



University of South Florida

Digital Commons @ University of South Florida

USF Tampa Graduate Theses and Dissertations

USF Graduate Theses and Dissertations

December 2022

Diagnosis of Neurodegenerative Diseases Using Higher Order Statistical Analysis of Electroencephalography Signals

Seyed Alireza Khoshnevis
University of South Florida

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>

Scholar Commons Citation

Khoshnevis, Seyed Alireza, "Diagnosis of Neurodegenerative Diseases Using Higher Order Statistical Analysis of Electroencephalography Signals" (2022). *USF Tampa Graduate Theses and Dissertations*. <https://digitalcommons.usf.edu/etd/9642>

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact scholarcommons@usf.edu.

Diagnosis of Neurodegenerative Diseases Using Higher Order Statistical Analysis of
Electroencephalography Signals

by

Seyed Alireza Khoshnevis

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering
Department of Electrical Engineering
College of Engineering
University of South Florida

Major Professor: Ravi Sankar, Ph.D.
Stephen Saddow, Ph.D.
Ismail Uysal, Ph.D.
Marvin Andujar, Ph.D.
Kandethody Ramachandran, Ph.D.

Date of Approval:
November 12, 2021

Keywords: Parkinson's disease, Event related potential, Brain computer interface, Brain source localization, Higher order statistics, Bispectrum, Bicepstrum, Quadratic phase coupling

Copyright © 2021, Seyed Alireza Khoshnevis

Dedication

This dissertation is dedicated to my parents, Seyed Abbas Khoshnevis and Farahnaz Aziziasl and all my family and friends who supported and encouraged me from afar during my studies.

Acknowledgments

I would like to sincerely thank my advisor Prof. Ravi Sankar, for providing constant support and guidance throughout my research. I would also like to thank all the committee members, whose knowledge and constructive feedback allowed me to improve and grow as a researcher.

I am very grateful to Dr. Ehsan Sheybani, with whom I have collaborated on different occasions, for his support in different ways during my studies.

I would also like to thank Dr. Murugappan, Dr. Cavanagh, all the participants and everyone else who was involved in collecting and creating the datasets which allowed me to conduct this study.

Table of Contents

List of Tables	iii
List of Figures	iv
Abstract	v
Chapter 1: Introduction	1
1.1 Background	1
1.2 Applications of Electroencephalography	3
1.2.1 Event Related Potentials	4
1.2.2 Background Electroencephalography Signals	7
1.3 Motivations and Research Objectives.....	9
1.4 Contributions	12
1.5 Dissertation Organization.....	13
Chapter 2: Review of Diagnosis of Neurodegenerative Diseases Using Electroencephalography	15
2.1 Neurodegenerative Diseases.....	15
2.1.1 Parkinson’s Disease.....	15
2.2 Diagnosis Process Using Electroencephalography Signals	19
2.3 State-of-the-Art in Diagnosis of Parkinson’s Disease Using Electroencephalography.....	22
2.4 Summary	24
Chapter 3: Higher Order Statistics	26
3.1 Background	26
3.2 Bispectrum Estimation.....	30
3.2.1 Conventional Bispectrum Estimation.....	30
3.2.2 Parametric Bispectrum Estimation.....	31
3.3 Quadratic Phase Coupling.....	32
3.4 Higher Order Statistics in Electroencephalography Signal Processing	33
3.5 Summary	38
Chapter 4: Higher Order Statistical Analysis Framework for Classification of Parkinson’s Disease.....	39
4.1 Background	39
4.2 Proposed Framework	41
4.3 Data Acquisition	43
4.4 Preprocessing and Artifact Removal	45

4.5 Feature Extraction.....	46
4.5.1 Lower Order Statistical Features.....	47
4.5.2 Higher Order Statistical Features.....	48
4.5.3 Proposed New Higher Order Statistical Features.....	49
4.6 Summary	52
 Chapter 5: Diagnosis of Parkinson’s Disease and Classification of Stages Using the Proposed Approach	53
5.1 Diagnosis of Parkinson’s Disease from Healthy Control	53
5.2 Classification of Early-Stage vs Late-Stage.....	57
5.3 Comparison with State-of-the-Art Methods.....	62
5.4 Discussion	64
5.5 Summary	66
 Chapter 6: Supervised Classification of Parkinson’s Disease Using Shallow and Deep Neural Networks	67
6.1 Background	67
6.1.1 Convolutional Neural Network.....	67
6.1.2 Residual Neural Network	68
6.2 Preprocessing.....	69
6.3 Classification Performance	69
6.4 Comparison with the Proposed Method.....	73
6.5 Summary	73
 Chapter 7: Conclusion and Future Research	75
7.1 Conclusion.....	75
7.2 Future Research	78
7.2.1 Higher Order Statistic in Electroencephalography.....	78
7.2.2 Other Applications of Higher Order Statistics.....	79
 References	80
 Appendix A: Copyright Permissions	89

List of Tables

Table 4.1	Summaries of studies conducted on category 1 arranged based on the artifacts and their sources.	39
Table 4.2	Summaries of studies conducted on category 2 that contain methods using HOS features.	40
Table 4.3	Summaries of studies conducted on category 3 of Higher Order Statistics applications in EEG analysis.	41
Table 4.4	General information on the collected EEG signals in dataset #1.	44
Table 4.5	General information on the collected EEG signals in dataset #2.	45
Table 5.1	The average of the feature extracted from the EEG signals for each class.	54
Table 5.2	The average of the feature extracted from the alpha band for each class.	55
Table 5.3	The average of the feature extracted from the beta band for each class.	56
Table 5.4	The average diagnosis performance of all ensemble classifiers.	57
Table 5.5	The accuracy of diagnosis using bagged trees ensemble for all three sets.	58
Table 5.6	Average of the normalized lower order features for each stage of PD.	59
Table 5.7	The average of the normalized 3rd - 7th order standardized moments for each stage of PD.	60
Table 5.8	Average of the normalized proposed features (Complex H 1-5, Real Bicepstrum H 1-5 and Complex Bicepstrum H 1-5) for each stage of PD.	61
Table 5.9	Performance of the classifiers including the AUC and the overall accuracy.	61
Table 5.10	Summary of previous studies conducted on automated diagnosis of PD and classification of its stages.	63
Table 6.1	The highest overall test accuracy of the 6-layer networks.	69

List of Figures

Figure 1.1 Typical P300 speller matrix.....	5
Figure 2.1 Usual course of symptom development in PD patients from early to late-stage.....	16
Figure 2.2 The effect of PD on SN area and its location in the brain.....	17
Figure 2.3 Flow of dopamine between the presynaptic and postsynaptic neurons for normal vs PD.....	18
Figure 2.4 A block diagram of disease diagnosis using EEG signals collected from patients.....	20
Figure 2.5 The 3D graph of a 15-channel sample EEG signal.....	21
Figure 3.1 The region of non-redundancy.....	27
Figure 3.2 Magnitude and phase of the bispectrum of the sample EEG, estimated by the direct method.....	31
Figure 3.3 Estimated bispectrum using parametric, ARMA model.....	32
Figure 4.1 Proposed hierarchy approach for diagnosis of PD and classification of the stages.....	42
Figure 4.2 The process of PD diagnosis for the all-band vs alpha band vs beta band.....	46
Figure 5.1 The diagram describing classification with ensemble learning.....	57
Figure 6.1 The train and test accuracies of the 6-layer shallow CNN.....	70
Figure 6.2 The train and test loss of the 6-layer shallow CNN.....	70
Figure 6.3 The train and test accuracies of the 6-layer shallow ResNet.....	71
Figure 6.4 The train and test loss of the 6-layer shallow ResNet.....	71
Figure 6.5 The proposed one dimensional 6-layer shallow CNN.....	72

Abstract

The field of signal processing has many applications, one of which is in the field of biomedical engineering where it has improved the performance of biomedical devices and the accuracy of medical diagnosis. One of the areas that have benefited from this field is the diagnosis of Parkinson's disease (PD). This disease is one of the most common neurodegenerative diseases of the central nervous system. Nearly one million Americans suffer from PD, and this number goes up to over ten million people worldwide. The main symptoms of PD include bradykinesia, rest tremor, rigidity, and impaired balance. There are many types of biomedical data that have been used in the diagnosis process of PD, however, most of the applied biomedical signals rely on the presence of motor symptoms which means in most cases, by the time that the patients are diagnosed, they are likely to have lost the majority of the dopaminergic neurons in their brains.

One of the signals that have been used for the diagnosis of PD is electroencephalography (EEG). The neurons in the brain communicate with each other through electrical potentials that appear at the synapses. EEG is a noninvasive method that collects the small voltages that appear on the scalp caused by large clusters of neurons using multiple electrodes; therefore, EEG recordings are multi-channel signals where each channel is corresponding to a specific region of the brain. There are two main types of EEG signals, background EEG, and event-related potential (ERP). Background EEG is the signal collected during the rest state and contains the regular activity of the brain when it is not provoked, whereas ERP is the changes in the background EEG, resulting from a specific stimulus. While background EEG signals are better suited for diagnosis

purposes, they are highly nonlinear, non-stationary and non-Gaussian signals; hence, to extract relevant information from them, advanced methods of signal processing are required. The background EEG is a random signal which indicates that regular features such as time locked features or peaks of the signal do not carry much information. For random signals, statistics is usually the most appropriate method for analysis.

The word statistic is usually used to refer to first and second order statistics. Higher order statistics (HOS) is defined as a more general term that covers higher order of statistical features. the field of HOS analysis is usually employed for highly complex signals, where the first and second order statistics failed to adequately define the system. Due to the highly random nature of background EEG signals, HOS has been employed by many researchers for more detailed analysis. In this study, a range of HOS features have been used to improve the diagnosis performance of PD patients from healthy control (HC) and classification of stages of PD after a positive diagnosis. A detailed analysis of the features was performed to find the best combination for this application and a number of new HOS features were developed to improve the performance of the model. Concurrently, based on previous research a spectral analysis of the data was performed to investigate the effect of PD on brain rhythms where HOS features were extracted from multiple rhythms and used separately in the diagnosis process.

The classification stage was performed by a range of conventional, ensemble and deep learning algorithms while employing the leave-one-trial-out (LOTO), leave-one-subject-out (LOSO) cross validation (CV) methods. To preserve the balance of the data, a new CV approach, leave-two-subjects-out (LTO) was also employed (one from each class). For diagnosis of PD, a comparison between the features extracted from different brain rhythms and different classifiers was performed. The result was then compared to several deep learning methods and other state-

of-the-art approaches. The performance of the PD stage classification was also compared to other studies in this field. Together these two methods create a unified hierarchy model to diagnose and identify the stages of PD.

Chapter 1: Introduction

1.1 Background

Human bodies contain vast quantities of information about our health. By making use of different sensors, it is possible to capture parts of this information through raw data. Biomedical sensors are able to collect different types of data from various regions of the body such as in electrocardiography (ECG), electroencephalography (EEG), magnetoencephalography (MEG), etc. The quantity of the captured data along with the complexity of the signals causes physicians to rely on a very small portion of information based on very few readings. Biomedical signal processing involves computational methods that aim to analyze large quantities of data and extract useful information from them which can later be used by physicians. One of the main methods for recording data from the human body is EEG.

EEG is a noninvasive method for monitoring the changes in the electrical potential of the brain from the scalp. The neurons in the brain communicate through two types of synapses, chemical and electrical. In communication through chemical synapses, charged chemical substances known as neurotransmitters flow from the transmitter neuron to the receiver neuron. This flow occurs due to the changes in the electrical potential between the sender and receiver. There are over 100 different neurotransmitters, however most of the work is done by seven neurotransmitters, dopamine, gamma-aminobutyric acid, glutamate, histamine, acetylcholine, norepinephrine, and serotonin. EEG uses several electrodes that are placed on the scalp to record synaptic potential between the neurons inside the brain. Each of the electrodes, records the signals from a different location on the scalp and together they form a multi-channel signal which is the

EEG recording. For a normal person the amplitude of this signal is usually around 10-100 μ volts and the frequency is between 1-100 Hz. Therefore, according to the Nyquist theorem, the frequency used to sample EEG recordings should at least be 200 Hz. There are two types of EEG signals based on the reference point for the electrical potential of the electrodes. In monopolar EEG, an electrode is placed on the earlobe which acts as the reference, while in bipolar EEG a certain channel is selected and acts as the reference for other channels [1]. EEG signals are usually contaminated with large quantities of noise and have a relatively low signal to noise ratio (SNR), which makes them unusable without several prior preprocessing steps. They are random, highly nonlinear, non-Gaussian and non-stationary which makes extracting useful information from them, extremely difficult; therefore, usually for analyzing them, advanced signal processing methods are employed [2].

Waveforms that are associated with certain neural activities and are caused by certain behaviors are known as brain rhythms and are usually recognized by their amplitude, shape, frequency, and location among other characteristics. The six main brainwaves or “neural oscillations” are alpha (8-12 Hz), beta (12-30), gamma (30-100 Hz), delta (0.5-4 Hz), theta (4-8 Hz) and sigma (12-16 Hz). Each brainwave is associated with certain activities for instance alpha rhythm is usually associated with the state of wakeful relaxation usually with closed eyes, while beta rhythm is associated with busy or active concentration and is seen to be related to the motor cortex area of the brain [3, 4].

The EEG signals that are collected during rest state (usually with eyes closed) and show the natural state of the brain are known as background EEG signals. In the presence of an external stimuli, certain changes are added to the background signal; these changes can be extracted as a separate signal, known as event related potential (ERP) signal. The ERP signals are the reaction

of the brain to a certain stimulus and are temporally in sync with and related to the event. When stimulus is manually (usually in a lab) performed, the collected ERP is known as evoked potential (EP). ERP signals are used in many fields including the development of brain computer interface (BCI) systems while background EEG signals are mostly used for medical purposes such as diagnosis of neurological disorders [5-7].

1.2 Applications of Electroencephalography

The human brain does not act as a fully deterministic system since it is able to create new things. It is also not a complete stochastic system because of its ability to learn and to repeat exact sequences of thoughts and actions. The answer to this paradox is the nonlinear nature and deterministic chaos of the human brain [8]. This is one of the reasons that analyzing brain signals have proven to be extremely difficult. Researchers have used EEG analysis in many different areas. In medicine, it has a variety of applications from detecting brain injuries and illnesses, level of pain, depth of anesthesia measurements to amnesia and memory loss. However, medicine is not the only field that EEG is useful for and another area that has benefited is the field of Robotics. The EEG based brain controlled mobile robots can be used to improve the living conditions of many disabled people [9]. Researchers in academia and companies have already used these methods to build EEG controlled wheelchairs and prosthetics. There are also headbands that collect EEG data, translate them to simple commands and let people move predetermined objects with their mind similar to the way it is done in science fiction movies; even some companies in the gaming industry are studying EEG signals in order to make brain controlled virtual reality (VR) games in the future. Due to the variety of applications, EEG signals have been widely investigated by many researchers over the years.

1.2.1 Event Related Potentials

ERPs are the foundation of BCI systems as the method of evoking potentials determines the paradigm of the resulting BCI, which is the main application of ERP signals. These action potentials are minuscule voltages resulting from thought or stimulation of the human brain. These are the features used in BCI systems categorized by time of occurrence post stimulus, amplitude, or being otherwise transformed for extraction. Current methods seek to solve the issue of low SNR through filtering of noise and transformation methods.

The original device used to create a BCI system was a simple electrode made of conductive pins attached to a plate inserted directly into the gray brain matter. The invasive nature of this method known as electrocorticography (ECoG) or intracranial electroencephalography (iEEG) meant that the method was a last resort for patients with “locked-in” syndrome. A more recent development in invasive BCI technology is the combination of stents and electrodes coined stentodes allowing for minimally invasive surgery placing a stent in the brain to collect signals [10]. Due to ethics concerns, using ERP from EEG signals is a much better option. As a result, BCI systems developed with this noninvasive method achieved similar accuracies despite the much lower SNRs

ERPs are commonly induced for BCI systems via external visual stimuli or internal motor imagination stimuli. Since control of a computer is the main application for BCI, visual stimuli are the simplest for eliciting ERP. The first ever noninvasive BCI using ERP was called the P300 speller and published by Farwell and Donchin [11]. Later on, Li et al. [12-14] improved the newly developed BCI systems by developing a new single trial P300 extraction method. This method uses a matrix of flashing symbols to evoke an ERP that appears at 300 milliseconds after visual cue. An example of this speller setup can be seen in Figure 1.1. The other type of visual stimuli

BCI is known as steady state visually evoked potentials (SSVEP). The method uses light-emitting diodes (LED) flashing at a constant low frequency which produces a corresponding result in the visual cortex. Based on the length and frequency of the signal, the desired command can be initiated [15].



Figure 1.1 Typical P300 speller matrix.

Internal stimuli-based ERPs are the source of two categories of BCI known as motor imagery (MI) and reflexive semantic conditioning (RSC). Motor imagery BCI correlates imagined limb movements with commands by reading signals from the motor cortex. Due to the varying nature of brain signals, MI methods always require a training period where the user is cued to imagine body movements and the resulting signal can be captured for classifier training. RSC BCI use features that result from the meaning of the stimulus. This method requires a training period as well, after which the user can simply think of the same meaning to trigger the desired command. These BCI systems do not have a set efficiency due to their multi-modular nature in the stages of preprocessing, feature extraction, and classification. This variability is readily apparent in papers like Doud et al. [16] and Taheri et al. [17]. When researchers focus on the task, there may not be enough time to achieve high accuracies as in the former paper an accuracy of 70% was achieved while the latter focusing solely on novel classification methods achieved an average of 96%. The number of mathematical anomalies in brain signals make current methods out to be crude

constructs in a field that is nearly half a century old [6]. As such, until commercially marketable systems emerge to set industry standards, research in the field remains somewhat scattered due to its magnitude.

ERP extraction methods from EEG signals have evolved greatly over time. Older methods such as the Fourier transform and canonical correlation analysis (CCA) transform continuous data sets into discrete chunks to be processed for features like magnitude and phase in Fourier transform or matrix coefficients in CCA. Newer advanced methods have been created catering directly to the needs of BCI systems rather than just grouping data. These methods coined brain source localization (BSL) seek to either find the location of signal sources in the brain or predict the EEG signals dependent on the location of the signal sources. The methods known as the inverse and forward problems respectively have found great success being used widely in the global BCI competition organized by Benjamin Blankertz [18, 19]. The most popular method here known as common spatial patterns (CSP) merges traditional blind source separation (BSS) methods with BSL to isolate signals while simultaneously assigning their most probable point of origin. The method has won many of the subsets in the competition since its first appearance in BCI Competition 2003.

Gauging the efficacy of feature extraction methods as it is only one component of a complex system. The classification stage originally started with a simple threshold comparison of features. Later methods linearize datasets like Fischer's linear discriminant analysis (LDA) and subsequent derivatives. As the field has grown, machine learning methods have shown more potential with some of the more common methods including support vector machine (SVM), K nearest neighbor (KNN), and forest ensemble methods. A few problems remain in machine learning including that training datasets need to be sufficiently large and increasing the number of

classes for a given BCI system lowers the accuracy. This limits the number of commands BCI systems have available as each command is considered a class.

Despite the variety of parameters to measure each stage of the process, the unknowable nature of the target signal in a BCI system means that swapping methods in individual stages is the only way to judge effectiveness. As a result, the most common parameters for judging BCIs are false positive rates, classification accuracy, and information transfer rates. This does not mean options such as the number of classes or speed of decision making should not be considered but rather that most papers will not include them if they are irrelevant.

1.2.2 Background Electroencephalography Signals

As mentioned, background EEG signals show the natural state of the brain. Although there are many applications for which background EEG signals have been used, the majority of them have been for biomedical purposes.

Epilepsy is one of the common neurological disorders that affects over 50 million people [20] and it is usually marked by the occurrence of seizures. These disorders are caused by sudden electrical discharges in the cerebral cortex which disturbs the normal functions of the brain. Researchers used to believe that epileptic seizures occur suddenly, only moments before the clinical attacks. However, studies of EEG signals of patients showed changes in brain activity up to several hours before the attacks. Over the past few years, researchers have focused on using EEG signals for predicting and analyzing epileptic seizures. Kiyimik et al. [21] compared the results of short-time Fourier transform (STFT) and continuous wavelet transform (CWT) to evaluate seizure activity in epileptic patients. The STFT method produced better results in processing real-time signals because of its low computational complexity and could be useful for real-time diagnosis, but STFT has limited frequency resolution. Contrarily, wavelet methods had

better performance in the multi-resolution analysis for noise elimination and could be more beneficial for clinical interpretation due to their high resolution [1]. However, it must be kept in mind that spectral analysis, in general, is more susceptible to noise and artifacts compared to temporal statistical measures. Haider et al. [22] have proved that a panel of multiple quantitative Electroencephalogram (qEEG) can be used to identify seizures in critically ill adults with reasonable sensitivity, significantly reducing the review time compared to standard EEG signal interpretation. However, it is still necessary to confirm the suspected seizures with intermittent EEG signal assessment. Fu et al. [23] proposed a new spectrum (time-frequency) analysis algorithm for identifying seizures from EEG signals. This method uses the Hilbert marginal spectrum (HMS) analysis, which is based on Hilbert-Huang transform (HHT). HHT uses the empirical mode decomposition (EMD), which decomposes the signal into basic components called intrinsic mode functions (IMFs) instead of sinusoids and wavelets. EMD decomposes the signal in the time domain and is mostly beneficial in cases of nonlinear and nonstationary signals. Ensemble empirical mode decomposition (EEMD) is the enhanced version of EMD; in this method some of the problems of the traditional EMD such as mode mixing problem are fixed. Swiderski et al. [24] used the Lyapunov exponents (LE) of EEG signal for the identification of epileptic seizures. The result showed a noticeable difference between the largest Lyapunov exponents (LLE) of normal and epileptic EEG signals.

One of the most important factors in a successful surgery is the general anesthesia which is administrated by a specialist. One of the applications of EEG signals is for measuring the depth of anesthesia for patients during surgery. Many researchers have conducted studies on this subject such as Zhang et al. [25] and Liang et al. [26] who have used the entropy of background EEG signals for determining the depth of anesthesia.

EEG signals have also been used for emotion recognition. In their studies, Li et al. [27] focused on exploring linear and nonlinear EEG features that could be used for emotion recognition. Other studies such as the research conducted by Song et al. [28] focused more on application of deep learning for classification of different emotions from background EEG signals. Natarajan et al. [34] evaluated some nonlinear parameters such as Hurst exponent (HE), correlation dimension (CD), LLE and Approximate entropy (ApEn) of EEG signals for different mental states. Afterward, they applied the analysis of variance (ANOVA) test on the result and achieved excellent 'p' values in all cases. They found that the alpha wave becomes stronger under the influence of music and reflexology; where at the same time they noticed a decrease in CD, LLE, and H_{1-5} features.

Another application of EEG signal is for patients who suffer from neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington disease (HD), and such. These signals can be used in the diagnosis process, for measuring the progression of the disease or monitoring symptoms and other aspects of the patients' lives such as their sleep, emotions, etc. The purpose of this research was to develop new methods of EEG analysis that would improve the diagnosis and stage classification of neurodegenerative disease patients. Aside from the applications mentioned in this section, EEG signals have many other applications in the medical field, health monitoring, recreational, and military.

1.3 Motivations and Research Objectives

The primary objective of this project is to apply the higher order statistics (HOS) analysis on EEG signals to diagnose and identify the stages of neurodegenerative diseases. To achieve this objective, the focus of this research was on PD, however approaches based on other biomedical signals, EEG analysis of different neurodegenerative diseases can be very similar since from the

neurological point of view they are very similar and in fact are mistaken for one another constantly. The contribution of this study is to create a hierarchy model based on newly developed HOS features of EEG signals to offer an objective aid in the diagnosis process of neurodegenerative diseases with the focus on PD. In this hierarchy model first PD vs healthy control (HC) participants are classified and once the positive results were achieved it moves on to the second step, to identify the stages of PD by classifying the PD patients as early or late stages. Analysis of brain signals is the only approach that does not rely on symptoms since it monitors the source of the disease and hence it has the potential of diagnosing the disease at early stages.

Over the past decades, there have been many studies conducted on using gait [29, 30], tremor [31, 32], speech [33, 34], etc. signals to diagnose PD patients. The main issue of some of these approaches is that they are dependent on the manifestation of symptoms. This can be a major problem since not everyone develops the same symptoms (and at the same rate) and also because by the time that the motor symptoms occur, most of the dopamine-producing (dopaminergic) neurons are already dead [35].

Recently, there has been an increasing number of studies that have focused on the diagnosis of PD using EEG signals. Using biomedical signals that are collected from the brain allows researchers to inspect the source of the disease and creates possibilities where the diagnosis occurs before the development of the symptoms. However, since EEG recordings are random multi-channel signals, they require advanced signal processing methods in order to use them for classification purposes. One method for analyzing EEG recordings is through HOS analysis. HOS is the generalization of second order statistics for higher orders and includes higher order features that are defined in temporal, spectral, and cepstral domains among others.

As it was previously mentioned, EEG signals are highly non-Gaussian random signals, meaning the probability distribution function (PDF) of this random signal does not resemble a Gaussian process. The similarity of a distribution with the Gaussian process can be quantified and the resulting parameter is known as the measure of non-Gaussianity. The first measure of non-Gaussianity is kurtosis which is the 4th order standardized moment. In general, HOS analysis is one of the best approaches for analyzing non-Gaussian signals. At the same time, during our preliminary studies, it was noticed that the non-Gaussianity of the EEG signals for PD patients is higher compared to HC. The death of neurons caused by neurodegenerative diseases reduces the connectivity of the brain, which in turn increases the non-Gaussianity of the EEG signals. Therefore, HOS analysis of EEG signals was chosen as the main approach in this study.

The first impact of this project is on PD patients. As it was mentioned 10 million people suffer from PD worldwide and the number of people affected by the disease will undoubtedly increase in the future. The diagnosis of different stages of PD can help physicians create better and more effective treatment plans for their patients; it might even affect the process of finding a cure. The second impact of the project will be felt by other neurodegenerative disease patients; all neurodegenerative diseases cause loss of neurons and usually the only difference between them is the type of neurons that they affect. Therefore, this project has the ability to be expanded to other neurodegenerative diseases. The new features that were discovered through the course of this research (not to mention our other findings) will undoubtedly be useful in diagnosis of other similar diseases (perhaps indirectly). Finally, the HOS features that were discovered and used for classification of PD are still mathematical features similar to others such as entropy. There are many other random signals that contain a high amount of noise similar to EEG; therefore, it is a

high possibility that the discovered features have applications in other areas of biomedical signal processing and statistical analysis in general.

1.4 Contributions

A new automated model was developed based higher order features of the alpha frequency band of EEG signals that uses ensemble learning for classification of neurodegenerative diseases including PD patients.

- A new hierarchy model was developed for diagnosis and identification of the stages of neurodegenerative diseases. To test this model two PD datasets were employed.
- Fifteen new HOS features known as CH₁-CH₅, RBH₁-RBH₅ and CBH₁-CBH₅ were developed specifically for classification of EEG signals of neurodegenerative disease patients.
- Compared the efficacy of the developed features and selected the most suitable ones for this application.
- An exhaustive study on other features used for classification of PD including, HOS and nonlinear features was conducted and the best combination was selected and implemented. Created a feature pool with the selected features from the state-of-the-art studies and the developed HOS features.
- Employed and compared multiple classification algorithms to find the most suitable for this application.
- Employed conventional classification methods and ensemble learning approaches for classification of the created feature pool and compared their performances.

- Implemented multiple deep and shallow neural network architecture including convolutional neural networks (CNN) and residual neural networks (ResNet) and compared their results with the proposed approach.
- Decomposed the EEG signals into frequency bands and calculated the selected features for each band along with the whole signal. Compared the classification performance of HOS features for alpha and beta bands of EEG signals.
- Compared the performance of the proposed ideas including the model, HOS features, combined feature pool, frequency band and classification approach with state-of-the-art methods.

1.5 Dissertation Organization

Chapter 2 is dedicated to basic information about neurodegenerative diseases where it starts with a brief background followed by information about PD such as prevalence, symptoms and diagnosis. After the subsection on PD, the diagnosis process using EEG signals is explained followed by a brief discussion on the state-of-the-art methods related to classification of neurodegenerative diseases using EEG signals.

The focus of Chapter 3 is on the mathematical aspect of HOS, where first some background including algorithms and features are discussed. The conventional and parametric estimation methods for bispectrum are presented followed by a brief explanation of quadratic phase coupling (QPC). In the last subsection, the applications of HOS analysis in EEG signal processing are discussed.

In Chapter 4, HOS analysis of EEG signals for diagnosis of PD is discussed, where after a brief background, the proposed framework, the datasets used in this study followed by the preprocessing steps and the proposed features are discussed.

In Chapter 5, the extracted features along with the performance of the proposed approach is presented and compared to current state-of-the-art methods.

In Chapter 6, the supervised classification of EEG signals of PD vs HC using both shallow and deep CNN and ResNet are included and their performances are compared with each other and with the HOS based approach.

In Chapter 7, a summary of this research is presented. Then possible improvements and future directions for expanding this research by using the developed features for other applications are described.

Chapter 2: Review of Diagnosis of Neurodegenerative Diseases Using Electroencephalography¹

2.1 Neurodegenerative Diseases

Any structural, biochemical, or electrical oddities that occur in the nervous system including the brain, spinal cord, and nerves (peripheral nervous system) is considered to be a neurological disorder. Currently, there are more than 600 neurological disorders that affect around 50 million Americans annually [36]. Neurodegenerative diseases are one of the most important subcategories of neurological disorders. This type of disease occurs when over time, (certain) neurons of the brain (or nerve cells of the peripheral nervous system) become dysfunctional and die. The death of neurons disrupts the regular activities of the brain and can severely affect patient's day to day lives or in some cases be fatal. Among different neurodegenerative diseases, one of the most common ones is PD. Some of the other common neurodegenerative diseases are AD, HD, amyotrophic lateral sclerosis (ALS), motor neuron disease (MND), prion disease, spinocerebellar ataxia (SCA), and spinal muscular atrophy (SMA).

2.1.1 Parkinson's Disease

In 2010 the number of PD patients in the US was estimated by the Parkinson's Foundation Prevalence Project to be around 680,000 patients. In 2020, the estimated number went up by over 36% to 930,000. This number is estimated to increase by over 33% to 1,238,000 people by 2030. According to this study, the number of PD cases increase by age and it is more common in men

¹ This chapter was published in [6]. Permission is included in Appendix A.

compared to women [37]. Worldwide, the estimated number of people afflicted by PD exceeds 10 million people [38].

The main symptoms of PD are rest tremor, bradykinesia, muscle rigidity, speech alteration, and postural instability. Resting Tremor is the repeating, oscillating, involuntary muscle movement that can cause shaking and tremor of a body part (usually the limbs) and it occurs in almost 75% of PD patients [39]. Bradykinesia is the slowing of voluntary movements and is caused by the decline of the brain’s ability to articulate and execute instructions [38].

Muscle rigidity is caused by the increased resistance in the voluntary movements of the limbs which goes beyond the old age or arthritis. Speech alteration is another one of the PD symptoms which affects around 50-90% of PD patients, and can occur in different forms including having low voice or monotone pitch among others [40]. Loss of balance usually occurs during later stages of PD when the patients will experience difficulties with keeping their balance while standing or walking. Aside from the main ones, there are many other symptoms that are associated with PD such as small handwriting, insomnia, constipation, masked face, stooping, dizziness among others. Figure 2.1 shows the progression of some of the major symptoms of PD from early to late-stage.

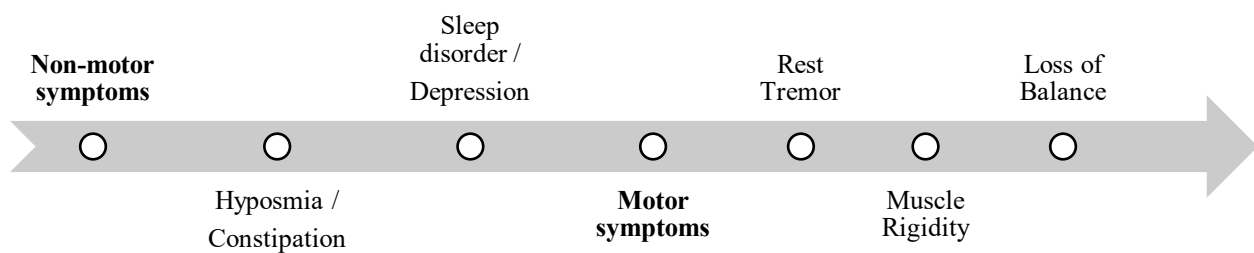


Figure 2.1 Usual course of symptom development in PD patients from early to late-stage.

The main cause for the occurrence of PD is not yet discovered, and although the effects of genetics in PD is widely accepted, the genome-wide screens have not been able to find any specific

genetic abnormality; aside from genetics, multiple studies have suggested that the development of PD is also affected by environmental factors [41]. Although the underlying causes of PD are not yet discovered, the symptoms of PD are caused by high deficit of dopamine and the presence of Lewy bodies in mesencephalon.

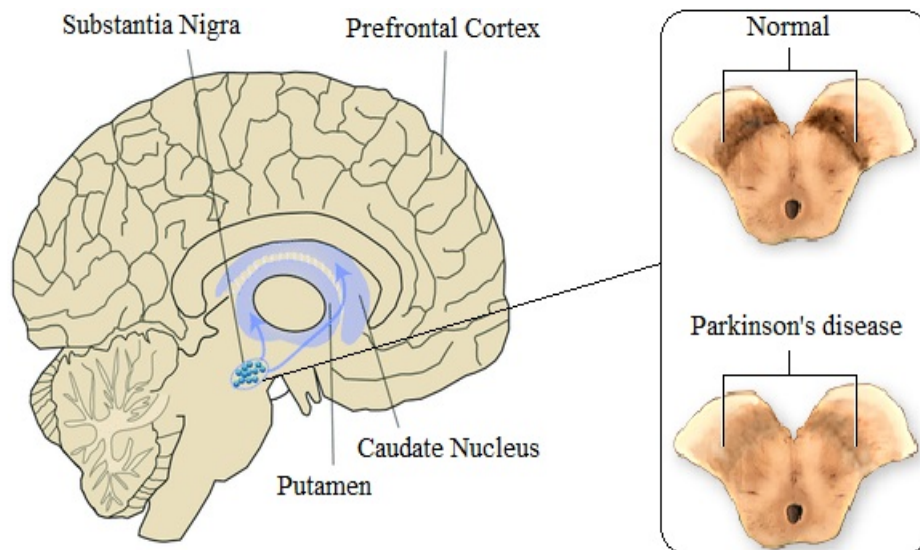


Figure 2.2 The effect of PD on SN area and its location in the brain.

The dopamine deficit occurs when large number dopaminergic neurons die. This type of neurons is mostly concentrated in the substantia nigra (SN) area where they can be seen as black spots. Figure 2.2 shows the location of SN in the brain along with the comparison of SN for normal people vs PD patients. The dark areas that represent the dopaminergic neuron are much smaller for PD patients (depending on the severity of the disease). It can also be seen that the SN area is located deep within the midbrain and cannot be physically accessed, making the diagnosis process even more complicated [42]. In fact, this process is so difficult that in some cases the definitive diagnosis occurs after the death of the patient, when there is physical access to the SN area during the autopsy.

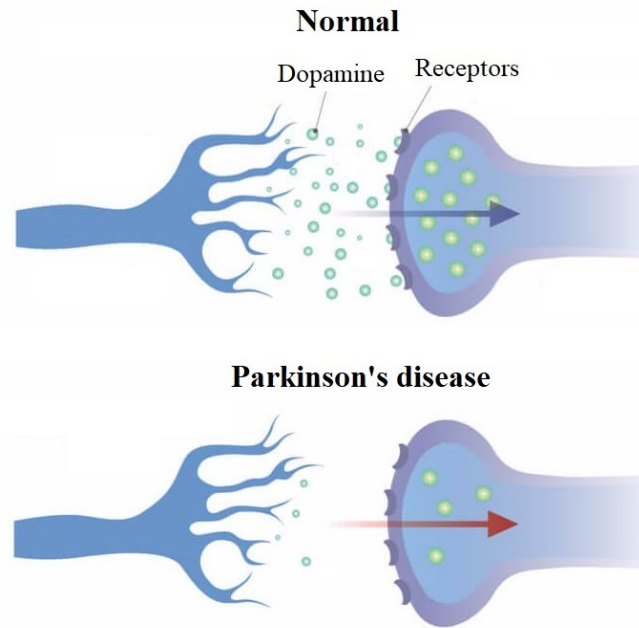


Figure 2.3 Flow of dopamine between the presynaptic and postsynaptic neurons for normal vs PD.

The decrease in the flow of dopamine causes the brain signals to not be delivered successfully. Figure 2.3 shows the flow of dopamine (neurotransmitter) between the presynaptic and postsynaptic neurons of a normal person vs a PD patient. The diagnosis of PD at the early stages plays a significant role in developing the treatment plan and ensuring the quality of life for the patients. Therefore, developing a novel method for diagnosis of the PD in the early stages cannot be overstated. As of now, there are no specific test that can give a definite diagnosis of PD, therefore neurologist usually keep the patients under observation and consider the results of multiple test such as dopamine transporter scan (DaTscan) before making a decision on the diagnosis. After a positive initial diagnosis, doctors use subjective clinical evaluation methods such as Parkinson's disease rating scale (UPDRS) [43] to measure the progression of the disease. Another such method is Hoehn and Yahr (H&Y) [44] that measures the severity of the disease and categorizes the patients into five different stages based on their symptoms.

As it was mentioned, researchers have used different biomedical signals such as speech, gait, tremor, etc. for classification of neurodegenerative diseases such as PD to provide an objective approach to the diagnosis process. Over the past few years, EEG signals have been extensively used for diagnosis purposes. However, it must be noted that the final diagnosis will still be performed by the neurologists and these methods, similar to other tests, aid doctors in the diagnosis process. In the following section, the diagnosis process using EEG signals, including the required signal processing steps are discussed.

2.2 Diagnosis Process Using Electroencephalography Signals

To perform diagnosis through EEG signals, there are several steps that must be followed, Figure 2.4 shows the steps that are needed for this process. This graph shows the general process of classification of EEG signals for diagnosis purposes. There are 7 steps included in this graph, however this number can be different based on the application or the type of the study. Some researchers add feature ranking as a separate step while some keep segmentation as a part of preprocessing.

There are several sources of artifacts that affect EEG signals and must be addressed. As it can be seen in Figure 2.4, in the second step, after data collection is artifact removal. The most important artifacts to be removed are eye movement and blinking artifact or electrooculography (EOG), muscular movement or electromyogram (EMG) and power line interference. There are many approaches for removing artifacts, some of the most important of which are ICA, wavelet-ICA, wavelet decomposition, CCA, EMD, principal component analysis (PCA), CSP and adaptive filtering, among others [2]. Some of these methods are BSS methods that can also be used for artifact removal by separating the sources where artifacts themselves are considered a separate source.

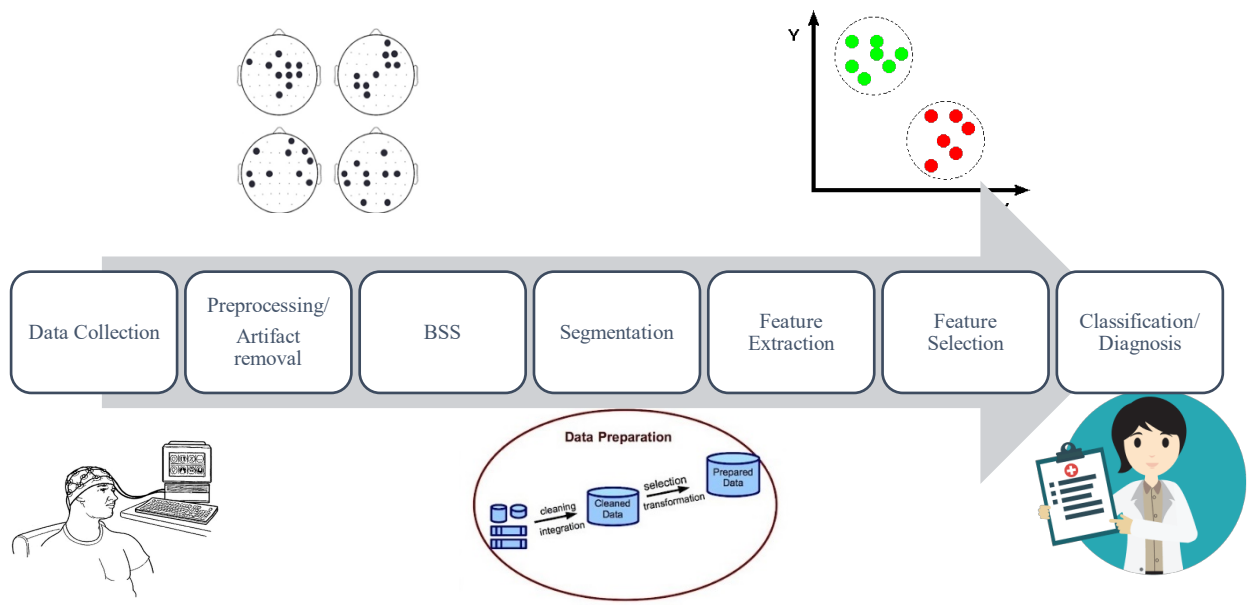


Figure 2.4 A block diagram of disease diagnosis using EEG signals collected from patients.

At each point in time, there are several main sources of electrical potential inside the brain which are observed through the electrodes that are placed on the scalp. The signal collected through each electrode is the observation signal from that location on the scalp and is a combination of all the sources for that point in time. For a certain electrode, the sources that are closer usually have a stronger influence while the sources that are further away affect it less. Due to this phenomenon, there is usually some correlation between different EEG channels. There are several approaches to reducing the correlation between the channels, such as BSS, brain source separation and BSL. Currently, one of the most popular methods for reducing the correlation between the channels is independent component analysis (ICA) which estimates the sources by assuming minimum Gaussianity for the sources. ICA uses the fourth order moment, kurtosis as the measure of non-Gaussianity. Figure 2.5 shows a 3D graph of a 15-channel EEG signal for a duration of 20 seconds. As it can be seen in this figure, the 15 channels are highly correlated; therefore, methods such as ICA are needed before this data is usable. This step of the process is marked as BSS in Figure 2.4;

however, many researchers consider both artifact removal and BSS as one step known as preprocessing.

Different methods of analysis that are usually used for other types of signals may not be suitable for non-stationary signals. Therefore, after artifact removal and BSS steps, the signal must be segmented to create the datapoints due to the non-stationary nature of EEG signals. Based on the frequency range of EEG signals and the low sampling frequencies, the duration of 2 s has been accepted as the benchmark segment size. The usual EEG devices have sampling frequencies of 128, 250 and 500 Hz which results in segments of 256, 500 and 1000 samples, respectively. Whereas some have used fixed-size segments based on the average stationarity of EEG signals, others have used more advanced methods such as spectral error measurement (SEM), generalized likelihood ratio (GLR), and nonlinear energy operator (NLEO) which automatically find the size of the semi-stationary segments [45].

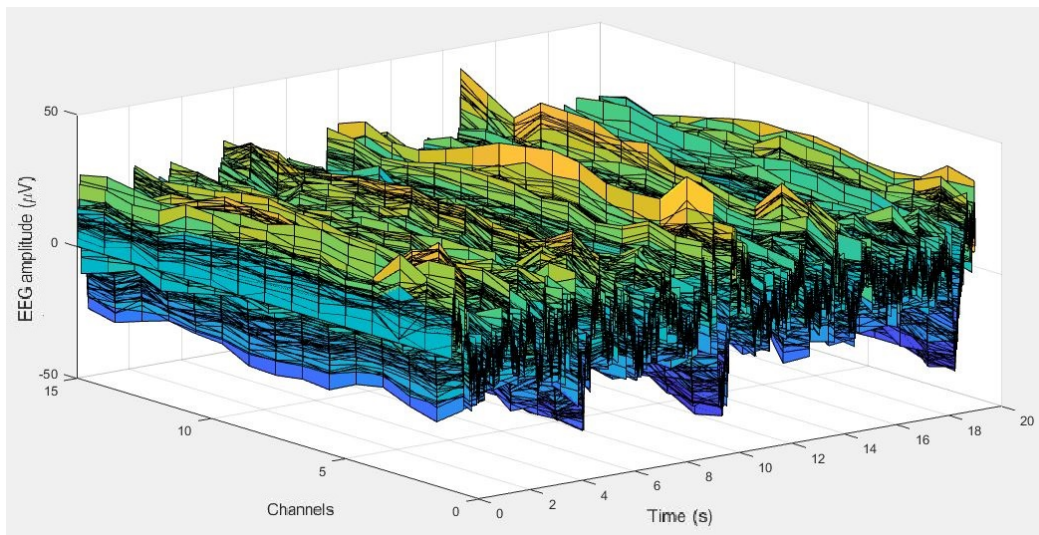


Figure 2.5 The 3D graph of a 15-channel sample EEG signal.

After segmenting the data, appropriate features must be extracted from each segment of the signal. Due to the random nature of EEG signals, statistical features have been shown to have

better performance compared to others. There are several ways to categorize the features, first by the linearity where the features are categorized as linear and nonlinear. Next by the domain they are calculated from such as temporal, spectral, wavelet, cepstral, etc. Another method to categorize the features is based on their order where there are second order statistical features, third order statistical features and so on. Usually features of the second order and lower are known as lower order features while third order and higher are known as higher order features. After extracting the features, some researchers use feature selection methods including feature ranking, manual selection and exhaustive search among others to reduce the number of features which in turn will reduce the dimension of the feature space and mitigate the curse of dimensionality.

When the features are extracted and the feature pool is created, machine learning methods can be used for classification of datapoints. In the classification phase, while the random cross validation (CV) is a popular method, for diagnosis applications it is more appropriate to use methods such as leave-one-trial-out CV (LOTO-CV), leave-one-subject-out CV (LOSO-CV) or similar approaches to avoid having datapoints belonging to one subject appearing in both the training and validation steps. This approach allows for a more balanced validation set while at the same time the patients are tested against multiple HC participants. Traditional classifiers such as SVM, KNN, decision tree (DT), naive Bayes (NB), LDA, etc. have been historically used for classification of the extracted features. Recently, deep learning methods have also been used for classification of EEG signals. In the next section, some of the current studies and state-of-the-art methods in diagnosis of PD are discussed.

2.3 State-of-the-Art in Diagnosis of Parkinson's Disease Using Electroencephalography

Researchers have been studying the effects of PD on the brain for several decades. Many of these studies have used EEG recordings to monitor the changes brought by PD. In 2011, Schlede

et al. found a significant correlation between the grand total EEG (GTE) score and the deteriorating of cognitive ability in late-stage PD patients [46]. In the same year, Swann used EEG recordings and ERP signals to calculate the stop signal reaction time (SSRT) of PD patients versus HC [47]. At the same time, Klassen et al. used Quantitative EEG recordings as a predictive biomarker for the progress of dementia in late-stage PD patients [48]. In 2014, Yuvaraj et al. used power and frequency of EEG recordings during the state of emotion to differentiate between PD patients and HC. They used 14-channel EEG recordings of 20 PD patients and 30 HC [49].

In 2017, Liu et al. [50] used the discrete wavelet transform (DWT) with sample entropy (SampEn) as their feature and tried to diagnose PD patients from EEG signals by using a three-way decision model. In their study, they used 10-channel EEG from 42 PD patients and 42 HC. In recent years, there have been some focus on the applications of deep learning in EEG analysis [51]. In 2018, Oh et al. [52] used a deep learning algorithm to classify 20 PD patients and 20 HC. They created a 13-layer CNN for classification. Their algorithm achieved an accuracy of 88.25 % and a specificity of 91.77 %. In 2014, Obukhov et al. calculated the hemispheric asymmetry in time-frequency characteristics of the central EEG electrodes, appearance of rhythms in the frequency range of 4-6 Hz and the disruption of dominant rhythm for HC and PD patients and used them as their features for detection of early stages of the disease [53]. As it was mentioned one of the main issues in diagnosis of PD is the inability to access SN part of the brain. Recently several researchers have attempted to simulate the electrical potential at the basal ganglia using different channels of real EEG data of PD patients [54]. In their studies, they employed the BESA toolbox and the location of basal ganglia to simulate the electrical activity at the SN location. After the simulation they used the total power of alpha and beta rhythms for classification of different stages of the disease. This study has taken place while practicing dopamine inducing activities to increase

the accuracy of the classification. Unlike their approach, the approach employed in this study used the rest state data of patients which has a much lower dopamine level. Having a lower level of dopamine will make the classification more difficult; however, it is much easier to attain rest state data. This study made use of much more advanced features instead of focusing on localization of channels.

Another method that has been widely utilized in EEG signal processing is HOS analysis. Based on previous studies, HOS features have been especially effective in classification of PD and possibly other neurodegenerative diseases. In this research, a hierarchy approach is employed for diagnosis and classification of stages of PD from EEG signals.

Previous studies in this area tend to focus on classification approaches and neglect to do an in dept analysis of the HOS features. Therefore, the focus of this work is to gather relevant statistical features for this application and combining the findings of previous studies and developing new HOS features based on the characteristics of the signal to improve the performance. In Chapter 3 some of the mathematical definitions and concepts of HOS analysis are discussed. These concepts are necessary to gain a good understanding of the new features that were developed and used in Chapter 4.

2.4 Summary

Neurodegenerative disease is one of the main branches of neurological disorders and PD is one of the most common on such diseases. Although diagnosis is mostly performed by neurologists, biomedical signal processing can act as an important aid in this process. Signal such as speech, gait, and EEG have been widely used to improve the accuracy of the diagnosis.

In this chapter, the process of using EEG signals as diagnostics tools was explained and some of the state-of-the-art in this field were investigated. The next chapter will explain some

fundamental knowledge on the field of HOS analysis which will be necessary to understand mathematical aspects and the purpose of this work in later chapters.

Chapter 3: Higher Order Statistics²

3.1 Background

The second order frequency spectrum is the Fourier transform of the second order autocorrelation function; it also goes by the name of the power spectrum and is extensively used in signal processing. Spectra of the higher orders are described in terms of HOS such as moments and cumulants of a signal. For instance, the third and fourth order standardized moments are skewness (S), and kurtosis (K) while the third and fourth order spectra are the Fourier transforms of the third and fourth order autocorrelation functions and are named bispectrum (B) and trispectrum (T), respectively. Spectra of the higher orders defined as the Fourier transform of the higher order autocorrelation and while features such as entropy are derived from the spectrum (and therefore can be called a HOS feature) others such as cumulants and moments are defined in time domain.

Bispectrum is widely used in the biomedical field especially EEG signal processing even now in methods such as Bispectral Index (BIS). BIS is a monitoring method used to supplement Guedel's classification for determining the depth of anesthesia and monitoring the level of consciousness. A detailed description of HOS applications in EEG analysis is presented in the next section. The property of bispectrum preserves the phase information and it is highly beneficial, both for signal reconstruction and for analyzing the nonlinear quadratic relations created by the coupling phenomenon amid different frequencies of the signal. The second, third and fourth order

² This chapter was published in [6]. Permission is included in Appendix A.

spectra which are the Fourier transforms of the matching autocorrelations (for periodic signals) are given in equations (1) - (3), respectively.

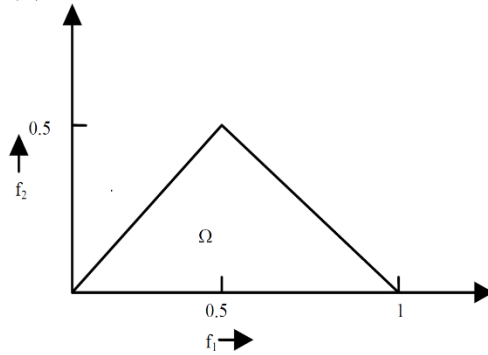


Figure 3.1 The region of non-redundancy.

$$PS_2^{\hat{x}}(f) = X(f)X^*(f) \quad (1)$$

$$PS_3^{\hat{x}}(f_1, f_2) = B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2) \quad (2)$$

$$PS_4^{\hat{x}}(f_1, f_2, f_3) = T(f_1, f_2, f_3) = X(f_1)X(f_2)X(f_3)X^*(f_1 + f_2 + f_3) \quad (3)$$

In these equations X is the Fourier transform of signal, x and is defined in the frequency domain rather than in the time domain, therefore, the variable also changes from n to f and X^* is the complex conjugate of X . Here, equations (1), (2), and (3) present the 2nd, 3rd and 4th order spectra which are the power spectrum, bispectrum and trispectrum. Bispectrum is defined in the bi-frequency domain; therefore, it is a function of two frequencies. Similarly, trispectrum is defined in tri-frequency and is a function of three frequencies. The bicoherence and tricoherence which are the normalized versions of bispectrum and trispectrum are given in equations (4) and (5) [55].

$$B_{norm}(f_1, f_2) = \frac{E[X(f_1)X(f_2)X^*(f_1+f_2)]}{\sqrt{P(f_1)P(f_2)P(f_1+f_2)}} \quad (4)$$

$$T_{norm}(f_1, f_2, f_3) = \frac{E[X(f_1)X(f_2)X(f_3)X^*(f_1+f_2+f_3)]}{\sqrt{P(f_1)P(f_2)P(f_3)P(f_1+f_2+f_3)}} \quad (5)$$

To classify different EEG signals, characteristics that can act as features are required; characteristics such as mean of magnitude and the phase entropy. The equations of these features are given in equations (6), (7), and (8).

$$\text{Mean of magnitude: } M_{ave} = \frac{1}{L} \sum_{\Omega} |B(f_1, f_2)| \quad (6)$$

$$\text{Phase entropy: } P_e = \sum_n p(\psi_n) \log p(\psi_n) \quad (7)$$

$$p(\psi_n) = \frac{1}{L} \sum_{\Omega} l(\varphi(B(f_1, f_2)) \in \psi_n) \text{ for } -\pi + \frac{2\pi n}{N} \leq \psi_n \leq -\pi + \frac{2\pi(n+1)}{N} \text{ and } n = 0, 1, \dots, N-1 \quad (8)$$

Where Ω indicates the region of f_1 and f_2 in Figure 3.1, φ is the phase and $l(\cdot)$ is the indicator function which gives “1” when the value of phase is within the range that is specified by ψ_n in (9) and “0” otherwise. The reason for using this region is that this section, which is a quarter of the bispectrum (both magnitude and phase), holds the actual data and the rest are redundancies. This region of non-redundancy can be seen in Figure 3.1.

Shannon entropy is used to measure the bispectrum phase entropy in equation (7). Mean magnitude in equation (6) can be used to differentiate between data with similar second order spectrum but it is susceptible to amplitude changes. This problem can be solved by normalization. For random systems the entropy is high and as they become more periodic and predictable the entropy decreases; until it gets to zero for a complete harmonic, periodic and predictable process.

Besides bispectrum phase entropy there are three other entropies which are imperative. Normalized bispectral entropy (BE_1), normalized bispectral squared entropy (BE_2) and normalized bispectral cubic entropy (BE_3) given in equations (9), (11), and (13), respectively. They are defined as entropies of normalized bicoherence and square normalized bicoherence given in equations (10), (12), and (14) [56].

$$BE_1 = - \sum_n P_n \log P_n \quad (9)$$

$$P_n = \frac{|B(f_1, f_2)|}{\sum_{\Omega} |B(f_1, f_2)|} \quad (10)$$

$$BE_2 = - \sum_i P_i \log P_i \quad (11)$$

$$P_i = \frac{|B(f_1, f_2)|^2}{\sum_{\Omega} |B(f_1, f_2)|^2} \quad (12)$$

$$BE_3 = - \sum_j P_j \log P_j \quad (13)$$

$$P_j = \frac{|B(f_1, f_2)|^3}{\sum_{\Omega} |B(f_1, f_2)|^3} \quad (14)$$

The phase of bispectrum does not change with a time-shift, the way Fourier phase does [57]. Equation (15) defines the bispectral invariant, P_a , which is the phase of the integrated bispectrum along the radial line with the slope equal to 'n'.

$$P_a(n) = \arctan\left(\frac{I_i(n)}{I_r(n)}\right) \quad (15)$$

$$I_r(n) = \text{Re} \left(\int_{f_1=0^+}^{\frac{1}{1+n}} B(f_1, nf_1) df_1 \right) \quad (16)$$

$$I_i(n) = \text{Im} \left(\int_{f_1=0^+}^{\frac{1}{1+n}} B(f_1, nf_1) df_1 \right) \quad (17)$$

The I_i and I_r used in equation (15) are defined in equations (16) and (17). As it was mentioned, HOS analysis achieves better results for signals with low SNR. The higher orders of a perfect Gaussian signal are zero; therefore, some higher order features can be used to measure the non-Gaussianity of the signal and to extract the independent non-Gaussian source signals from the mixture. This characteristic of HOS features can be used for detection and classification of non-Gaussian signals.

In 1982, Hinich developed an algorithm to test for Gaussianity and linearity of stationary time series. The basic notion was that if the third-order cumulant of a process is zero, then its Fourier transform or the bispectrum would also have to be zero, and therefore the bicoherence is zero. If the bicoherence is nonzero, the process is non-Gaussian and if it is a nonzero constant, the

process is linear, and non-Gaussian [58]. This test employs the mean of bicoherence power which is given in equation (18) as a reference [59].

$$P_s = \sum |B_{norm}(f_1, f_2)|^2 \quad (18)$$

3.2 Bispectrum Estimation

The estimation of power spectral density is one of the essential tools of signal processing. Similar methods can be used to estimate the bispectrum of the signal as well. Estimation methods for bispectrum are usually divided into two separate classes, conventional (or “Fourier type”) methods, and parametric methods; each of these classes includes various methods.

3.2.1 Conventional Bispectrum Estimation

Conventional methods include direct and indirect method [57]. Direct estimation tries to calculate the approximation of the definition of bispectrum given by equations (19) and (20) which are another form of equation (2).

$$X(f) = \int_{-\infty}^{\infty} e^{i2\pi ft} dZ(f) \quad (19)$$

$$E\{dZ(f_1)dZ(f_2)dZ^*(f_1 + f_2)\} = B(f)df_1df_2 \quad (20)$$

However, the indirect estimation methods try to calculate the approximation of the bispectrum based on equation (21).

$$B(f_1, f_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} R(n_1, n_2)e^{-2\pi i(f_1n_1+f_2n_2)} \quad (21)$$

These two estimates are different, however if they are calculated without using a window, practically they become identical. It has been shown that the conventional estimates are asymptotically unbiased and consistent. These methods usually have high variances, which means they require a large number of records to achieve smooth bispectral estimates. The bispectrum of our test EEG signal estimated by the direct method is illustrated in Figure 3.2.

The top plot in Figure 3.2 shows the magnitude of the estimated bispectrum, whereas the bottom plot shows the phase, and together they represent the third order spectra or bispectrum of the sample EEG signal. Since this method of estimation is based on the definition of bispectrum (the Fourier transform), it results in a more refined and smoother contour compared to parametric methods.

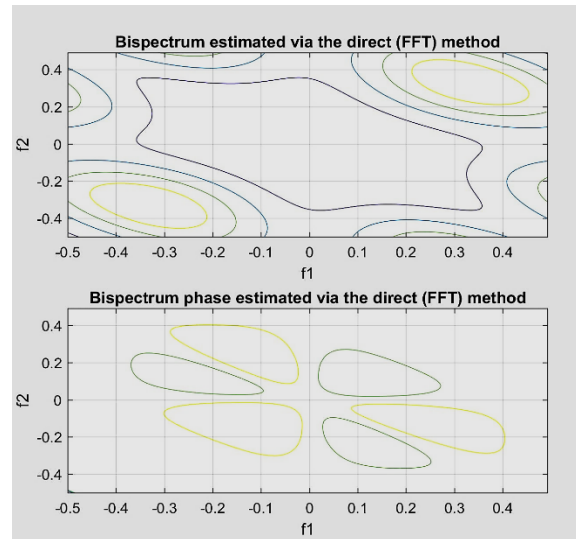


Figure 3.2 Magnitude and phase of the bispectrum of the sample EEG, estimated by the direct method.

Conventional methods have the advantages of ease of implementation and good estimate with very long data records; however, the abilities of the conventional methods are limited due to the “uncertainty principle” of the Fourier transform.

3.2.2 Parametric Bispectrum Estimation

Aside from the conventional methods, all other methods fall under the parametric estimator category. Therefore, the term parametric estimator covers a broad range of methods. Techniques such as the maximum-likelihood method of Capon and its modifications, cross-entropy methods (CE), and methods based on autoregressive-moving average (ARMA) models. Harmonic decomposition methods such as Pisarenko harmonic decomposition (PHD), Prony analysis,

multiple signal classification (MUSIC), singular value decomposition (SVD), etc. are also a part of parametric methods; however parametric methods are usually referred to the ones based AR, moving average (MA) and ARMA models [57, 60].

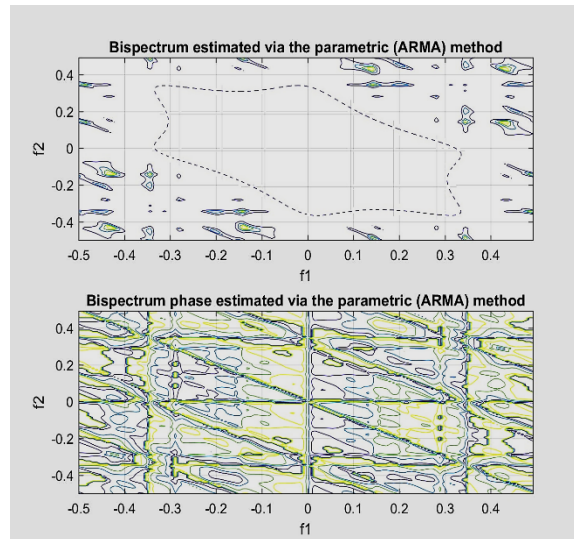


Figure 3.3 Estimated bispectrum using parametric, ARMA model.

The parametric estimation of the bispectrum of the sample signal using ARMA method is presented in Figure 3.3. Both these methods estimate the bispectrum, using the same data and since the ARMA method is iterative, it usually generates a number of small components compared to conventional methods. The outline of the contour of the direct method is fitted inside the magnitude to make it easier to spot the differences between Figures 3.2 and 3.3.

3.3 Quadratic Phase Coupling

In some cases when investigating a system, some interactions between the harmonics of the signal occur that are usually due to, the quadratic nonlinearity of the system. These interactions cause some changes in certain frequencies and phases of the output signal. Since the power spectrum loses the phase information, it is unable to detect these incidents. However, bispectrum is not phase-blind which means it can detect and quantify this phenomenon [57].

A simple example of this is given in [61], it states that if we have signal x as in equation (22), what would the phase be if x passes through a nonlinear system. This example considers a simple nonlinear system such as in equation (23). In this system, $h(t)$ (which is the output) will include the pairs of harmonics given in equation (24).

$$x(t) = A_1 \cos(2\pi f_1 t + \varphi_1) + A_2 \cos(2\pi f_2 t + \varphi_2) \quad (22)$$

$$h(t) = ax^2(t) \quad (23)$$

$$(2f_1, 2\varphi_1), (2f_2, 2\varphi_2), (f_1+f_2, \varphi_1 + \varphi_2), (f_1-f_2, \varphi_1 - \varphi_2) \quad (24)$$

The nonlinearity of the system causes an increase in the number of harmonics and therefore frequencies; this phenomenon is known as QPC [61]. Thus, the existence of QPC in a signal is a strong indication that the system is nonlinear.

As it was mentioned before, there are two main approaches to calculating the bispectrum, conventional and parametric methods. Overall, the conventional methods are more suited in regard with QPC and parametric approaches usually achieve higher resolution [61].

3.4 Higher Order Statistics in Electroencephalography Signal Processing

In the past, many studies such as [62] have used second order statistics and spectra for EEG signal analysis. However, after years of investigation researchers found out that due to the lack of phase information in second order statistics and the low SNR and high nonlinearity of EEG signals, HOS can achieve better analytical results. After that, the application of the HOS has become especially popular in the analysis of biological signals because of the ubiquity of inherently nonlinear characteristics of biological systems [63].

Third order statistics features are the most common HOS features for EEG analysis followed by fourth order statistics features. As it was mentioned in the previous section, these features usually include moments, cumulants, spectra, and different entropies. These features can

either be used to classify different types of EEG signals or to gain a better understanding of the normal and background EEG signal.

One of the applications of HOS in classifications of EEG is to detect epileptic seizures in patients. This application has been a popular subject of study among researchers in the past ten years [64-68]. Chua et al. [64] studied the properties of epileptic seizures in EEG signal using HOS. They used bicoherence patterns, entropies and other HOS features to analyze the EEG signals of both healthy subjects and epileptic patients. After feature extraction, they performed an ANOVA test where the features gave excellent p-values.

Zhou et al. [69] proposed a new feature extraction method based on HOS features to classify left/right-hand motor imagery from EEG signals. Motor imagery is a dynamic state where an individual mentally simulates the movement of a body part without the overt movement. They also employed LDA, SVM, and artificial neural network (ANN) for classification and achieved better results than 2003 winner on the same BCI dataset.

Hosseini et al. [70] used HOS analysis features along with SVM to classify emotional stress states in the two main areas of the valance-arousal space. Their analysis of the data showed that the radial basis function (RBF) kernel performs better than the others. In the same year, Ghandeharion et al. [71] devised an automatic ocular artifact suppression algorithm using WT coefficients and kurtosis to detect the artifact components of ICA. Concurrently, Acharya et al. [72] extracted HOS features from bispectrum and bicoherence during different stages of sleep (awake, rapid eye movement (REM), and four stages of sleep). Afterward, they fed these features to a Gaussian mixture model (GMM) in the classification step for automatic identification.

Javidi et al. [73] used kurtosis to develop a blind source extraction (BSE) algorithm to extract both circular and noncircular complex signals. BSE is a subcategory of BSS where instead

of dividing the signal into different sources, BSE attempts to find a specific component with certain characteristics. To achieve it, Javidi used HOS of latent sources, the deflation approach and sequential extraction based on the degree of kurtosis. They used this algorithm to remove EOG and EMG artifacts from EEG signals in real time.

Du et al. [68] tried to classify the epilepsy EEG signals based on higher order moments with another feature called the weighted center of bispectrum (WCB). In total they used 15 HOS features in their experiment and applied the PCA method to reduce the redundancies between the features. They employed eight machine-learning algorithms for classification, including multilayer perceptron artificial neural network (MLP-ANN), RBF network, random forest (RaF), rotation forest (RoF), logistic regression (LR), model trees (MT), simple logistic regression (SLR), and bagging (BA). They were able to achieve high accuracy, in some cases more than 98 %. The equations for WCB and absolute weighted center of bispectrum (aWCB) are given in equations (25) and (26).

$$WCB_x = \frac{\sum_{\Omega} f_1 B(f_1, f_2)}{\sum_{\Omega} B(f_1, f_2)}, WCB_y = \frac{\sum_{\Omega} f_2 B(f_1, f_2)}{\sum_{\Omega} B(f_1, f_2)} \quad (25)$$

$$aWCB_x = \frac{\sum_{\Omega} f_1 |B(f_1, f_2)|}{\sum_{\Omega} |B(f_1, f_2)|}, aWCB_y = \frac{\sum_{\Omega} f_2 |B(f_1, f_2)|}{\sum_{\Omega} |B(f_1, f_2)|} \quad (26)$$

The moment related feature that they used were the sum of logarithmic amplitudes of the bispectrum (H_1), the sum of logarithmic amplitudes of diagonal elements in the bispectrum (H_2), the first-order spectral moment of amplitudes of diagonal elements in the bispectrum (H_3), the second-order moment of amplitude of diagonal elements in the bispectrum (H_4), and the first-order spectral moment of amplitudes of the principal domain in the bispectrum (H_5) [68]. The equations of these features are given in equations (27) – (31), respectively.

$$H_1 = \sum_{\Omega} \log(|B(f_1, f_2)|) \quad (27)$$

$$H_2 = \sum_{\Omega} \log(|B(f_k, f_k)|) \quad (28)$$

$$H_3 = \sum_{k=1}^N k \log(|B(f_k, f_k)|) \quad (29)$$

$$H_4 = \sum_{k=1}^N (k - H_3)^2 \log(|B(f_k, f_k)|) \quad (30)$$

$$H_5 = \sum_{\Omega} \sqrt{i^2 + j^2} \log(|B(f_i, f_j)|) \quad (31)$$

Lay-Ekuakille et al. [74] used HOS features along with decimated signal diagonalization (DSD), which can process exponentially damped signals to classify epilepsy patients. Shafiul Alam et al. [75] also tried using HOS features to develop a seizure detection method. They used different moments and cumulants as features extracted from the EEG signal in the EMD domain. EMD is particularly well suited for analyzing nonstationary and nonlinear signals such as an EEG and combined it with an ANN for classification. Yuvaraj et al. [76] compared the power spectrum and HOS features to categorize EEG emotional states in PD patients. They used KNN and SVM algorithms for classification. The result showed higher accuracy when HOS features were used. Mahajan et al. [77] used modified multiscale Sample entropy, kurtosis, Wavelet decomposition, and ICA to develop a fast unsupervised and fully automatic algorithm for EOG artifact identification and removal, from EEG signal. Wang et al. [78] compared the Gaussianity of the EEG signal from a number of AD patients and healthy people by using kurtosis as the measure of non-Gaussianity. They found the average kurtosis of the EEG signals from the AD patients was to be much higher than that of healthy people indicating an abnormal dynamic within the AD patients' EEG pattern.

In 2014, Yuvaraj et al. [79] published a paper where they used EEG recordings of 20 PD patients and 20 HC during six basic emotions of happiness, sadness, fear, anger, surprise, and disgust to classify the emotional state of PD patients and HC. In their study, they employed three HOS features, mean of bispectral magnitude, BE_1 and BE_2 and used a SVM for classification. Continuing their previous work, in 2016 Yuvaraj et al. [80] used HOS features of EEG signals to

develop a novel Parkinson's disease diagnosis index (PDDI) for automated detection of PD. After analyzing thirteen different HOS features, they ranked them by their F value and used the first three highly rated features (H_1 , EB_1 , and H_2) to develop the PDDI. The mathematical equation they developed by trial-and-error is given in equation (32).

$$PPDI = \frac{\{(3.5*EB_1)+\{0.5*(H_1/H_2)\}\}}{10} \quad (32)$$

They employed nine different classifiers, DT, KNN, Fuzzy K nearest neighbor (FKNN), NB, probabilistic neural network (PNN) and SVM using three kernel functions, RBF and polynomial kernel functions order 2 and 3 (poly 2 and poly 3). They achieved similar results to Hosseini et al. in [70], and again, the SVM classifier using RBF kernel function (SVM-RBF) had the best performance and achieved the highest mean accuracy of 99.62 %, sensitivity of 100 % and specificity of 99.25 % while using higher order features [80]. Later in 2018 Oh et al. [81] published another paper using the same dataset for the same application; however, this time they employed 13-layer CNN for classification. They achieved the mean accuracy of 88.25 %, sensitivity of 84.71 % and specificity of 91.77 %. It can be seen that SVM-RBF combined with higher order features had a much better performance. Bairy et al. [82] used HOS features such as variance, kurtosis, normalized kurtosis, skewness, and normalized skewness along with linear predictive coding (LPC) and receiver operating characteristic algorithms to develop a computer-aided diagnosis (CAD) system to help the diagnosis of depression in patients.

Following his previous work in [70], Hosseini tried a hybrid approach for the recognition of different stages of epilepsy. He used HOS features along with a genetic algorithm (GA) and least square support vector machine (LS-SVM) with Gaussian and polynomial RBF kernels to recognize different epilepsy states. By comparing this algorithm and one without the GA, he has demonstrated that the existence of GA will improve the accuracy of the results [83]. During the

same time, Ikeda et al. [84] published a paper in which they used kurtosis along with exact low-resolution brain electromagnetic tomography (eLORETA) for Source estimation of epileptic activity in EEG signals. The eLORETA technique is a three-dimensional representation that shows the electrical activity of the brain. They compared this model with that of equivalent current dipole (ECD) and synthetic aperture magnetometry (SAM) and reported that eLORETA kurtosis analysis of epilepsy patients' EEG data might aid in the localization of spike activity sources.

As mentioned here, many studies that have been conducted on EEG analysis have used HOS features either directly or indirectly, and although HOS is not a new development, it is still a popular choice for those who study the field of EEG analysis.

3.5 Summary

The field of HOS analysis has been widely used in signal processing and have been especially beneficial for signals with random nature such as EEG signals. The generalized form of statistics is capable of analyzing more complex signals and provide information that would otherwise be inaccessible through second order statistics.

In this chapter, the mathematical aspect of HOS analysis along with different methods for estimation of bispectrum was explored. Then explanation of QPC and finally some applications of HOS analysis in EEG signal processing applications were provided. The next chapter will focus on applications of HOS analysis specifically for the diagnosis of neurodegenerative diseases and the proposed approach along with the newly developed HOS features will be introduced.

Chapter 4: Higher Order Statistical Analysis Framework for Classification of Parkinson's Disease³

4.1 Background

In the previous chapters it was shown that HOS analysis has many applications in EEG signal processing. Based on the previous studies on this subject, these applications are divided into 3 main categories [6]. Based on the focus of a study, HOS analysis might be applied in one or more of these categories.

Table 4.1 Summaries of studies conducted on category 1 arranged based on the artifacts and their sources.

Artifact type	Source	Reference	Artifact Removal Method
EOG artifact	Eye movement	[71, 73, 85]	ICA, wavelet, kurtosis, wavelet-ICA, automatic artifact suppression
EMG Artifact	Activity of the muscles near the electrodes	[85, 86]	Wavelet-ICA, ICA
EKG artifacts	Heart beat	[87]	CWT-ICA
Eye Blink Artifact	Blinking of the eye	[77]	Modified multiscale sample entropy, kurtosis, wavelet-ICA
Noise	Device, power line, environment, etc.	[73, 78]	Kurtosis-based BSE

The first category of applications of HOS in EEG analysis mostly involves identification and removal of artifacts or noise (preprocessing stage) [71, 73, 77, 78, 85, 87]. Some examples of such studies are shown in Table 4.1, along with the artifacts in question and their sources. These sources are explained in detail in [88] along with related studies. It must be noted that while BSS methods such as ICA and wavelet analysis are used for reducing correlation between components

³ This chapter was published in [125]. Permission is included in Appendix A.

of EEG, they can also be used for identifying and removing artifact and noise sources. The use of higher order features was shown to increase the accuracy rates in detecting EEG artifacts.

Table 4.2 Summaries of studies conducted on category 2 that contain methods using HOS features.

Method	Reference	HOS Features	Application
Kurtosis Based Adaptive Gradient Descent	[73]	K	Blind source extraction
JADE	[86]	K	Blind source separation
Fast-ICA	[89]	K	Blind source separation
Blind Deconvolution	[90]	S, K, 5 th and 6 th moments	Deconvolution
Jarque–Bera test	[91]	S, K	Goodness of fit test
D'Agostino's K-squared test	[92]	K	Goodness of fit test
Nonminimum Phase Detection	[93]	B	System detection
Infomax	[94]	K	Artificial neural network, blind source separation

In the second category, the HOS features are used to develop algorithms that can also be used in EEG analysis as well [63, 84] (and may also be beneficial to other fields). Since the features of the higher order are unaffected by Gaussian noise, they were used to develop general signal processing methods such as joint approximation diagonalization of eigen-matrices (JADE) and fast-ICA [86, 95-98] for BSS or kurtosis-based BSE [73] which have also proved useful in EEG analysis. Table 4.2 contains some of these methods along with their main applications. These methods are mostly used in the BSS and feature selection stages. Many of the methods mentioned in this table use kurtosis, which means that non-Gaussianity is used frequently for developing algorithms. This is not surprising considering non-Gaussianity plays an important role in determining the independency of signals.

The third category is where HOS features are directly used for classification problems and has been widely studied by researchers. In this section, different studies on classification of

different motor functions [69], emotional stress states [70], sleep stages [72], emotional states of PD patients [76] and diagnosis of PD [80], epilepsy [67, 68, 83, 84], and depression [82] are included. In Table 4.3, a summary of the previous studies conducted in this category along with features and classifiers that were used in each one is presented.

Table 4.3 Summaries of studies conducted on category 3 of Higher Order Statistics applications in EEG analysis.

Subject	Reference	HOS Features	Classifiers
Emotional stress states	[70]	B, B_{norm}, P_a	SVM-RFB
Motor function	[69]	H_1, H_2, H_3, H_4	LDA, SVM, NN
Sleep stages	[72]	$M_{ave}, EB_1, EB_2, WCB_x, WCB_y, H_1, H_2, H_3$	GMM
Emotional states of PD patients	[76]	M_{ave}, EB_1, EB_2	KNN, SVM-RBF
Diagnosis of PD	[80]	$B, B_{norm}, M_{ave}, P_e, EB_1, EB_2, WCB_x, WCB_y, aWCB_x, aWCB_y, H_1, H_2, H_3, H_4, H_5$	DT, FKNN, KNN, NB, PNN, SVM-poly 2, SVM-poly 3, SVM-RBF
Diagnosis of depression	[82]	K, K_{norm}, S, S_{norm}	LPC
Diagnosis of epilepsy	[67]	$B, B_{norm}, M_{ave}, P_e, EB_1, EB_2, EB_3, H_1, H_2, H_3, H_4, H_5$	DT, PNN, KNN, NB, PNN, SVM-poly 1, SVM-RBF
	[68]	$B, B_{norm}, M_{ave}, P_e, EB_1, EB_2, WCB_x, WCB_y, aWCB_x, aWCB_y, H_1, H_2, H_3, H_4, H_5$	MLP-ANN, BA, RaF, RoF, RBF-network, LR, MT, SLR
	[83]	$B, B^2, B_{norm}, B_{norm}^2$	LS-SVM

In this study, all categories of HOS applications (mentioned above) were investigated where HOS based methods have been used in the BSS, artifact removal, and feature extraction steps. However, the focus of the study was on the third category where HOS features have been directly used for classification purposes.

4.2 Proposed Framework

The proposed idea of this research is to perform a hierarchy classification of PD using newly developed HOS features to improve the classification performance. In this study, two

datasets were used, one for classification of PD vs HC and the other for classification of stages of PD. To perform this task, a total of 35 features were extracted, described in the previous section, for each data point of our datasets; some of these features were existing higher-order features, some lower-order features and some of them I have developed and are used here for the first time. Then, these extracted features were used to perform classification on both datasets; first for the PD vs HC and then for the stages of PD.

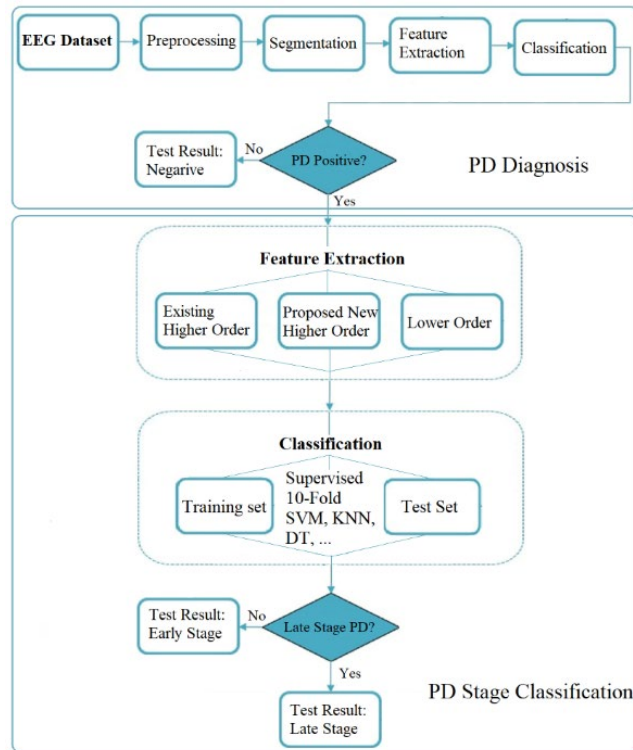


Figure 4.1 Proposed hierarchy approach for diagnosis of PD and classification of the stages.

The process of this model can be seen in Figure 4.1. In this model, the first phase describes previous studies that classify PD from healthy and the second phase describes this study, which aims to classify the stages of PD after obtaining positive result from the initial phase. Since the data points have labels, supervised classification algorithms were used. For this study, several conventional machine learning methods such as SVM, KNN and DT along with ensemble learning

methods were used to perform the classification. In this chapter, the methods employed in the model such as the data acquisition, preprocessing, feature extraction, and classification methods are discussed. Chapter five will include the resulted features and the classification performance along with the comparison with state-of-the-art methods.

Another area of this research work is dedicated to finding the most relevant frequency band for diagnosis of PD and extracting HOS features only from that band. The idea is that by finding the most relevant part of the signal and removing the unrelated parts, it could be possible to improve the classification performance and accuracy of the diagnosis. Several studies have pointed to the fact that certain changes occur in the alpha and beta rhythms in PD patients [99, 100], and the total power of alpha and beta rhythms collected from EEG signals have already been used for diagnosis of PD [54, 101, 102]. Alpha rhythm is associated with a state of relaxation usually when the eyes are closed. Beta rhythm is associated with a state of alertness, usually when the eyes are open. Therefore, in this study, besides using the whole signal, the HOS features were also extracted from alpha and beta frequency bands and compared their results for classification of PD vs HC.

4.3 Data Acquisition

Data acquisition is an important part of any experiment, and that is especially true for biomedical applications. As discussed, in this study two datasets were used which overall contain EEG data from 96 participants. Both datasets were collected after informed consent was given by the participants and all procedures were in accordance with the ethical standards.

The first dataset that was used in this study (dataset #1) is a part of a bigger open-source dataset of 28 PD patients and 28 HC participants. This dataset was originally a part of a study conducted by Cavanagh et al. [103], where they used this dataset for investigating the effects of PD on ERP components; however, in this study, only the rest state portion of their dataset was

used. The study conducted by Cavanagh et al. [103] was approved by the Institutional Review Board at the University of New Mexico and all participants were paid by the hour and gave written informed consents. The EEG signals were recorded using a 64 channel Brain Vision system (CPz reference, AFz Ground) where the signal of each electrode was sampled at 500 Hz. Table 4.4 shows some basic information about the device that was used for recording the EEG signals. In the portion of the data that was used in this study, all PD patients are OFF-medication. To record the OFF-medication EEG signals, the patients go through a 15-hour overnight withdrawal from their prescription medication. The dataset contains 18 male and 10 female PD patients and each PD patient has an HC match of the same gender and is relatively close in age. The average years since diagnosis for all PD patients is 5.53 yrs. The average UPDRS motor score of the PD participants is 23.78 with the lowest being 10 and the highest being 41.

Table 4.4 General information on the collected EEG signals in dataset #1.

Emotiv EPOC headset	
Number of channels	64 (including one EOG channel)
Reference electrodes	CPz reference, AFz Ground
Connectivity	Wired
Electrode type	Ag/Agcl
Channel names	VEOG, Fp1-Fp2, AFz, AF3-AF4, AF7-AF8, Fz, F1-F8, FCz, FC1-FC6, FT7-FT10, Cz, C1-C6, T7-T8, CP1-CP6, TP7-TP10, Pz, P1-P8, POz, PO3-PO4, PO7-PO8, Oz, O1-O2
Sampling rate	500 Hz (samples per second)

The second dataset that has been used in this study (dataset #2) was originally collected at the Hospital University Kebangsaan Malaysia (HUKM) medical center in Kuala Lumpur, Malaysia and contains EEG signals of 20 PD patients and 20 HC between the ages of 45 and 65 [79]. Based on the H&Y scale the PD patients are divided into, 2 stage one, 11 stage two and 7 stage three. Since there are only 2 stage one patients, for stage classification the third stage was considered as late-stage and the first two stages are combined to create the early-stage. This EEG

dataset was collected using a “Emotiv EPOC headset” and includes rest state EEG signals. The EEG signals were collected at 128 Hz sampling rate and each contain 14 channels. All PD patients and HCs reported to be right-handed and were also tested by Edinburgh handedness inventory (EHI). The PD patients were optimally medicated (ON state) during the testing session. Table 4.5 shows some basic information about the collected signals.

Table 4.5 General information on the collected EEG signals in dataset #2.

Emotiv EPOC headset	
Number of channels	14 plus
Reference electrodes	CMS/DRL references, P3/P4 locations
Connectivity	Proprietary Wireless, 2.4 Hz band
Electrode type	Ag/Agcl
Channel names	AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4
Sampling rate	128 Hz (samples per second)

4.4 Preprocessing and Artifact Removal

The EEG signals were recorded while the participants were restfully seated in a quiet room with their eyes closed to reach a relaxed state of mind. They were told to avoid any movement (including eye movement) as much as possible. In case any participant blinked by accident, the eye blinking artifacts were removed by removing the samples with amplitudes of more than 80 μ V. The signals were then filtered using a 6th order Butterworth bandpass filter (with cutoff frequencies of 1 and 49 Hz) to remove the high and low frequency noise; then the AAR 1.3 extension from EEGLAB was used to remove EMG and EOG artifacts. EMG is the artifact caused by muscle activity near the electrode and EOG is the artifact caused by blinking or movement of the eyes [104]. In the next phase of preprocessing, fast-ICA algorithm, an ICA technique was employed, with EEGLAB v2019.1 to extract the components from the channels. ICA is a BSS method that attempts to separate the independent components of the observation signals based on the assumption that the sources are independent and non-Gaussian. For each moment in time, there

are several sources of electrical potential inside the brain associated with different regions. The EEG electrodes measure the electrical potential on the surface of the scalp; however, the measured signal is a mixture of signals generated by the sources. Therefore, there is usually some correlation between different channels. To get an estimate of the original sources, BSS methods are required. In EEG signal analysis, ICA has been used by researchers as the main BSS method. After the BSS step each signal is moved from the channels to component domain and is divided into 1000 and 2000 sample segments. Each of the segments or datapoints will go through feature extraction and will then be used for classification.

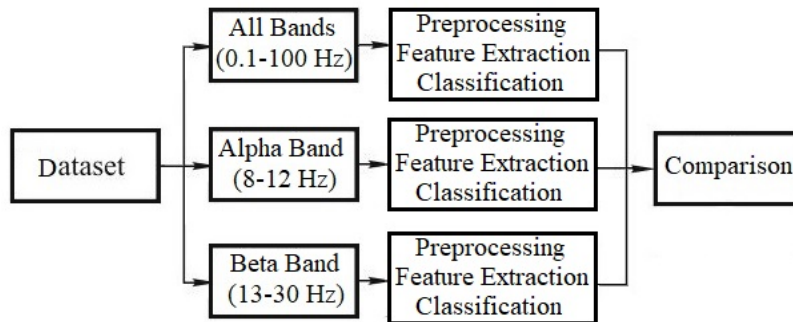


Figure 4.2 The process of PD diagnosis for the all-band vs alpha band vs beta band.

As it was mentioned in classification of PD vs HC, the alpha and beta bands were extracted and used to perform the classification. To do so, the signals were filtered using an 8-12 Hz bandpass infinite impulse response (IIR) filter to extract the alpha band and a 13-30 Hz bandpass IIR filter to extract the beta band. Each band then goes through the same preprocessing feature extraction and classification steps. This process is demonstrated in Figure 4.2.

4.5 Feature Extraction

To improve the accuracy of the model, 15 new higher order features were developed and combined with 14 existing higher order and 6 lower order features for classification of the stages of PD. In this section, each of these features will be explained along with reasons for their selection.

After calculating these features, they were normalized before being used for classification. Therefore, all the values in Tables 2 – 6 are averages of the normalized features.

4.5.1 Lower Order Statistical Features

There are many first and second order statistical features that have been used for EEG analysis. In this study, six of them were chosen based on previous studies, where each one was used directly in classification of EEG signals either for diagnosis of PD patients or other purposes. The first two features are the total power of alpha and beta rhythms. The powers of these rhythms have been previously used for both classification of PD [102] and classification of stages of PD [54]. Four nonlinear lower order features (LLE, HE, ApEn, and CD) that have been extensively used in EEG analysis were also used.

LLE is a nonlinear feature that expresses the convergence of the signal in phase space. This is one of the most useful quantification methods for chaotic dynamics of a system and is able to successfully detect patterns in EEG signals. The LLE feature of EEG has been extensively used in BCI systems [105, 106] and analysis of neurological disorders [107, 108].

HE is a nonlinear feature that measures and quantifies the correlation of different points of a signal in the time domain [109]. This parameter can be used to investigate the presence and degree of long-distance dependency of the signal. HE has been widely used for the classification of EEG signals in different applications [110, 111].

ApEn is a nonlinear feature that is obtained from the Kolmogorov entropy and quantifies the level of complexity of a signal in the time domain. This feature has been widely used in different areas of biomedical signal processing; it has been especially useful in EEG signal analysis [112, 113] since it is mostly suited for signals with low SNR [114].

CD is a feature that measures the complexity of a system by ascertaining the dimensionality of the signal's state space. This feature has been used in the classification of EEG signals for different applications such as prediction of treatment response for repetitive transcranial magnetic stimulation (rTMS) [115], diagnosis of autism [116], etc.

4.5.2 Higher Order Statistical Features

As it was mentioned in the previous sections, an overall of 29 higher order features were used in this study; out of all these higher order features, 23 were extracted from bispectrum and the other 6 are higher order standardized moments. The higher order moments are considered as pre-existing features however they have not been previously used in this field. From the 23 features that were extracted from bispectrum, 8 are existing features that have been used for EEG analysis of PD patients [117]. By investigating the higher order moments of different stages of PD, a sudden increase was noticed occurring in the average of the third stage in standardized moments.

The next set of features that was selected for this study were the bispectral entropies. The BE_1 , BE_2 , and BE_p features were selected. Each of these entropies was calculated from the bispectrum and are considered as third order features.

$$BE_1 = - \sum_{n=0}^{N-1} P_1 \log P_1 \text{ where } P_1 = \frac{|B(f_1, f_2)|}{\sum_{\Omega} |B(f_1, f_2)|} \quad (33)$$

$$BE_2 = - \sum_{n=0}^{N-1} P_2 \log P_2 \text{ where } P_2 = \frac{|B(f_1, f_2)|^2}{\sum_{\Omega} |B(f_1, f_2)|^2} \quad (34)$$

In equations (33) and (34), B is the bispectrum of the frequencies f_1 and f_2 , Ω is the region of non-redundancy of bispectrum [6], and N is the total number of points in Ω .

$$BE_p = \sum_{n=0}^{N-1} p(\psi_n) \log p(\psi_n) \text{ where } p(\psi_n) = \frac{1}{N} \sum_{\Omega} l(\varphi(B(f_1, f_2) \in \psi_n)) \quad (35)$$

In equation (35), φ is the phase, $l(\cdot)$ is a function where its value is 1 when the phase falls within ψ_n (and 0 otherwise), and ψ_n is described as in equation (36).

$$\psi_n = \left[-\pi + \frac{2\pi n}{N}, -\pi + \frac{2\pi(n+1)}{N}\right] \quad (36)$$

The next set of features used in this study were the sum of the log of bispectrum amplitudes (H_1), the trace of the log of bispectrum amplitudes (H_2), the first-order moment of the diagonal elements of bispectrum amplitudes (H_3), the second-order central moment of the diagonal elements of bispectrum amplitudes (H_4), and the first-order moment of the log of bispectrum amplitudes in the principal domain (H_5) [68]. The bispectral entropies and H_1 - H_5 features have been previously used by researchers for PD classification and other areas of EEG analysis [117].

$$H_1 = \sum_{\Omega} \log(|B(f_i, f_j)|) \quad (37)$$

$$H_2 = \sum_{\Omega} \log(|B(f_n, f_n)|) \quad (38)$$

$$H_3 = \sum_{\Omega} n \cdot \log(|B(f_n, f_n)|) \quad (39)$$

$$H_4 = \sum_{\Omega} (n - H_3)^2 \log(|B(f_n, f_n)|) \quad (40)$$

$$H_5 = \sum_{\Omega} \sqrt{i^2 + j^2} \log(|B(f_i, f_j)|) \quad (41)$$

Similar to equation (33), in equations (37) – (41), B is the bispectrum and Ω is the region of non-redundancy. Together the H_1 - H_5 features, three BE features and six μ^n features are the pre-existing higher order features that were selected for this study. As it was mentioned before, the first two set of features (H and BE) have been previously used in EEG analysis for PD classification, however the use of μ^n is a novel approach.

4.5.3 Proposed New Higher Order Statistical Features

Up until this point all the features that were mentioned already exist and have either been used in PD classification or in other areas of EEG analysis. In this section, 15 new HOS features that were developed will be described and used for PD stage classification. In practice, each of these features improves the final accuracy. These features can be divided into three sets of five, all sets were inspired from the H_1 - H_5 features.

The first set is essentially the same as the H₁-H₅ features, with the only difference being that they are calculated from the complex bispectrum where H₁-H₅ features are calculated from the amplitude of bispectrum. In the new features the phase of the bispectrum can also affect the values. If the bispectrum of each point is considered in the exponential form of equation (42), then equations (37) - (41) are calculated based on the log of A for each point. However, by calculating the log of bispectrum instead of the log of bispectrum amplitude equation (43) can be achieved.

$$B(f_1, f_2) = A_{i,j} e^{i\varphi_{i,j}} \quad (42)$$

$$\log B(f_i, f_j) = \log(A_{i,j} e^{i\varphi_{i,j}}) = \log(A_{i,j}) + \log(e^{i\varphi_{i,j}}) = \log(A_{i,j}) + i \cdot \varphi_{i,j} \quad (43)$$

In (42) and (43), $A_{i,j}$ is the amplitude and $\varphi_{i,j}$ is the phase of B for f_i and f_j , and result of equation (43) is a complex number. The magnitude of log bispectrum of equation (44) can be calculated based on equation (43). By replacing the log of bispectrum amplitudes in equation (44), equations (37) – (41) can be rewritten to get equations (45) – (49). I named these features complex H (CH) features.

$$|\log B(f_i, f_j)| = |\log(A_{i,j}) + i \cdot \varphi_{i,j}| = \sqrt{(\log(A_{i,j}))^2 - (\varphi_{i,j})^2} \quad (44)$$

$$CH_1 = \sum_{\Omega} \sqrt{(\log(A_{i,j}))^2 - (\varphi_{i,j})^2} \quad (45)$$

$$CH_2 = \sum_{\Omega} \sqrt{(\log(A_{n,n}))^2 - (\varphi_{n,n})^2} \quad (46)$$

$$CH_3 = \sum_{\Omega} n \cdot \sqrt{(\log(A_{n,n}))^2 - (\varphi_{n,n})^2} \quad (47)$$

$$CH_4 = \sum_{\Omega} (n - H_3)^2 \sqrt{(\log(A_{n,n}))^2 - (\varphi_{n,n})^2} \quad (48)$$

$$CH_5 = \sum_{\Omega} \sqrt{i^2 + j^2} \cdot \sqrt{(\log(A_{i,j}))^2 - (\varphi_{i,j})^2} = \sqrt{(i^2 + j^2) \cdot ((\log(A_{n,n}))^2 - (\varphi_{n,n})^2)} \quad (49)$$

As discussed, the H₁-H₅ and CH₁-CH₅ features are calculated from the two functions of the log of bispectrum amplitudes and the amplitude of the log bispectrum. The new features are calculated based on the 2D inverse Fourier transforms of these functions. By applying the inverse Fourier transform, the function domains are changed from log bispectrum to bicepstrum. Applying the 2D inverse Fourier transform to the log of bispectrum amplitudes (the function of H₁-H₅ features) forms the real bicepstrum of the signal and by using it, features similar to H can be calculated. These features which I have named real bicepstral H (RBH) features can be calculated using equations (50) – (54).

$$RBH_1 = \sum_{\Omega} \mathcal{F}^{-1}(\log(|B(f_i, f_j)|)) \quad (50)$$

$$RBH_2 = \sum_{\Omega} \mathcal{F}^{-1}(\log(|B(f_n, f_n)|)) \quad (51)$$

$$RBH_3 = \sum_{\Omega} n \cdot \mathcal{F}^{-1}(\log(|B(f_n, f_n)|)) \quad (52)$$

$$RBH_4 = \sum_{\Omega} (n - H_3)^2 \mathcal{F}^{-1}(\log(|B(f_n, f_n)|)) \quad (53)$$

$$RBH_5 = \sum_{\Omega} \sqrt{i^2 + j^2} \mathcal{F}^{-1}(\log(|B(f_i, f_j)|)) \quad (54)$$

Similar to RBH₁-RBH₅ features, complex bicepstral H (CBH) features were computed using the CH₁-CH₅ features; by using equation (44), these features can be obtained through equations (55) – (59).

$$CBH_1 = \sum_{\Omega} \mathcal{F}^{-1} \left(\left| \log \left(B(f_i, f_j) \right) \right| \right) \quad (55)$$

$$CBH_2 = \sum_{\Omega} \mathcal{F}^{-1} \left(\left| \log \left(B(f_n, f_n) \right) \right| \right) \quad (56)$$

$$CBH_3 = \sum_{\Omega} n \cdot \mathcal{F}^{-1} \left(\left| \log \left(B(f_n, f_n) \right) \right| \right) \quad (57)$$

$$CBH_4 = \sum_{\Omega} (n - H_3)^2 \mathcal{F}^{-1} \left(\left| \log \left(B(f_n, f_n) \right) \right| \right) \quad (58)$$

$$CBH_5 = \sum_{\Omega} \sqrt{i^2 + j^2} \mathcal{F}^{-1} \left(\left| \log \left(B(f_i, f_j) \right) \right| \right) \quad (59)$$

The H_1 - H_5 features are defined based on the real bispectrum, however these features were recalculated based on the real and complex bispectrum to create new features. By using the principles of HOS, it was possible to create more comprehensive feature sets. The diagonal slice of bispectrum has already been used in several areas of signal processing, however this is the first time HOS features are defined in cepstral (complex or real) domain.

4.6 Summary

The HOS analysis of EEG signals has been used for many different applications such as denoising and artifact removal, developing methods such as ICA and finally classification and more specifically, diagnosis of diseases.

In this chapter, the proposed framework for diagnosis of PD vs HC and then stages of PD from HOS analysis of EEG signals were explained. The two datasets that were used in this study were introduced as well. To achieve the goals of this study, a number of previously developed statistical features including higher order were extracted from the datasets. Besides these features, 15 new features were also introduced that were specifically designed to improve the performance of the proposed model. In the next chapter, a number of classifiers were employed to compare the performance of the proposed method with that of the state-of-the-art methods in EEG classification of PD.

Chapter 5: Diagnosis of Parkinson's Disease and Classification of Stages

Using the Proposed Approach

5.1 Diagnosis of Parkinson's Disease from Healthy Control

By performing an exhaustive search, the features were ranked and 22 features out of the total 35 that were extracted were chosen for classification of PD from HC. It was observed that for this application the CH and CBH features do not have a noticeable contribution to the classification performance. Here, all the features that were used for classification are listed.

1. Lower Order Features (LLE, HE, ApEn, CD)
2. Bispectral Features (H_1 - H_5)
3. Bicepstral Features (RBH_1 - RBH_5)
4. Bispectral Entropies (BE_1 , BE_2 , BE_p)
5. Higher Order Moments (μ_3 - μ_7)

The details on each of these features were discussed in Chapters 3 and 4. For this part of the experiment, a total of 4 lower order and 18 higher order features were used. Table 5.1 shows the average of all the features that were extracted from the EEG signals for each class (PD and HC). Due to previous studies' positive conclusions on the effects of PD on brain rhythms, in this study the HOS features of alpha and beta frequency bands were also computed. The average of the extracted features for the alpha band is presented in Table 5.2 and for the beta band in Table 5.3.

In this section, 5 ensemble classifiers were used to perform the classification. The reason for this study to focus on ensemble learning is due to the results of the previous experiment where

it became evident that ensemble classifiers have better performance than a single base learner; this fact is also supported by many other previously published works [118].

Table 5.1 The average of the feature extracted from the EEG signals for each class.

Features	HC	PD
CD	7.0726E+05	6.6531E+05
ApEn	2.7191E+03	2.5472E+03
LLE	8.8651E+04	8.3069E+04
HE	3.7263E-01	5.0374E-01
H ₁	3.0196E+06	2.8484E+06
H ₂	2.3446E+01	2.3849E+01
H ₃	1.1327E+01	1.0713E+01
H ₄	1.3635E+03	1.2752E+03
H ₅	5.1717E-01	4.5143E-01
RBH ₁	3.0278E+06	2.8526E+06
RBH ₂	2.4085E+02	2.3514E+02
RBH ₃	-1.7472E+10	-2.2215E+10
RBH ₄	1.6644E+01	7.9868E+00
RBH ₅	1.9559E+00	1.9555E+00
BE ₁	1.5567E+00	1.5134E+00
BE ₂	-8.0891E+01	-7.7823E+01
BE _P	7.0219E-01	7.6176E-01
μ ₃	5.1809E-03	-7.2330E-04
μ ₄	3.6470E+00	4.1244E+00
μ ₅	3.2122E-01	-1.3928E+00
μ ₆	4.4180E+01	1.0940E+02
μ ₇	8.1807E+01	-8.8468E+02

Ensemble learning is a general term used to describe methods that fuse the result of multiple machine learning models (mostly supervised) to improve the overall prediction performance. In many machine learning applications, ensemble methods are viewed as state-of-the-art methods. In most cases, these methods improve the performance of the model by using multiple training sessions and integrating their resulting predictions [119]. The base learner can be any machine learning model such as DT, LDA, KNN, etc. In the ensemble approach, the errors of a single model are most likely to be compensated by others, improving the performance of this

model compared to a single base learner. Over the past few years, researchers have increasingly utilized ensemble learning methods for different applications; the research in this field has resulted in new techniques that have exceptionally high performance compared to other methods. Ensemble methods are able to improve performance due to three main reasons, overfitting avoidance, computational advantage, and representation [119].

Table 5.2 The average of the feature extracted from the alpha band for each class.

Features	HC	PD
CD	7.8011E+05	7.3852E+05
ApEn	3.0890E+03	2.9260E+03
LLE	9.9353E+04	9.4097E+04
HE	1.1706E+00	1.1729E+00
H ₁	3.5096E+06	3.3289E+06
H ₂	4.2473E+01	4.1843E+01
H ₃	1.3911E+01	1.3345E+01
H ₄	1.5527E+03	1.4812E+03
H ₅	6.7332E-01	6.5171E-01
RBH ₁	3.5248E+06	3.3625E+06
RBH ₂	1.5020E+02	1.5034E+02
RBH ₃	-1.7688E+13	-8.9717E+12
RBH ₄	2.8931E+01	6.7276E+01
RBH ₅	1.8107E+00	1.8134E+00
BE ₁	4.0435E-01	4.0472E-01
BE ₂	1.6706E+02	1.6708E+02
BE _P	1.9711E+00	1.9632E+00
μ ₃	-4.9668E-04	4.7778E-05
μ ₄	4.7610E+00	4.7557E+00
μ ₅	-4.2279E-03	6.0883E-04
μ ₆	3.2728E+01	3.2672E+01
μ ₇	-4.1204E-02	9.1805E-03

Another advantage of this model is the use of parallel computing. Figure 5.1 presents the block diagram of ensemble learning and how these methods combine the results of multiple classifiers. Since the base classifiers in this figure operate separately from each other, they can work in parallel; therefore, the time that ensemble classification requires to be trained is not much

higher than a single classifier. Bagging, AdaBoost, RUSboost, and random subspace are some of the most applied ensemble methods.

Table 5.3 The average of the feature extracted from the beta band for each class.

Features	HC	PD
CD	7.9892E+05	7.5922E+05
ApEn	2.9608E+03	2.8063E+03
LLE	9.6700E+04	9.1701E+04
HE	2.0979E+00	2.1060E+00
H ₁	3.5159E+06	3.3441E+06
H ₂	4.1779E+01	4.1222E+01
H ₃	1.3322E+01	1.2751E+01
H ₄	1.4906E+03	1.4206E+03
H ₅	7.0062E-01	6.7743E-01
RBH ₁	3.5307E+06	3.3669E+06
RBH ₂	1.9099E+02	1.9085E+02
RBH ₃	-1.3535E+13	-6.8312E+12
RBH ₄	2.1055E+01	4.4771E+01
RBH ₅	1.7158E+00	1.7223E+00
BE ₁	2.5238E-01	2.5959E-01
BE ₂	2.0193E+02	2.0251E+02
BE _P	1.9719E+00	1.9717E+00
μ ₃	7.5935E-04	-1.4909E-04
μ ₄	1.6895E+01	1.6701E+01
μ ₅	1.5098E-01	-8.6415E-03
μ ₆	4.4575E+02	4.3905E+02
μ ₇	5.7329E+00	9.8680E-02

Another important factor in most biomedical classification is for the train and test sets to be completely separate; therefore, for segmented data (such as in our study), there should be no overlap between the training and testing groups. In this work for each method of classification, the training was performed 28 times with datapoints from 27 pairs of participants (PD and HC) and the datapoints from the remaining pair were used as test. By using this method, a modified CV approach was developed, where there was no overlap between the participants in the training and testing groups; similar to LOSO-CV but for two (one PD and one HC).

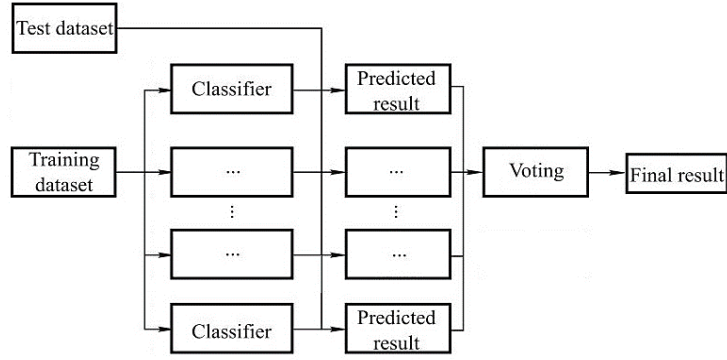


Figure 5.1 The diagram describing classification with ensemble learning.

As mentioned earlier, the features from alpha and beta bands were extracted as well in this study. Then each set of features was used separately for classification and their results were compared. For each of the five methods, the average testing performance of the leave-two-subjects-out CV (LTSO-CV) scheme is presented in Table 5.4. It can be seen that the highest performance was achieved by Bagged DT for the alpha band. Therefore, the overall testing accuracy of this classifier for each pair of the test group is presented in Table 5.5.

Table 5.4 The average diagnosis performance of all ensemble classifiers.

Ensemble Classifier	All Bands (%)		Alpha Band (%)		Beta Band (%)	
	Sen	Spe	Sen	Spe	Sen	Spe
AdaBoost DT	69.82	73.39	98.75	98.39	98.75	97.85
Bagged DT	65.17	76.25	99.10	99.28	98.21	98.57
RUSBoosted	58.75	63.03	95.35	93.92	73.21	97.32
Subspace KNN	48.39	56.07	96.42	86.60	86.25	78.92
Subspace discriminant	68.92	71.96	74.10	64.28	63.75	69.64

5.2 Classification of Early-Stage vs Late-Stage

In this section, a total of 35 features were extracted for each data point of the second dataset that was previously described and were used to classify the early (combined 1st and 2nd stages)

and late (3rd stage) stages of PD. All of the aforementioned features contributed to the classification performance improvement which is why unlike for the diagnosis of PD (the first part of the hierarchy model) all of the features were used in this section. The classification of stages of PD using HOS analysis is the second part of the hierarchy model described in Figure 4.1.

Table 5.5 The accuracy of diagnosis using bagged trees ensemble for all three sets

Test Group PD_ID/HC_ID	All Bands (%)	Alpha Band (%)	Beta Band (%)
801/894	92.5	97.5	100
802/908	57.5	100	100
803/8010	77.5	100	100
804/906	100	97.5	100
805/903	100	100	97.5
806/8060	90	100	95
807/893	40	100	95
808/909	95	100	97.5
809/911	50	95	97.5
810/895	87.5	100	97.5
811/913	80	100	100
813/900	52.5	100	97.5
814/896	92.5	97.5	97.5
815/899	50	100	100
816/914	52.5	100	97.5
817/910	70	100	97.5
818/890	60	100	97.5
819/891	50	100	100
820/912	52.5	100	100
821/905	97.5	100	97.5
822/904	27.5	100	100
823/892	70	100	97.5
824/902	62.5	100	100
825/901	60	100	97.5
826/898	50	100	100
827/897	80	97.5	100
828/8070	97.5	92.5	100
829/907	85	100	95
Average	70.71	99.19	98.39

In the overall model, the first step is to perform the classification of PD from HC and once a positive diagnosis is made, then the next step is to identify the stages of the disease by classifying the early vs late stages. Table 5.6 shows the averages of the normalized lower features that were discussed in this subsection for each stage of the disease

Table 5.6 Average of the normalized lower order features for each stage of PD.

Features	HC	Early-Stage	Late-Stage
CD	3.72E-01	4.09E-01	3.27E-01
ApEn	4.60E-01	4.45E-01	5.99E-01
LLE	5.72E-01	5.57E-01	6.77E-01
HE	1.22E-01	2.25E-01	5.13E-02
P_α	3.69E-03	3.35E-03	4.73E-03
P_β	2.75E-03	2.01E-03	2.15E-03

The increase in the higher order moments of the third stage is not unanimous between all patients therefore these features cannot be used to classify late-stage of PD on their own. The standardized moments up to 20th order were computed, however, after the 7th order the accuracy did not improve any further, which is why only the 3rd order up to 7th order were chosen in this study. The averages of the higher order standardized moments, the bispectral entropies and the H₁-H₅ features are given in Table 5.7 for each class.

In Table 5.8, the averages of the proposed features for each stage are provided; as it can be seen there are three sets, each with 5 features. The mathematical description of these features can be found in the previous chapter.

After feature extraction, since the data points are labeled, again supervised classification algorithms were used. In this section, several methods such as SVM, KNN and DT were chosen to perform the classification; aside from the conventional methods, ensemble classifiers such as boosted trees, RUS boosted trees and subspace KNN were also employed. All of these classifiers

were applied using the classification learner app in Matlab. For both sections, a 10-fold CV approach was used for all classification algorithms.

Table 5.7 The average of the normalized 3rd - 7th order standardized moments for each stage of PD.

Features	HC	Early-Stage	Late-Stage
μ''_3	-4.25E-04	6.09E-03	-4.27E-03
μ''_4	1.73E-02	1.59E-02	1.58E-02
μ''_5	7.47E-04	-6.29E-05	2.43E-05
μ''_6	1.13E-03	3.79E-04	3.09E-04
μ''_7	7.15E-04	-3.95E-05	1.03E-05
BE_P	9.64E-01	9.64E-01	9.68E-01
BE_1	-2.62E+00	-2.46E-03	-3.96E+01
BE_2	9.24E-01	9.23E-01	9.32E-01
H_1	2.65E-01	3.07E-01	2.33E-01
H_2	2.45E-01	2.86E-01	2.13E-01
H_3	2.45E-01	2.86E-01	2.13E-01
H_4	2.97E-01	3.20E-01	2.58E-01
H_5	2.66E-01	3.08E-01	2.34E-01

When using SVM several different kernels were employed and it was observed that quadratic kernel provided the best performance. The kernel scale for this classifier was set on automatic. Next, different KNN approaches were used and the weighted KNN method was found to have the best performance; for this classifier, the distance metric was set to Minkowski, distance weight was set to squared inverse, the standardized data was set to true, and 10 number of neighbors were considered. For the DT classifier the maximum number of splits was set to 18 and the split criterion to Gini's diversity index.

As mentioned earlier, several ensemble classifiers were used besides conventional classifiers. Ensembles use a set of different learning algorithms and combine the results using weighted or unweighted voting to obtain higher predictive performance. These methods are usually more accurate than individual constituent classifiers [120, 121].

Table 5.8 Average of the normalized proposed features (Complex H 1-5, Real Bicepstrum H 1-5 and Complex Bicepstrum H 1-5) for each stage of PD.

Features	HC	Early-Stage	Late-Stage
CH ₁	2.91E-01	3.29E-01	2.64E-01
CH ₂	2.80E-01	3.17E-01	2.53E-01
CH ₃	2.81E-01	3.17E-01	2.53E-01
CH ₄	2.48E-01	2.86E-01	2.20E-01
CH ₅	2.91E-01	3.29E-01	2.65E-01
RBH ₁	5.13E-01	5.38E-01	4.91E-01
RBH ₂	3.15E-01	3.55E-01	2.80E-01
RBH ₃	4.45E-01	4.66E-01	4.21E-01
RBH ₄	9.42E-02	1.24E-01	6.88E-02
RBH ₅	5.00E-01	5.17E-01	4.84E-01
CBH ₁	4.72E-01	5.06E-01	4.64E-01
CBH ₂	3.29E-01	3.68E-01	3.02E-01
CBH ₃	4.23E-01	4.48E-01	4.01E-01
CBH ₄	1.04E-01	1.33E-01	8.17E-02
CBH ₅	4.58E-01	4.88E-01	4.57E-01

The first ensemble classifier that was selected for this task was the boosted trees. For this classifier the ensemble method was set to AdaBoost, learner type to DT, maximum number of splits to 100 and number of learners to 100. The next ensemble classifier that was employed is the RUSBoosted trees where the ensemble method was set to RUSBoosted, learner type to DT, maximum number of splits to 200, and number of learners to 200. The final ensemble classifier that was used in this section is the subspace KNN. For this classifier the ensemble method was set to subspace, learner type to KNN, number of learners to 100 and the subspace dimension to 18.

Table 5.9 Performance of the classifiers including the AUC and the overall accuracy.

Classifier	AUC	Accuracy (%)
DT	0.82	77
SVM	0.83	74
KNN	0.82	67
Ensemble Boosted Trees	0.87	82
Ensemble Subspace KNN	0.81	67
Ensemble RUSBoosted Trees	0.90	87

The overall performance of all the classifiers used in this section, including area under the curve (AUC) and overall accuracy, can be seen in Table 5.9. By comparing the performance of the methods, it can be clearly seen that ensemble classifiers had better overall performance.

5.3 Comparison with State-of-the-Art Methods

In the previous section, first PD was classified from HC using five ensemble methods and then the stages of PD were classified using six classifiers, three of which were ensemble methods. Ensemble classifiers achieved better overall performance which is why these methods have been the focus of this study. For PD vs HC, the bagged trees ensemble had the highest performance while for stages of PD RUSBoosted trees achieved the highest accuracy. RUSBoost is a hybrid sampling/boosting algorithm that aims to achieve higher performance for datasets with the class imbalance problem [122]. As it can be seen the second dataset that was used in this study is unbalanced, which is why RUSBoosted tree ensemble classifier was able to achieve the highest accuracy. For the diagnosis of PD by using PCA, certain redundancies within the feature pool were noticed which were removed by using feature ranking and selection. In the classification of stages, all 35 features contributed to the performance. All classifications were performed using the LOTO-CV, LOSO-CV and LTSO-CV methods. The performances shown in Tables 5.4 and 5.5 were achieved through LTSO-CV method, while for Table 5.9 LOTO-CV method was used. Using the LOSO-CV approach for RUSBoosted trees (classifier with the highest accuracy) an average of 0.89 for AUC and 86.4% accuracy (at 20% false positive) was achieved, proving that our approach is independent of the dataset. The combination of the binary classifications of PD vs HC and early vs late stages as described in Figure 4.1, create a fully automated model for diagnosis and identification of the stages of PD.

Table 5.10 Summary of previous studies conducted on automated diagnosis of PD and classification of its stages.

Target	Authors	Features	Classifiers	Accuracy (%)	Task
PD Diagnosis	Han et al. [123] (2013)	α , β , δ and θ band powers and wavelet packet entropies (WPE)	-	-	Investigating abnormalities in early PD.
	Yuvaraj et al. [80] (2016)	Bave, H ₁ -H ₅ , BE ₁ -BE ₂ , BE _p , aW _x , aW _y	DT, KNN, NB, SVM	99.62	Developing an index for classification of PD
	Vanegas et al. [124] (2018)	FFT of visual ERP	Logistic regression, DT, Extra tree	99.81	Classification of PD using spectrum of the steady state visual evoked potential
	Oh et al. [52] (2018)	-	13-layer CNN	88.25	Deep learning approach to classification of PD
	Koch et al. [125] (2019)	FFT coeff, Autocorrelation coeff, Occipital peak frequency, Total beta power, Total peak frequency	RF	91	Automated vs hand crafted features for classification of PD
	Stage-1 (This work)	Features extracted from alpha freq. band: μ_{3-7} , H ₁ -H ₅ , BH ₁ -BH ₅ , BE ₁ -BE ₂ , BE _p , CD, ApEn, LLE, HE	Ensemble DT, Ensemble KNN, Ensemble LDA	99.19	Extracting HOS features from alpha and beta freq. bands for classification of PD
PD Stage Classification	Aldea et al. [101] (2016)	CD, LLE, HE of α , β , δ , θ , γ bands	-	-	Evaluating differences between HC, PD and epileptic subjects with nonlinear features
	Naghsh et al. [16] (2019)	Total power of α , β bands	SVM, Kmeans	95	BSL for classification of stages of PD
	Stage-2 (This work)	H ₁ -H ₅ , CH ₁ -CH ₅ , BH ₁ -BH ₅ , CBH ₁ -CBH ₅ , BE ₁ -BE ₂ , BE _p , Total power of α , β bands, μ_{3-7} , CD, ApEn, LLE, HE	DT, SVM, KNN, Ensemble DT, Ensemble KNN	87	Classification of stages of PD using new HOS features

In Table 5.10, some of the main studies in the classification of PD were compared with the proposed method; it must be noted that the studies mentioned in this table do not use the same datasets, and many of the ones with high accuracy have used ERP based datasets. While some researchers such as Oh et al. [52] have attempted to use deep learning approaches for classification of PD, most researchers have employed traditional machine learning methods. Generally, deep learning methods require a large training dataset to achieve good performance while many biomedical datasets (including EEG) tend to be on the smaller side; combining that with the fact that EEG signals are random in nature, makes deep learning unsuitable for classification of EEG signals. In our previous study [126] it was noticed that conventional machine learning classifiers were outperformed by ensemble methods for EEG classification; this was also supported by many other studies in EEG processing [127, 128]; therefore, ensemble learning is currently the best approach for EEG classification which is why for this study the focus was mostly on ensemble learning for the classification step.

5.4 Discussion

In this study, a hierarchy approach was used as described in Figure 4.1. The first stage of this study is similar to the work of Yuvaraj et al. [80] where the focus is on classification of PD vs HC. The proposed method of using HOS features of alpha rhythm achieved very high accuracy for this stage. The second stage of this research is focused on classifying early-stage vs late-stage PD and together they create the complete model.

This approach creates a more specialized classification and allows us to use features based on the occurrence of each specific symptom. The binary hierarchy classification is able to achieve higher accuracy than is possible through conventional methods. For the first stage, the features were collected from the alpha rhythm achieved the highest accuracy which was not true for the

second stage. In this work, nonlinear and statistical features from both the time and frequency domains were used. The temporal statistical features (such as the higher standardized moments, HE, etc.) carry useful information about the nonstationary and non-Gaussian characteristics of the signal while the features defined in the frequency domain (such as frequency band powers, higher-order entropies and etc.) hold information about the energy (distribution) of the frequency bands of the signal. Using both the temporal and spectral features (in the case of higher-order features, bifrequency) has allowed us to achieve higher accuracy than would be possible with using features from only one domain.

To achieve the goal of this research, the main focus was on the development of new higher-order features. The newly developed 15 higher-order features that are based on existing H_1 – H_5 higher-order features are grouped into three sets. The first set (RBH_1 – RBH_5) was defined in the bicepstrum domain instead of the bispectrum. The cepstral domain's ability to transform temporal convolution to addition allows this domain to approach the features from a new angle. The bicepstral domain goes even further and expands the signal across two quefreny axes (creating the biquefreny domain). The other feature sets (CH_1 – CH_5 and CBH_1 – CBH_5) are very similar to H_1 – H_5 and RBH_1 – RBH_5 , the only difference being that instead of calculating them from the real bispectrum and bicepstrum, they were computed based on the complex values of bispectrum and bicepstrum. Therefore, in these features, the phase is also included. The developed features improved the accuracy of classification in this study; however, they are statistical features that may have applications in other areas as well, therefore more in-depth studies are needed to uncover their applications in other areas.

5.5 Summary

In this chapter, the features used for classification of PD vs HC and early stage vs late stage of PD including second order, existing higher order, and the newly developed higher order features were computed and presented. Concurrently, these features were also extracted from the alpha and beta rhythms and were used to perform the diagnosis, and the performance of the different methods were compared. While for this part of the research, multiple CV approaches were employed, all the accuracies reported in this section were achieved using the LTSO-CV approach. The highest performance achieved for each classification was then compared with current state-of-the-art methods.

Chapter 6: Supervised Classification of Parkinson's Disease Using Shallow and Deep Neural Networks

6.1 Background

6.1.1 Convolutional Neural Network

Evolutionary computation (EC) is a growing field having been in development since at least the 1960's [129]. As the field has grown, machine learning has emerged as an extremely useful subset. Machine learning employs the use of algorithms in series called layers to process data. The most basic machine learning method known as feedforward neural networks perform a series of calculations to quantify the characteristics of a dataset. This artificial neural network can learn and make decisions on its own without human interaction [130]. With the development of deep networks, conventional neural networks with only a few layers became known as shallow networks. CNNs are a marked improvement on this paradigm being one of the best in understanding image content related tasks. CNNs have the ability to utilize spatial correlations within data being a multilayered feedforward hierarchical network [131]. Each layer performs convolution transformations to output features or averages the data to reduce the size and therefore computational requirements [132]. Using these averaging and data manipulation methods deep layered CNNs are able to quickly process large datasets like high resolution images.

While deep networks with as many layers as possible seem to be the trending choice for increasing accuracy and performance, they also require sufficiently large datasets. For certain applications like medical prognoses these datasets are often difficult to obtain either because they require permissions from patients or because they simply do not exist as the population affected is

too small. An unexpected problem arises in performance when applying very deep networks to shallow problems. As depth increases, network accuracy becomes quickly saturated during training. This is indicative that the network has stopped learning and gains no benefit from further data in training. This phenomenon known as overfitting causes low accuracies and adding more layers to a model that is already deep enough to handle the input data will increase training error [133]. Even by copying a shallow network with higher accuracy to the deeper network, accuracy will not improve [134]. The idea follows then that neural network sizes must be custom fit for each application with no one size fits all solution.

6.1.2 Residual Neural Network

In an effort to reduce these difficulties in training, ResNet was proposed by He et al. 2015 [133]. By stacking layers of residual blocks with shortcuts to skip layers 2 at a time, training was incredibly easy. They observed that skipping single or triple layers did not improve performance in the same manner, but rather than following the rule of twos was vital to lowering training and testing errors. It was so successful that they trained a 1001 layer network which was previously near impossible [135]. The original team proposed that residual units hold residual mappings aiding in the ability to represent identity mappings, thus decreasing degradation significantly [136]. Regardless of the reason, the results allow neural networks to be useful in a wider variety of applications. For example, ResNet can be used to predict mental disorders using data pulled from EEG signals. A common problem in EEG data capture is obtaining patient consent and sufficient patients with the disorder for analysis. Having the ability to train with relatively little data, ResNet provides a substantial solution for EEG analysis. Researchers have been able to definitively predict patients with epilepsy before symptoms appeared [137] and diagnose schizophrenia using small data sets from individual lobes of the patients' brains [138]. ResNet

lowering training requirements has been remarkable in removing limitations for neural network application.

6.2 Preprocessing

For this phase of our research, dataset #1 described in section 4.3 was used and the same preprocessing steps for removing artifacts and the BSS phase were applied to the data. The data was then segmented into 1000 sample segments where each segment represents 2 seconds of data. Unlike before there is no feature extraction and all channels of the signal are used at the same time. Therefore, the inputs of the networks are multi-channel 2D time domain signals with the dimensions of 64×1000 where each row represents one ICA component.

6.3 Classification Performance

The focus of this work is to compare the performance of the two CNN and ResNet approach for the classification of PD vs HC from the EEG signals. To perform this task, different sizes for each network were employed to perform the classification. Since the dataset used in this study is comparatively small, networks with a lower number of layers achieved better overall performance.

Table 6.1 The highest overall test accuracy of the 6-layer networks.

Model	Overall Accuracy (%)
6-Layer CNN	92.26
6-Layer ResNet with 1 Skip Layer	71

To investigate the sizes of the network and their effect on performance, 18 was chosen as the number of layers for both networks since that is the smallest prebuilt ResNet network. While the CNN model achieved some low classification accuracy (<70%), the ResNet would quickly get overfitted and was not able to produce any meaningful results. This is why the size of the networks

and the number of layers were reduced until the highest performance was achieved using 6-layer networks.

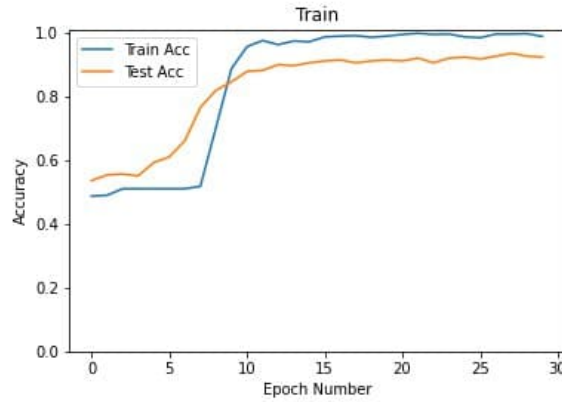


Figure 6.1 The train and test accuracies of the 6-layer shallow CNN.

Figure 6.1 shows the validation accuracy of the 6-layer shallow CNN over 30 epochs and Figure 6.2 shows the loss of this network. In both figures, the graphs for the train set and test set can be seen. Similarly, Figures 6.3 and 6.4 show the validation accuracy and loss of the train set and test set of the ResNet. Table 6.1 shows the overall accuracies of the two networks using the LOTO-CV approach.

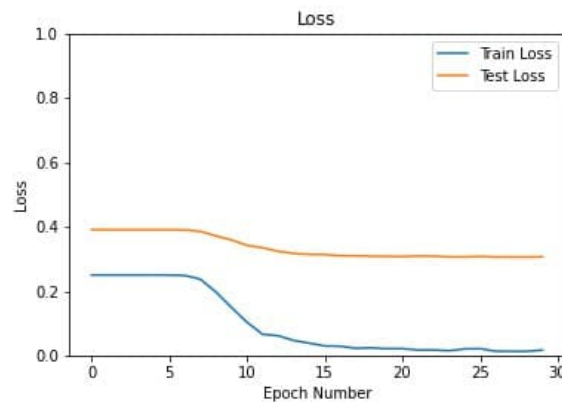


Figure 6.2 The train and test loss of the 6-layer shallow CNN.

The ResNets usually require large datasets that are absent in most biomedical applications which is why in this study, the shallow CNN approach outperformed the ResNet. Also, the idea behind ResNets is built on a large number of layers and it was observed that for smaller networks, they fail to provide acceptable results.

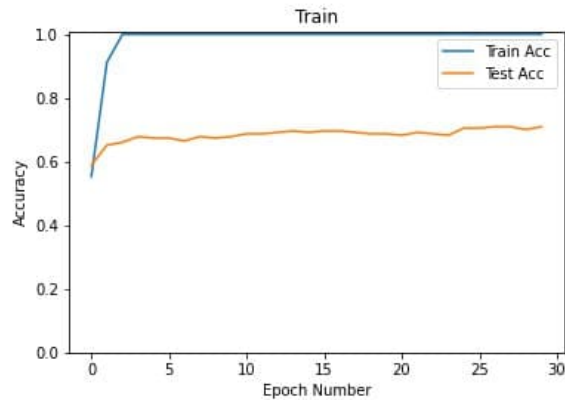


Figure 6.3 The train and test accuracies of the 6-layer shallow ResNet.

It can be seen that in ResNet, the training stops after only two or three epochs where for CNN it continues until around 25. This is why the test loss shown in Figure 6.2 continues to decrease whereas the loss shown in Figure 6.4 stops decreasing after the 2nd or 3rd epoch.

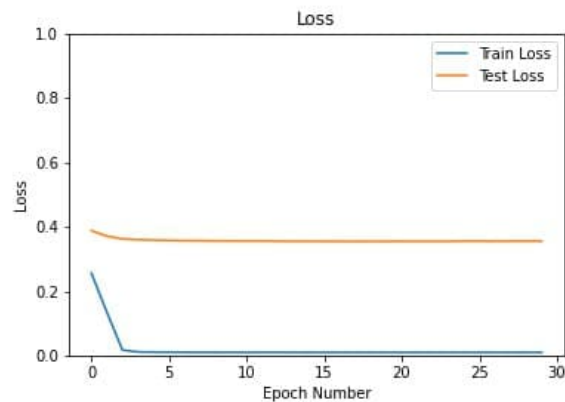


Figure 6.4 The train and test loss of the 6-layer shallow ResNet.

Figure 6.5 shows the diagram of the 6-layer CNN model that was used in this section. It includes details such as the size of the input data, the number and type of layers, activation function, pooling method, etc. This model includes, 4 convolutional layers with one input and output fully connected layers (6 in total).

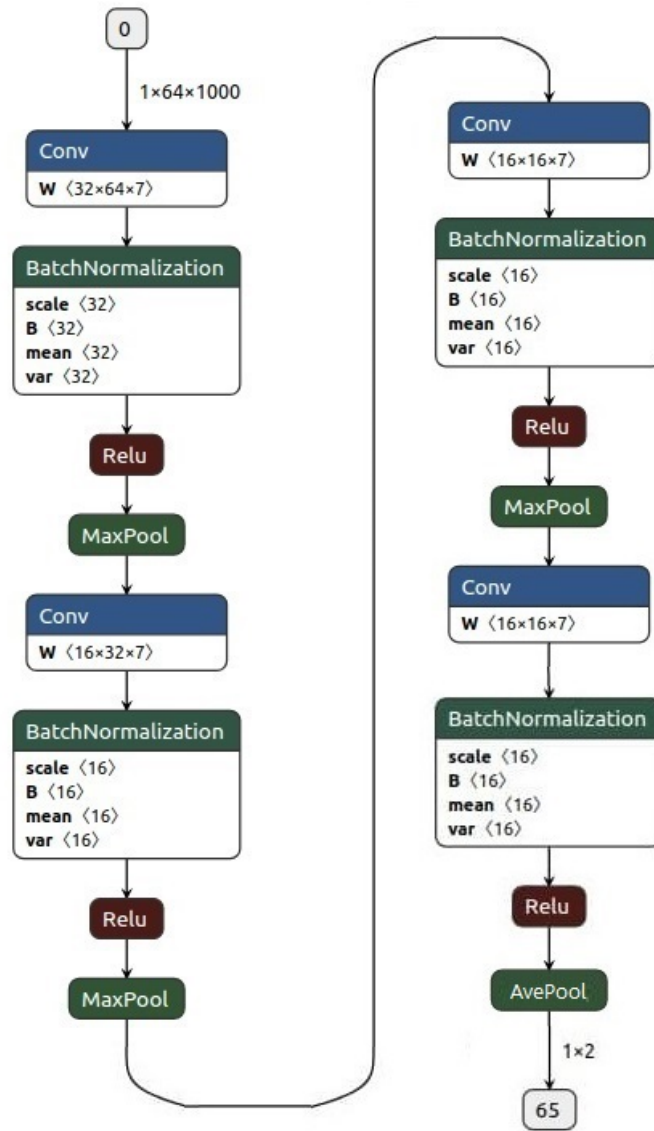


Figure 6.5 The proposed one dimensional 6-layer shallow CNN.

Overall, it appears that CNN can be used for EEG classification and would be able to achieve high performance while ResNet does not seem to be a particularly good match for this

type of signal. Obviously, the size of the network and the size of the dataset can have a huge effect on the performance, but since the datasets used in biomedical applications are generally small, shallow CNN methods seem to be the current best option.

6.4 Comparison with the Proposed Method

The results achieved by deep learning models were significantly lower than what was achieved using HOS features of alpha rhythm and ensemble learning. While using the LTSO-CV method, over 99 % accuracy was achieved. However, the accuracy reduced to 92 % while using LOTO-CV, which has much more flexibility. This means that the results are vastly different and the deep learning methods used here were not able to compete with the proposed methods. There are two main reasons as to why the proposed method outperformed both the CNN and the ResNet models. The first reason is due to the random nature of EEG signals. Therefore, networks that try to look for patterns within the data do not achieve good results for this type of data which is the main reason why the focus was on statistical analysis in this work. The second reason is the size of the dataset; deep learning methods require large datasets that usually do not exist in biomedical signal processing. To continue this research, one approach could be to focus on reinforcement learning which looks for long-term dependencies between the samples and is usually a good option for analysis of time domain data. The second approach could be to transform the dataset to a different domain before using deep learning methods to avoid the first problem. This issue will be discussed further in the future works section of the next chapter.

6.5 Summary

Over the past few years, deep learning algorithms including CNN and ResNet have gained the attention of many researchers and have been used for many different applications. Previously, for the proposed method, the classifiers that were employed were chosen from conventional and

ensemble methods. In this research, CNN and ResNet algorithms were used for the classification of PD vs HC and their results were compared with the proposed model. It was observed that while deep learning methods are able to achieve acceptable results, their accuracy is still considerably lower than the proposed method. This goes to show that while deep learning approaches are great machine learning tool, they might not be the best fit for every application.

Chapter 7: Conclusion and Future Research

7.1 Conclusion

The diagnosis of neurodegenerative diseases has been studied for several decades. The methods that have been developed to aid the neurologists in the process are mostly based on the changes in biomedical signals caused by the symptoms of the diseases; therefore, they rely on the existence of the symptoms which in many cases occur after the initial phases. Using biomedical signals collected from the brain and using the source of the disease for classification has the potential for very early diagnosis. Among different methods for monitoring brain activity, EEG is the most popular due to its noninvasiveness and low cost.

In this research, a hierarchical approach was taken using the EEG signals of the patients to diagnose PD patients from HC and then classification of stages of PD after a positive diagnosis was made. To improve the accuracy and the overall performance, a detailed HOS analysis of the EEG signals was performed and multiple classification approaches were compared. In the HOS analysis phase of this work, prior features used for this application and similar applications were analyzed and from them new HOS features were developed.

In this study, 15 new HOS features were developed to improve the performance of the model based on the existing H_1 - H_5 higher order features. The new features are calculated based on the complex bispectrum along with real and complex bicepstrum (instead of real bispectrum). The CH_1 - CH_5 are calculated similar to H_1 - H_5 but based on the complex bispectrum; therefore, the phase is also included in these features. Similarly, the RBH_1 - RBH_5 features were calculated based on the real bicepstrum and the CBH_1 - CBH_5 features were calculated based on the complex bicepstrum of

the signals. There are very few features that have been defined in the bispectral domain and the characteristics of this domain have mostly been left undiscovered. The developed features improved the accuracy of classification in this study; however, they are statistical features that may have applications in other areas as well, therefore more in-depth studies are needed to uncover their applications in other areas.

The performance of the classifiers confirms the advantages of using HOS features, especially the developed bicepstral RBH features in the classification of PD; also, it can be seen that they have the highest performance when they are extracted from the alpha rhythm which proves our assumption that not all the frequencies of the EEG signals contribute to the classification of PD. The other frequency components act as unwanted signals that can reduce the classification performance. As it was mentioned, several studies have pointed to the irregularity of alpha rhythm in PD patients [99, 100]; in this study, it was shown that those changes can be quantified by using HOS features where they can be used for the classification of PD. The existing features that were selected for this study have all been previously used for the classification of EEG signals either for diagnosis of PD or other diseases.

It was observed that the performance of the classification using the HOS features from alpha rhythm was higher for male subjects compared to females, where the algorithm was able to classify all data points accurately for 11 of the 18 male test pairs compared to 2 out of 10 female test pairs. According to a study conducted by Barry et al. [139] in the alpha band, the spatial and temporal correlation is higher in male subjects compared to female subjects, which explains why for the alpha band, the proposed algorithm had a higher performance for male subjects. The results achieved from the beta band were different, where the classifiers were able to classify all data points accurately for 9 out of 10 female test pairs compared to 13 out of 18 for male test pairs.

While the Bagged DT ensemble had the highest performance for the alpha band, AdaBoost DT ensemble had the highest performance for all bands and beta band. By comparing the proposed method with the current state-of-the-art methods, it was observed that for this application HOS analysis of the alpha band yields better results compared to other methods. In the classification of the stages of PD, 6 classifiers were employed, 3 of which were ensemble classifiers. The KNN both as an individual classifier and the ensemble gave the lowest performance compared to other methods. SVM as a single classifier had higher overall performance compared to DT; however, the DT ensembles were able to achieve much higher performance. The highest performance was achieved by RUSBoosted trees ensemble classifier. RUSBoost is a hybrid sampling/boosting algorithm that aims to achieve higher performance for datasets with the class imbalance problem [122].

One important issue regarding the previous studies conducted on applications of HOS analysis of EEG signals is that due to mathematical and computational complexity, there is little focus on the features themselves and the rationale behind the higher order features are rarely investigated individually. In this study, a detailed investigation of each of the features was conducted which had been missing in previous works. It should also be noted that HOS analysis has several main disadvantages; besides the computational complexity, increasing the order of the spectrum will increase the dimension. This new dimension will help us in finding new features, however by adding a dimension to the spectrum the computational burden will increase drastically. Therefore, an initial analysis of lower order statistics along with a detailed analysis of each HOS feature is crucial for this task. In this research, temporal, spectral, and cepstral statistical analyses of different orders, not only higher orders but also the lower orders were performed and each of the proposed HOS features was carefully analyzed before being used in the proposed model. This

is one of the main reasons why the proposed model was able to achieve very high performance compared to other state-of-the-art methods.

7.2 Future Research

One of the main issues in many areas of biomedical signal processing is the limited access to data which usually results in smaller datasets. There is currently no standardized dataset for EEG of neurodegenerative diseases, which means the experiments are conducted on different datasets; this, in turn, reduces the efficacy and reliability of comparisons between methods. Creating a unified dataset that can be used to compare methods would greatly improve the accuracy and credibility of methods developed in this area.

As it was mentioned, one of the advantages of the proposed method is the ability to be expanded to other neurodegenerative diseases. Similar to PD, irregularities have been observed in several other neurodegenerative diseases. Thus, investigating similar properties for other neurodegenerative diseases allow this model to be expanded to other neurodegenerative diseases.

7.2.1 Higher Order Statistic in Electroencephalography

The general interest in HOS analysis has increased over the past 20 years and yet many applications of this field are yet to be discovered. As it was mentioned the application of HOS analysis in EEG signal processing can be divided into three main categories. In the first category, one of the most beneficial algorithms is WICA. This has been used for the removal of various artifacts in EEG in different research studies. One of the possible future developments in this category could be to gather this information to create a single algorithm for detecting and removing all types of EEG artifacts by modifying WICA or similar algorithms. The algorithm can be optimized before adding to EEG toolboxes for public use. The second category has the least number of studies available in the literature but may have the potential for the most impact in this

field. There are not that many algorithms developed using HOS. However, these algorithms have been widely used in almost every area of signal processing. One of the possible directions of this field in the future could be in developing better measures of non-Gaussianity or more robust algorithms for BSS. The rapid advancement of the field of computer science in machine learning coupled with the development of more innovative and robust feature selection methods will increase the research studies in the third category; this will result in more reliable diagnoses of neurological diseases and disorders.

7.2.2 Other Applications of Higher Order Statistics

EEG signals are not the only type of biomedical signals that have been affected by HOS analysis; electrocardiography, surface electromyogram (sEMG), lung sound, heart sound, bowel sounds, somatosensory, different medical imaging, etc. [140]. With the development of more capable processing units, new applications for HOS analysis are being discovered. One of the recent applications of HOS analysis is in speech signal processing where AlBadawy et al. [141] have used bispectral analysis to distinguish AI synthesized speech vs real human speech.

Furthermore, biomedical signal processing is not the only field that has been affected. HOS has applications in many various areas such as lidar, radar, laser, plasma physics, seismic data processing, image processing, harmonic retrieval, time-delay estimation, adaptive filtering, array processing, blind equalization, etc. [142].

References

- [1] D. P. Subha, P. K. Joseph, R. Acharya U, and C. M. Lim, "EEG Signal Analysis: A Survey," *Journal of Medical Systems*, vol. 34, no. 2, pp. 195-212, 2010/04/01 2010.
- [2] M. R. Lakshmi, T. Prasad, and D. V. C. Prakash, "Survey on EEG signal processing methods," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 1, 2014.
- [3] C. S. Nayak and A. C. Anilkumar, "Eeg normal waveforms," *StatPearls*, 2019.
- [4] M. D. Binder, N. Hirokawa, and U. Windhorst, *Encyclopedia of neuroscience*. Springer Berlin, Germany, 2009.
- [5] A. Yousefiankalareh, M. A. Manoochehri, S. A. Khoshnevis, and S. R. Surakanti, "Improve the Frequency Identification in SSVEP based BCI Systems with Moving Windows Algorithm," in *2019 6th International Conference on Control, Instrumentation and Automation (ICCIA)*, 2019, pp. 1-5: IEEE.
- [6] S. A. Khoshnevis and R. Sankar, "Applications of Higher Order Statistics in Electroencephalography Signal Processing: A Comprehensive Survey," *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 169-183, 2020.
- [7] H.-J. Hwang, S. Kim, S. Choi, and C.-H. Im, "EEG-Based Brain-Computer Interfaces: A Thorough Literature Survey," *International Journal of Human-Computer Interaction*, vol. 29, no. 12, pp. 814-826, 2013/12/02 2013.
- [8] W. Klonowski, "Everything you wanted to ask about EEG but were afraid to get the right answer," *Nonlinear biomedical physics*, vol. 3, no. 1, pp. 1-5, 2009.
- [9] L. Bi, X.-A. Fan, and Y. Liu, "EEG-Based Brain-Controlled Mobile Robots: A Survey," *IEEE Transactions on human-machine systems*, vol. 43, no. 2, pp. 161-176, 2013.
- [10] D. Hurley, "Implanted Endovascular Device Provides ALS Patients With Independent Access to Computer Navigation," *Neurology Today*, vol. 20, no. 24, pp. 8-10, 2020.
- [11] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," (in eng), *Electroencephalogr Clin Neurophysiol*, vol. 70, no. 6, pp. 510-23, Dec 1988.
- [12] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "Single trial independent component analysis for P300 BCI system," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 4035-4038.
- [13] K. Li, V. Narayan Raju, R. Sankar, Y. Arbel, and E. Donchin, "Advances and Challenges in Signal Analysis for Single Trial P300-BCI," Berlin, Heidelberg, 2011, pp. 87-94: Springer Berlin Heidelberg.
- [14] K. Li, R. Sankar, K. Cao, Y. Arbel, and E. Donchin, "A new single trial P300 classification method," *International Journal of E-Health and Medical Communications (IJEHMC)*, vol. 3, no. 4, pp. 31-41, 2012.
- [15] R. Zerafa, T. Camilleri, O. Falzon, and K. P. Camilleri, "To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs," *Journal of Neural Engineering*, vol. 15, no. 5, p. 051001, 2018/07/02 2018.

- [16] A. J. Doud, J. P. Lucas, M. T. Pisansky, and B. He, "Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface," *PloS one*, vol. 6, no. 10, p. e26322, 2011.
- [17] S. Taheri, M. Ezoji, and S. M. Sakhaei, "Convolutional neural network based features for motor imagery EEG signals classification in brain-computer interface system," *SN Applied Sciences*, vol. 2, no. 4, p. 555, 2020/03/05 2020.
- [18] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41-56, 2007.
- [19] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814-825, 2011.
- [20] B. Litt and J. Echauz, "Prediction of epileptic seizures," *The Lancet Neurology*, vol. 1, no. 1, pp. 22-30, 2002.
- [21] M. K. Kıymık, İ. Güler, A. Dizibüyük, and M. Akın, "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," *Computers in biology and medicine*, vol. 35, no. 7, pp. 603-616, 2005.
- [22] H. A. Haider *et al.*, "Sensitivity of quantitative EEG for seizure identification in the intensive care unit," *Neurology*, vol. 87, no. 9, pp. 935-944, 2016.
- [23] K. Fu, J. Qu, Y. Chai, and T. Zou, "Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals," *Biomedical Signal Processing and Control*, vol. 18, pp. 179-185, 2015.
- [24] B. Swiderski, S. Osowski, and A. Rysz, "Lyapunov exponent of EEG signal for epileptic seizure characterization," in *Proceedings of the 2005 European Conference on Circuit Theory and Design, 2005.*, 2005, vol. 2, pp. II/153-II/156 vol. 2: IEEE.
- [25] X. S. Zhang, R. J. Roy, and E. W. Jensen, "EEG complexity as a measure of depth of anesthesia for patients," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 12, pp. 1424-1433, 2001.
- [26] Z. Liang *et al.*, "EEG entropy measures in anesthesia," *Frontiers in computational neuroscience*, vol. 9, p. 16, 2015.
- [27] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Frontiers in neuroscience*, vol. 12, p. 162, 2018.
- [28] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532-541, 2018.
- [29] S. V. Perumal and R. Sankar, "Gait monitoring system for patients with Parkinson's disease using wearable sensors," in *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, 2016, pp. 21-24: IEEE.
- [30] S. A. Khoshnevis, S. B. Appakaya, E. Sheybani, and R. Sankar, "Compression of Gait IMU signals Using Sensor Fusion and Compressive Sensing," in *2020 Wireless Telecommunications Symposium (WTS)*, 2020, pp. 1-5: IEEE.
- [31] S. V. Perumal and R. Sankar, "Gait and tremor assessment for patients with Parkinson's disease using wearable sensors," *ICT Express*, vol. 2, no. 4, pp. 168-174, 2016/12/01/ 2016.
- [32] A. B. Oktay and A. Kocer, "Differential diagnosis of Parkinson and essential tremor with convolutional LSTM networks," *Biomedical Signal Processing and Control*, vol. 56, p. 101683, 2020.

- [33] S. B. Appakaya and R. Sankar, "Parkinson's Disease Classification using Pitch Synchronous Speech Segments and Fine Gaussian Kernels based SVM," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 236-239: IEEE.
- [34] S. B. Appakaya and R. Sankar, "Classification of Parkinson's disease Using Pitch Synchronous Speech Analysis," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1420-1423: IEEE.
- [35] F. L. Pagan, "Improving outcomes through early diagnosis of Parkinson's disease," *The American journal of managed care*, vol. 18, no. 7 Suppl, pp. S176-82, 2012.
- [36] M. Mott and W. Koroshetz, "Bridging the gap in neurotherapeutic discovery and development: the role of the National Institute of Neurological Disorders and Stroke in translational neuroscience," *Neurotherapeutics*, vol. 12, no. 3, pp. 651-654, 2015.
- [37] C. Marras *et al.*, "Prevalence of Parkinson's disease across North America," *npj Parkinson's Disease*, vol. 4, no. 1, p. 21, 2018/07/10 2018.
- [38] (2019). *Parkinson's Foundation Announces State Level Prevalence Medicare Costs of Parkinson's Disease*. Available: <https://www.parkinson.org/about-us/Press-Room/Press-Releases/Parkinsons-Foundation-Announces-State-Level-Prevalence-Medicare-Costs-of-Parkinsons-Disease>
- [39] W. H. Organization, *Neurological disorders: public health challenges*. World Health Organization, 2006.
- [40] D. J. Surmeier, J. A. Obeso, and G. M. Halliday, "Selective neuronal vulnerability in Parkinson disease," *Nature reviews Neuroscience*, vol. 18, no. 2, p. 101, 2017.
- [41] M. Hallett and W. Poewe, *Therapeutics of Parkinson's disease and other movement disorders*. John Wiley & Sons, 2008.
- [42] J. C. Bridi and F. Hirth, "Mechanisms of α -synuclein induced synaptopathy in Parkinson's disease," *Frontiers in neuroscience*, vol. 12, p. 80, 2018.
- [43] M. D. S. T. F. o. R. S. f. P. s. Disease, "The unified Parkinson's disease rating scale (UPDRS): status and recommendations," *Movement Disorders*, vol. 18, no. 7, pp. 738-750, 2003.
- [44] C. G. Goetz *et al.*, "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease," *Movement disorders*, vol. 19, no. 9, pp. 1020-1028, 2004.
- [45] L. Wong and W. Abdulla, "Time-frequency evaluation of segmentation methods for neonatal EEG signals," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, 2006.
- [46] N. Schlede *et al.*, "Clinical EEG in cognitively impaired patients with Parkinson's Disease," *Journal of the Neurological Sciences*, vol. 310, no. 1, pp. 75-78, 2011/11/15/ 2011.
- [47] N. Swann *et al.*, "Deep Brain Stimulation of the Subthalamic Nucleus Alters the Cortical Profile of Response Inhibition in the Beta Frequency Band: A Scalp EEG Study in Parkinson's Disease," *The Journal of Neuroscience*, vol. 31, no. 15, p. 5721, 2011.
- [48] B. T. Klassen *et al.*, "Quantitative EEG as a predictive biomarker for Parkinson disease dementia," *Neurology*, vol. 77, no. 2, p. 118, 2011.
- [49] R. Yuvaraj *et al.*, "On the analysis of EEG power, frequency and asymmetry in Parkinson's disease during emotion processing," *Behavioral and Brain Functions*, vol. 10, no. 1, p. 12, 2014/04/09 2014.

- [50] G. Liu *et al.*, "Complexity analysis of electroencephalogram dynamics in patients with Parkinson's disease," *Parkinson's Disease*, vol. 2017, 2017.
- [51] S. Sakhavi, C. Guan, and S. Yan, "Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5619-5629, 2018.
- [52] S. L. Oh *et al.*, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Computing and Applications*, 2018/08/30 2018.
- [53] Y. V. Obukhov *et al.*, "Electroencephalograms features of the early stage Parkinson's disease," *Pattern recognition and image analysis*, vol. 24, no. 4, pp. 593-604, 2014.
- [54] E. Naghsh, M. F. Sabahi, and S. Beheshti, "Spatial analysis of EEG signals for Parkinson's disease stage detection," *Signal, Image and Video Processing*, pp. 1-9, 2019.
- [55] M. Raghuvver and C. Nikias, "Bispectrum estimation: A parametric approach," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 1213-1230, 1985.
- [56] C. L. Nikias and J. M. Mendel, "Signal processing with higher-order spectra," *IEEE Signal processing magazine*, vol. 10, no. 3, pp. 10-37, 1993.
- [57] C. L. Nikias and M. R. Raghuvver, "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE*, vol. 75, no. 7, pp. 869-891, 1987.
- [58] M. J. Hinich, "Testing for Gaussianity and linearity of a stationary time series," *Journal of time series analysis*, vol. 3, no. 3, pp. 169-176, 1982.
- [59] S. A. Hosseini, M. A. Khalilzadeh, M. B. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of EEG signals in emotional stress states," in *2010 Second International Conference on Information Technology and Computer Science*, 2010, pp. 60-63: IEEE.
- [60] S. J. Orfanidis, *Optimum signal processing: an introduction*. Macmillan publishing company, 1988.
- [61] W. Kiciński and A. Szczepański, "Quadratic phase coupling phenomenon and its properties," *Hydroacoustics*, vol. 7, pp. 97-106, 2004.
- [62] T. Inouye *et al.*, "Quantification of EEG irregularity by use of the entropy of the power spectrum," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 3, pp. 204-210, 1991.
- [63] P. Venkatakrisnan, R. Sukanesh, and S. Sangeetha, "Detection of quadratic phase coupling from human EEG signals using higher order statistics and spectra," *Signal, Image and Video Processing*, vol. 5, no. 2, pp. 217-229, 2011.
- [64] K. C. Chua, V. Chandran, R. Acharya, and C. M. Lim, "Higher Order Spectral (HOS) Analysis Of Epileptic EEG Signals," in *29th Annual International Conference of the IEEE*, 2007.
- [65] K. C. Chua, V. Chandran, U. Rajendra Acharya, and C. M. Lim, "Analysis of epileptic EEG signals using higher order spectra," *Journal of Medical Engineering & Technology*, vol. 33, no. 1, pp. 42-50, 2009.
- [66] U. R. Acharya, S. Vinitha Sree, and J. S. Suri, "Automatic Detection Of Epileptic EEG Signals Using Higher Order Cumulant Features," *International Journal of Neural Systems*, vol. 21, no. 5, pp. 403-414, 2011.
- [67] U. R. Acharya *et al.*, "Automated Diagnosis Of Epilepsy Using CWT, HOS And Texture Parameters," *International Journal of Neural Systems*, vol. 23, no. 3, 2013.

- [68] X. Du, S. Dua, R. U. Acharya, and C. K. Chua, "Classification of Epilepsy Using High-Order Spectra Features and Principle Component Analysis," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1731-1743, 2012.
- [69] S.-M. Zhou, J. Q. Gan, and F. Sepulved, "Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface," *Information Sciences*, vol. 178, no. 6, pp. 1629-1640, 2008.
- [70] S. A. Hosseini, M. A. Khalilzadeh, and M. B. Naghibi-Sistani, "Higher Order Spectra Analysis of EEG Signals in Emotional Stress States," in *Mohammad Bagher*, Kiev, 2010.
- [71] H. Ghandeharion and A. Erfanian, "A fully automatic ocular artifact suppression from EEG data using higher order statistics: Improved performance by wavelet analysis," *Medical Engineering and Physics*, vol. 32, no. 7, pp. 720–729 2010.
- [72] U. R. Acharya, E. C.-P. Chua, K. C. Chua, L. C. Min, and T. Tamura, "Analysis and Automatic Identification of Sleep Stages Using Higher Order Spectra," *International Journal of Neural Systems* vol. 20, no. 6, pp. 509–521, 2010.
- [73] S. Javidi, D. P. Mandic, C. C. Took, and A. Cichocki, "Kurtosis-based blind source extraction of complex non-circular signals with application in EEG artifact removal in real-time," *Frontiers in Neuroscience*, vol. 5, 2011.
- [74] A. Lay-Ekuakille *et al.*, "Multidimensional analysis of EEG features using advanced spectral estimates for diagnosis accuracy," in *IEEE International Symposium on Medical Measurements and Applications*, Gatineau, 2013.
- [75] S. M. Shafiul Alam and M. I. H. Bhuiyan, "Detection of Seizure and Epilepsy Using Higher Order Statistics in the EMD Domain," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 312 - 318, 2013.
- [76] R. Yuvaraj *et al.*, "Emotion classification in Parkinson's disease by higher-order spectra and power spectrum features using EEG signals: A comparative study," *Journal of Integrative Neuroscience*, vol. 13, no. 1, pp. 89-120, 2014.
- [77] R. Mahajan and B. I. Morshed, "Unsupervised Eye Blink Artifact Denoising of EEG Data with Modified Multiscale Sample Entropy, Kurtosis, and Wavelet-ICA," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 158 - 165, 2015.
- [78] G. Wang, S. J. Shepherd, C. B. Beggs, N. Rao, and Y. Zhang, "The use of kurtosis denoising for EEG analysis of patients suffering from Alzheimer's disease," in *4th International Conference on Biomedical Engineering and Biotechnology*, Shanghai, 2015.
- [79] R. Yuvaraj *et al.*, "Detection of emotions in Parkinson's disease using higher order spectral features from brain's electrical activity," *Biomedical Signal Processing and Control*, vol. 14, pp. 108-116, 2014/11/01/ 2014.
- [80] R. Yuvaraj, U. R. Acharya, and Y. Hagiwara, "A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals," *Neural Computing and Applications*, vol. 30, no. 4, pp. 1225-1235, 2016.
- [81] S. L. Oh *et al.*, "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Computing and Applications*, pp. 1–7, 2018.
- [82] G. M. Bairy *et al.*, "Automated Diagnosis of Depression Electroencephalograph Signals Using Linear Prediction Coding and Higher Order Spectra Features," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 8, pp. 1857-1862, 2017.
- [83] S. A. Hosseini, "A Hybrid Approach Based on Higher Order Spectra for Clinical Recognition of Seizure and Epilepsy," *Basic and Clinical Neurosciense*, vol. 8, no. 6, pp. 479-492, 2017.

- [84] S. Ikeda, R. Ishii, L. Canuet, and R. D. Pascual-Marqui, "Source estimation of epileptic activity using eLORETA kurtosis analysis," *BMJ Case Reports*, 2017.
- [85] G. Inuso, F. LaForesta, N. Mammone, and F. C. Morabito, "Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings," in *International Joint Conference on Neural Networks*, Orlando, 2007.
- [86] W. D. Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical Correlation Analysis Applied to Remove Muscle Artifacts From the Electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2583-2587, 2006.
- [87] M. Bagheri Hamaneh, N. Chitravas, K. Kaiboriboon, S. D. Lhatoo, and K. A. Loparo, "Automated Removal of EKG Artifact From EEG Data Using Independent Component Analysis and Continuous Wavelet Transformation," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1634-1641, 2014.
- [88] J. Urigüen and B. Garcia-Zapirain, "EEG artifact removal-state-of-the-art and guidelines.," *Journal of Neural Engineering*, vol. 12, no. 3, 2015.
- [89] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.
- [90] J. Y. Lee and A. K. Nandi, "Blind Deconvolution of Impacting Signals Using Higher-Order Statistics," *Mechanical Systems and Signal Processing*, vol. 12, no. 2, pp. 357-371, 1998.
- [91] K. O. Bowman and L. R. Shenton, "Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 ," *Biometrika*, vol. 62, no. 2, pp. 243-250, 1975.
- [92] F. J. Anscombe and W. J. Glynn, "Distribution of the kurtosis statistic b_2 for normal statistics," *Biometrika*, vol. 70, no. 1, pp. 227-234, 1983.
- [93] Giannakis, G. B., and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 360-377, 1989.
- [94] J. P. Nadal and N. Parga, "Sensory Coding: Information Maximization and Redundancy Reduction," in *Neuronal Information Processing: From Biological Data to Modelling and Application* Singapore: World Scientific, 1999, pp. 164-171.
- [95] F. Oveisi, "EEG signal classification using nonlinear independent component analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, 2009.
- [96] T. Ning and N. Trinh, "Cross-spectral and cross-bispectral analysis of hippocampal EEG with ICA processing," in *Proceedings of the IEEE 30th Annual Northeast Bioengineering Conference*, Springfield, 2004.
- [97] W. Lu and J. C. Rajapakse, "ICA with Reference," *Neurocomputing*, vol. 69, no. 16-18, pp. 2244-2257, 2006.
- [98] S. Makeig, T.-P. Jung, A. J. Bell, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences*, vol. 94, no. 20, pp. 10979-10984, 1997.
- [99] J. Wan, G. Yi, and J. Wang, "EEG Sub-band Abnormality of Early-stage Parkinson's Disease with Mild Cognitive Impairment," in *2020 39th Chinese Control Conference (CCC)*, 2020, pp. 2856-2861: IEEE.
- [100] N. H. Yılmaz, P. Çalışoğlu, B. Güntekin, and L. Hanoğlu, "Correlation between alpha activity and neuropsychometric tests in Parkinson's disease," *Neuroscience Letters*, vol. 738, p. 135346, 2020.


- [101] R. T. Aldea, O. Geman, I. Chiuchisan, and A. M. Lazar, "A comparison between healthy and neurological disorders patients using nonlinear dynamic tools," in *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)*, 2016, pp. 299-303: IEEE.
- [102] N. Jackson, S. R. Cole, B. Voytek, and N. C. Swann, "Characteristics of waveform shape in Parkinson's Disease Detected with Scalp Electroencephalography," *Eneuro*, vol. 6, no. 3, 2019.
- [103] J. F. Cavanagh, P. Kumar, A. A. Mueller, S. P. Richardson, and A. Mueen, "Diminished EEG habituation to novel events effectively classifies Parkinson's patients," *Clinical Neurophysiology*, vol. 129, no. 2, pp. 409-418, 2018.
- [104] K. Yasoda, R. Ponmagal, K. Bhuvaneshwari, and K. Venkatachalam, "Automatic detection and classification of EEG artifacts using fuzzy kernel SVM and wavelet ICA (WICA)," *Soft Computing*, vol. 24, no. 21, pp. 16011-16019, 2020.
- [105] A. Hekmatmanesh, R. M. Asl, H. Wu, and H. Handroos, "EEG control of a bionic hand with imagination based on chaotic approximation of largest Lyapunov exponent: A single trial BCI application study," *IEEE Access*, vol. 7, pp. 105041-105053, 2019.
- [106] A. Hekmatmanesh, R. M. Asl, H. Handroos, and H. Wu, "Optimizing largest lyapunov exponent utilizing an intelligent water drop algorithm: A brain computer interface study," in *2019 5th International Conference on Event-Based Control, Communication, and Signal Processing (EBCCSP)*, 2019, pp. 1-5: IEEE.
- [107] I. E. Kutepov *et al.*, "EEG analysis in patients with schizophrenia based on Lyapunov exponents," *Informatics in Medicine Unlocked*, vol. 18, p. 100289, 2020.
- [108] S. Khoshnoud, M. A. Nazari, and M. Shamsi, "Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals," *Journal of integrative neuroscience*, vol. 17, no. 1, pp. 17-30, 2018.
- [109] S. Madan, K. Srivastava, A. Sharmila, and P. Mahalakshmi, "A case study on Discrete Wavelet Transform based Hurst exponent for epilepsy detection," *Journal of medical engineering & technology*, vol. 42, no. 1, pp. 9-17, 2018.
- [110] C. W. Yean *et al.*, "Emotional states analyze from scaling properties of EEG signals using hurst exponent for stroke and normal groups," in *Symposium on Intelligent Manufacturing and Mechatronics*, 2019, pp. 526-534: Springer.
- [111] S. Lahmiri, "Generalized Hurst exponent estimates differentiate EEG signals of healthy and epileptic patients," *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 378-385, 2018.
- [112] N. Ji, L. Ma, H. Dong, and X. Zhang, "EEG signals feature extraction based on DWT and EMD combined with approximate entropy," *Brain sciences*, vol. 9, no. 8, p. 201, 2019.
- [113] J. Ramakrishnan and B. R. Kanagaraj, "Analysis of non-seizure and seizure activity using intracranial EEG signals and empirical mode decomposition based approximate entropy," *Biomedical Research (0970-938X)*, 2018.
- [114] A. Sharmila, S. Aman Raj, P. Shashank, and P. Mahalakshmi, "Epileptic seizure detection using DWT-based approximate entropy, Shannon entropy and support vector machine: a case study," *Journal of medical engineering & technology*, vol. 42, no. 1, pp. 1-8, 2018.
- [115] F. Hasanzadeh, M. Mohebbi, and R. Rostami, "Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal," *Journal of affective disorders*, vol. 256, pp. 132-142, 2019.

- [116] D. Abdolzadegan, M. H. Moattar, and M. Ghoshuni, "A robust method for early diagnosis of autism spectrum disorder from EEG signals based on feature selection and DBSCAN method," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 482-493, 2020.
- [117] R. Yuvaraj, U. Rajendra Acharya, and Y. Hagiwara, "A novel Parkinson's Disease Diagnosis Index using higher-order spectra features in EEG signals," *Neural Computing and Applications*, vol. 30, no. 4, pp. 1225-1235, 2018/08/01 2018.
- [118] S. Sun, C. Zhang, and D. Zhang, "An experimental evaluation of ensemble methods for EEG signal classification," *Pattern Recognition Letters*, vol. 28, no. 15, pp. 2157-2163, 2007.
- [119] J. Lepping, "Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery," 2018.
- [120] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, 2000, pp. 1-15: Springer.
- [121] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of artificial intelligence research*, vol. 11, pp. 169-198, 1999.
- [122] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 185-197, 2009.
- [123] C.-X. Han, J. Wang, G.-S. Yi, and Y.-Q. Che, "Investigation of EEG abnormalities in the early stage of Parkinson's disease," *Cognitive neurodynamics*, vol. 7, no. 4, pp. 351-359, 2013.
- [124] M. I. Vanegas, M. F. Ghilardi, S. P. Kelly, and A. Blangero, "Machine learning for EEG-based biomarkers in Parkinson's disease," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 2661-2665: IEEE.
- [125] M. Koch, V. Geraedts, H. Wang, M. Tannemaat, and T. Bäck, "Automated Machine Learning for EEG-Based Classification of Parkinson's Disease Patients," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 4845-4852.
- [126] S. A. Khoshnevis and R. Sankar, "Classification of the stages of Parkinson's disease using novel higher-order statistical features of EEG signals," *Neural Computing and Applications*, vol. 33, no. 13, pp. 7615-7627, 2021.
- [127] C. Wei, L.-l. Chen, Z.-z. Song, X.-g. Lou, and D.-d. Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomedical Signal Processing and Control*, vol. 58, p. 101756, 2020.
- [128] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal emotion classification using EEG signal with sparse discriminative ensemble," *IEEE Access*, vol. 7, pp. 40144-40153, 2019.
- [129] J. A. Goguen, "SOME CONSIDERATIONS ON EVOLUTIONARY ALGORITHMS," CALIFORNIA UNIV BERKELEY DEPT OF MATHEMATICS1966.
- [130] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-260, 2015.
- [131] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097-1105, 2012.
- [132] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455-5516, 2020/12/01 2020.

- [133] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770-778.
- [134] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," pp. 1026-1034.
- [135] S. Li, J. Jiao, Y. Han, and T. Weissman, "Demystifying resnet," *arXiv preprint arXiv:1611.01186*, 2016.
- [136] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," pp. 630-645: Springer.
- [137] P. N. Bhagat, K. S. Ramesh, V. G. R. Matcha, and S. T. Patil, "Robust Prior Stage Epileptic Seizure Diagnosis System using Resnet and Backpropagation Techniques," *International Journal*, vol. 8, no. 5, 2020.
- [138] A. Shalbaf, S. Bagherzadeh, and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4, pp. 1229-1239, 2020.
- [139] R. J. Barry, A. R. Clarke, R. McCarthy, M. Selikowitz, S. J. Johnstone, and J. A. Rushby, "Age and gender effects in EEG coherence: I. Developmental trends in normal children," *Clinical neurophysiology*, vol. 115, no. 10, pp. 2252-2258, 2004.
- [140] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Application of higher order statistics/spectra in biomedical signals—A review," *Medical Engineering and Physics*, vol. 32, no. 7, pp. 679-689, 2010.
- [141] E. A. AlBadawy, S. Lyu, and H. Farid, "Detecting AI-Synthesized Speech Using Bispectral Analysis," in *CVPR Workshops*, 2019, pp. 104-109.
- [142] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278 - 305, 1991.

Appendix A: Copyright Permissions

The permission below is for the use of published content used in Chapters 2 and 3.



Requesting permission to reuse content from an IEEE publication

Applications of Higher Order Statistics in Electroencephalography Signal Processing: A Comprehensive Survey
Author: Seyed Alireza Khoshnevis
Publication: IEEE Reviews in Biomedical Engineering
Publisher: IEEE
Date: 2020
Copyright © 2020, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

[BACK](#) [CLOSE WINDOW](#)

The permission below is for the use of published content in Chapters 4 and 5.

9/11/2021

RightsLink Printable License

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 11, 2021

This Agreement between Mr. Seyed Alireza Khoshnevis ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5146040566454
License date	Sep 11, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Neural Computing & Applications
Licensed Content Title	Classification of the stages of Parkinson's disease using novel higher-order statistical features of EEG signals
Licensed Content Author	Seyed Alireza Khoshnevis et al
Licensed Content Date	Nov 24, 2020
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be	no