol. 11, pp. 43-49, 2004.

[162] C. Babiloni, A. Brancucci, F. Vecchio, L. Arendt-Nielsen, A. C. N. Chen, and P. M. Rossini, "Anticipation of somatosensory and motor events increases centro-parietal functional coupling: An EEG coherence study," *Clinical Neurophysiology,* vol. 117, pp. 1000-1008, 2006.

[163] C.-F. V. Latchoumane, F. Vialatte, A. Cichocki, and J. Jeong, "Multiway Analysis of Alzheimer's Disease: Classification based on Space-frequency characteristics of EEG time series," presented at the World Congress on Engineering 2008,, London, United Kingdom, 2008.

[164] F. Nobili, F. Copello, P. Vitali, T. Prastaro, S. Carozzo, G. Perego, and G. Rodriguez, "Timing of Disease Progression by Quantitative EEG in Alzheimer's Patients," *Journal of Clinical Neurophysiology,* vol. 16, p. 566, 1999.

[165] E. Pucci, N. Belardinelli, G. Cacchiò, M. Signorino, and F. Angeleri, "EEG power spectrum differences in early and late onset forms of Alzheimer's disease," *Clinical Neurophysiology,* vol. 110, pp. 621-631, 1999.

[166] J. Dauwels, F. Vialatte, C. Latchoumane, J. Jeong, and A. Cichocki, "EEG synchrony analysis for early diagnosis of Alzheimer's disease: a study with several synchrony measures and EEG data sets," *IEEE Engineering in Medicine and Biology Society,* vol. 2009, pp. 2224-2227, 2009.

[167] F. Abdollahi and A. Motie-Nasrabadi, "Combination of frequency bands in EEG for feature reduction in mental task classification," *Conf Proc IEEE Eng Med Biol Soc,* vol. 1, pp. 1146-9, 2006.

[168] C. Besthorn, R. Zerfass, C. Geiger-Kabisch, H. Sattel, S. Daniel, U. Schreiter-Gasser, and H. Förstl, "Discrimination of Alzheimer's disease and normal aging by EEG data," *Electroencephalography and Clinical Neurophysiology,* vol. 103, pp. 241-248, 1997.

[169] P. J. Uhlhaas, F. Roux, W. Singer, C. Haenschel, R. Sireteanu, and E. Rodriguez, "The development of neural synchrony reflects late maturation and restructuring of functional networks in humans," *Proceedings of the National Academy of Sciences,* vol. 106, pp. 9866-9871, June 16, 2009 2009.

[170] M. Le Van Quyen, J. Foucher, J.-P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, and F. J. Varela, "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *Journal of Neuroscience Methods,* vol. 111, pp. 83-98, 2001.

[171] P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, and H. J. Freund, "Detection of n:m Phase Locking from Noisy Data: Application to Magnetoencephalography," *Physical Review Letters,* vol. 81, pp. 3291-3294, 1998.

[172] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," *Expert Systems with Applications,* vol. 36, pp. 1329-1336, 2009.

[173] A. Özerdem, B. Güntekin, E. Saatçi, Z. Tunca, and E. Başar, "Disturbance in long distance gamma coherence in bipolar disorder," *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* vol. 34, pp. 861-865, 2010.

[174] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "“Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research,* vol. 12, pp. 189-198, 1975.

[175] H. S. Levin, "A compendium of neuropsychological tests: Administration, norms, and commentary," *Archives of Neurology,* vol. 50, pp. 451-451, 1993.

[176] A. L. Benton, *The revised visual retention test : clinical and experimental applications*. New York :: The Psychological Corporation, 1963.

[177] G. A. Talland and M. Ekdahl, "Psychological studies of Korsakoff's psychosis: IV. The rate and mode of forgetting narrative material," *J Nerv Ment Dis,* vol. 129, pp. 391-404, 1959.

[178] M. Nuwer, "IFCN standards for digital recording of clinical EEG," *Electroencephalography and Clinical Neurophysiology,* vol. 106, pp. 259-261, 1998.

[179] S. Sanei and J. A. Chambers, *EEG Signal Processing* Wiley-Blackwell, 2007.

[180] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing (2nd ed.)*: Prentice-Hall, Inc., 1999.

[181] U. Rajendra Acharya, S. Vinitha Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework," *Expert Systems with Applications,* vol. 39, pp. 9072-9078, 2012.

[182] B. Güntekin, E. Saatçi, and G. Yener, "Decrease of evoked delta, theta and alpha coherences in Alzheimer patients during a visual oddball paradigm," *Brain Research,* vol. 1235, pp. 109-116, 2008.

[183] C. H. M. Brunia, "Neural aspects of anticipatory behavior," *Acta Psychologica,* vol. 101, pp. 213-242, 1999.

[184] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology,* vol. 110, pp. 1842-1857, 1999.

[185] E. Arnáiz and O. Almkvist, "Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease," *Acta Neurologica Scandinavica,* vol. 107, pp. 34-41, 2003.

[186] J. Dauwels, F. Vialatte, and A. Cichocki, " Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing?," *Current Alzheimer Research,* vol. 7, pp. 487-505, September 2010.