Adopting Machine Learning Technology for the Classification of Parkinson's Disease

A Thesis Submitted in
Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

By

Farhan Mohammed

In

Faculty of Engineering and Information Technology

School of Electrical and Data Engineering

UNIVERSITY OF TECHNOLOGY SYDNEY

AUSTRALIA

Submitted JANUARY, 2020

UNIVERSITY OF TECHNOLOGY SYDNEY
SCHOOL OF ELECTRICAL AND DATA ENGINEERING

The undersigned hereby certifies that he has read this thesis entitled **"Adopting Machine Learning Technology for the Classification of Parkinson's Disease"** by **Farhan Mohammed** and that in his opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy

Production Note:
Signature removed prior to publication.

. . . . . .

Prof. Xiangjian (Sean) He

Principal Supervisor

Certificate of Authorship/Originality

I, Farhan Mohammed, declare that this thesis, is submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology here at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Trainig Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 20/07/2020

ABSTRACT

Adopting Machine Learning Technology for the Classification of Parkinson's Disease

by

Farhan Mohammed

Parkinson's Disease (PD) is the second most common neuro-degenerative disorder affecting approximately 1% of the population. Major symptoms include tremor, bradykinesia and freezing of gait. The precise diagnosis of PD remains a challenge for clinicians due to the similarity of PD symptoms with other disorders. Although diagnosis is based on clinical symptoms, PD is associated with a plethora of non-motor symptoms adding to its overall disability.

Research into early diagnosis of PD has taken advantage of machine learning-based image analysis. Neuroimaging modalities, like Single Photon Emission Computed Tomography (SPECT), have shown to aid in early diagnosis of PD. SPECT images are powerful tools that depict dopaminergic deficits in brain. Dopamine transporter (DAT) loss is a crucial feature necessary for PD identification. Regular scans are identified by intense and symmetric DAT binding that appear as two "comma-shaped" regions. Any asymmetry of this shape implies an abnormal finding. This thesis proposes three machine learning approaches for the classification of PD.

Firstly, we developed a neural network that classified PD patients from healthy controls. SPECT images were used to train our network. 10-fold cross-validation was used to evaluate our network's performances. Experimental results indicated that our approach outperformed the benchmark studies.

To classify PD patients into different stages, PDStageNet was implemented which learned several features from the images and their associated stages. The affected re-

gions were enhanced using image segmentation process. Experimental results showed that the proposed approach achieved a very high accuracy in classifying PD patients into five clinical stages of PD progression.

Finally, a classification model was utilised to streamline the process of identifying PD patients for surgical treatment using only clinical data. A feature selection process was used to identify the essential features and determine if the accuracy could be improved. Two experiments were carried out to test this hypothesis. In experiment 1, the best classifier for PD classification was identified. In experiment 2, feature selection process was implemented to determine the essential features. Experimental results indicated that, with only 60% of the features, the accuracy was higher than the benchmark studies.

The significance of these studies is that we propose effective machine learning-based approaches to diagnose PD at its earlier stages, so that the management and prognosis of PD patients can be significantly improved. Given the high performance of our approaches, we believe that the early diagnosis of PD can be done, which will revolutionize PD diagnosis and management.

Acknowledgements

This research would not have been possible without the guidance and the help of many people.

First of all, I would like to thank my supervisor, Prof. Xiangjian He, for his guidance, support and encouragement in my academic career. I have learned so much from him in regards to research and academic writing. I know that Prof. Xiangjian He has always been doing his best for me and believe his knowledge, experience and vision enabled him to help me do my best in my academic career.

I would also like to express my gratitude to my co-supervisors, Dr. Yiguang Lin and Prof. Jinjun Chen. Dr. Yiguang Lin has provided me with valuable and constructive comments that helped me refine my writing skills. I believe that what I have learned from him will be a valuable asset for me in the future. In addition, their guidance, comments and suggestions during the course of my PhD research have been invaluable. Prof. Xiangjian He, Dr. Yiguang Lin and Prof. Jinjun Chen are nice and humble and have a very good reputation in our research community.

Finally, I would like to express my love and gratitude to my parents and my family. I am extremely grateful for their love, prayers and moral support to complete my PhD studies. Being a member of such a wonderful family is the most beautiful thing that happened to me. Thank you very much for everything you have done for me.

Farhan Mohammed

Author's Publications for the PhD

Book chapter

B-1. **F. Mohammed**, X. He and Y. Lin,"Applications of Machine Learning Techniques in the Diagnosis of Parkinsons Disease: Promises and Challenges," *mHeath for Belt and Road Initiative: mHealth for Parkinson's Disease*, (pending publication)

Journal Papers

J-1. **F. Mohammed**, X. He and Y. Lin, "Easy-to-use Deep Learning Model for Highly Accurate Diagnosis of Parkinson's Disease using SPECT Images," *Computerized Medical Imaging and Graphics*, (pending publication).

Conference Papers

C-1. **F. Mohammed**, X. He, Y. Lin and J. Chen, "A Novel Model for Classification of Parkinsons Disease: Accurately Identifying Patients for Surgical Therapy, *Proc. of 52nd Hawaii International Conference on System Sciences (HICSS-52)*, Hawaii, USA, January, 2019. (ERA Tier A conference)

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Parkinson's Disease

Parkinson's Disease (PD) is the second most common neuro-degenerative disorder, after Alzheimer's disease, affecting approximately 1% of the population aged 55 or older. The precise diagnosis of PD, at its early stages, remains a challenge for modern clinicians. The difficulty to differentiate PD from other neuro-degenerative disorders is high due to the similarities in symptoms with other disorders. Although clinical diagnosis is primarily rested on the presence of typical clinical manifestations such as bradykinesia, tremor and other cardinal motor features, PD is associated with a plethora of non-motor symptoms adding to its overall disability.

1.1 Diagnosing Parkinson's Disease

The cause for PD is unknown, and there is presently no cure. All current treatment options, such as medication and surgery, aim in the management of symptoms only. Patho-physiologically, PD involves the malfunction and death of vital nerve cells that produce dopamine in the basal ganglia of the brain. As PD progresses, the amount of dopamine produced in the brain decreases, leading to difficulty in controlling movement normally [1].

PD occurs mostly in the age of 50 and above, and it is difficult to identify in the earlier stages. It is primarily characterised by the cardinal motor impairments like rigidity, tremor and bradykinesia. Tremor is the well-known symptom of PD, resulting in an automatic movement of arms, lips and hands. Rigidity is another common symptom of PD. It causes severe stiffness of the muscles. Some other

symptoms include problems with walking and slowness of movements. Levodopa is prescribed at the early stages of the disease. When the disease advances, the symptoms also progresses rapidly. When medicinal treatments become ineffective, surgical treatment is the only alternative. Therefore, diagnosing PD at an earlier stage is critical for improving the quality of a patient's life [2].

According to the Global Burden of Disease, Injuries and Risk Factors Study (GBD) in 2018 [3], PD was the fastest growing reason in overall deaths caused by a neurological disorder. In that study, the overall number of people affected by the disease was estimated to have more than doubled globally from 1990 to 2017. The increase in deaths from PD was more significant than the rise in prevalence, and the considerable variation in death rates between countries was suggestive of a change in practices rather than higher death rates among PD cases. Figure 1.1 illustrates the number of deaths in major countries from 1990 to 2017. It can be seen that deaths due to PD increases over the years, especially in China and India.

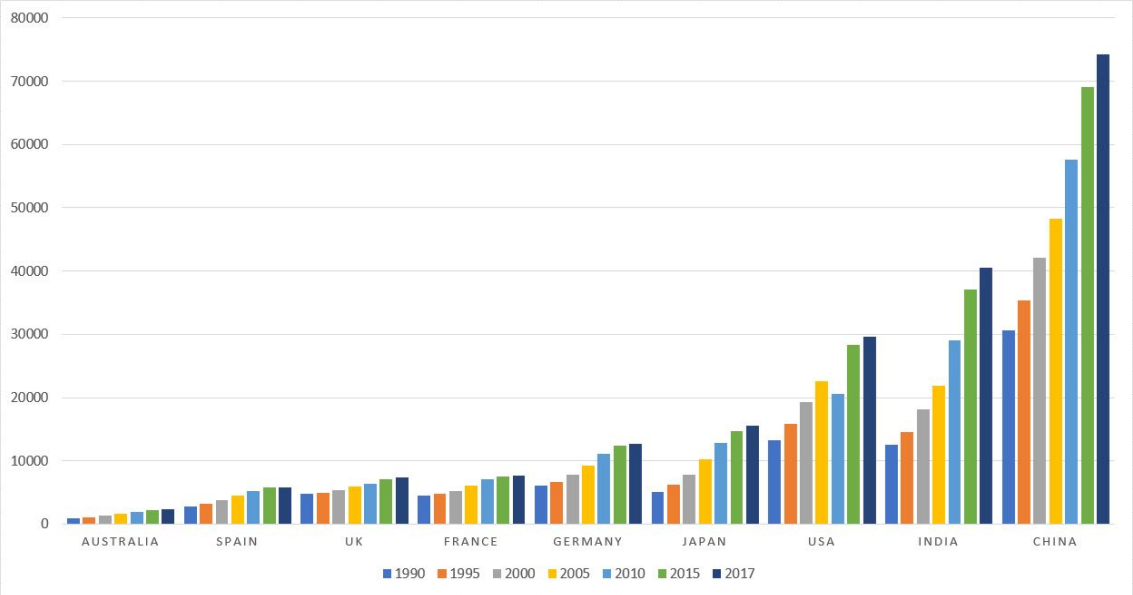
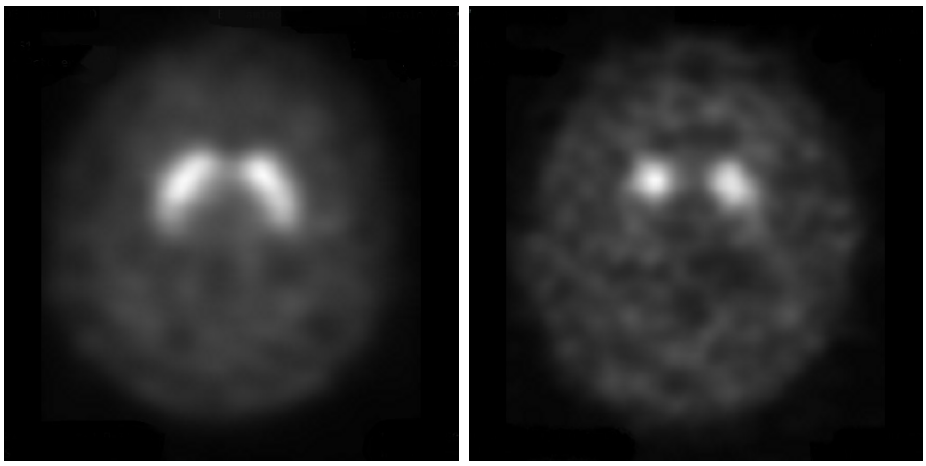


Figure 1.1 : PD deaths in major countries from 1990 - 2017

Visualisation of the striatal dopamine depletion in PD was a breakthrough in

molecular nuclear imaging [4, 5]. Since then, the field of neuroimaging has seen dramatic advances that are becoming increasingly relevant to the diagnosis of PD [6, 7, 8, 9, 10]. Several neuroimaging techniques such as dopaminergic imaging using Single Photon Emission Computed Tomography (SPECT) with 123I-Ioflupane (DaTSCAN) have demonstrated the detection of even early stages of the disease [11, 12, 13]. Dopaminergic imaging discriminates patients with PD by identifying presynaptic dopaminergic deficits in the caudate and putamen with high sensitivity and specificity [14]. Regular SPECT scans are characterised by full and symmetric DAT binding in the caudate nucleus and putamen on both hemispheres that appear as two "comma-shaped regions" (see Figure 1.2a). Any asymmetry or distortion of this shape implies an abnormal finding (see Figure 1.2b).



(a) (b)

Figure 1.2 : SPECT scans showing examples of (a) Non-PD patient and (b) PD patient

1.2 Health Expenditures Associated with Parkinson's Disease in Australia

Health expenditure data for PD was sourced from the Australian Institute of Health and Welfare (AIHW) [16]. The AIHW derives its expenditure estimates from an extensive top-down process developed in collaboration with the National Centre for Health Program Evaluation for the Disease Costs and Impact Study (DCIS). The approach measures health services utilisation and expenditure for specific diseases and disease groups in Australia. The health expenses of PD patients in Australia per year are relatively higher than many other diseases such as prostate cancer and breast cancer, mainly due to the higher use of residential aged care. However, individual expenses are relatively lower than dementia, cancers of all types, multiple sclerosis, infectious and parasitic diseases. The lifetime financial cost of a patient

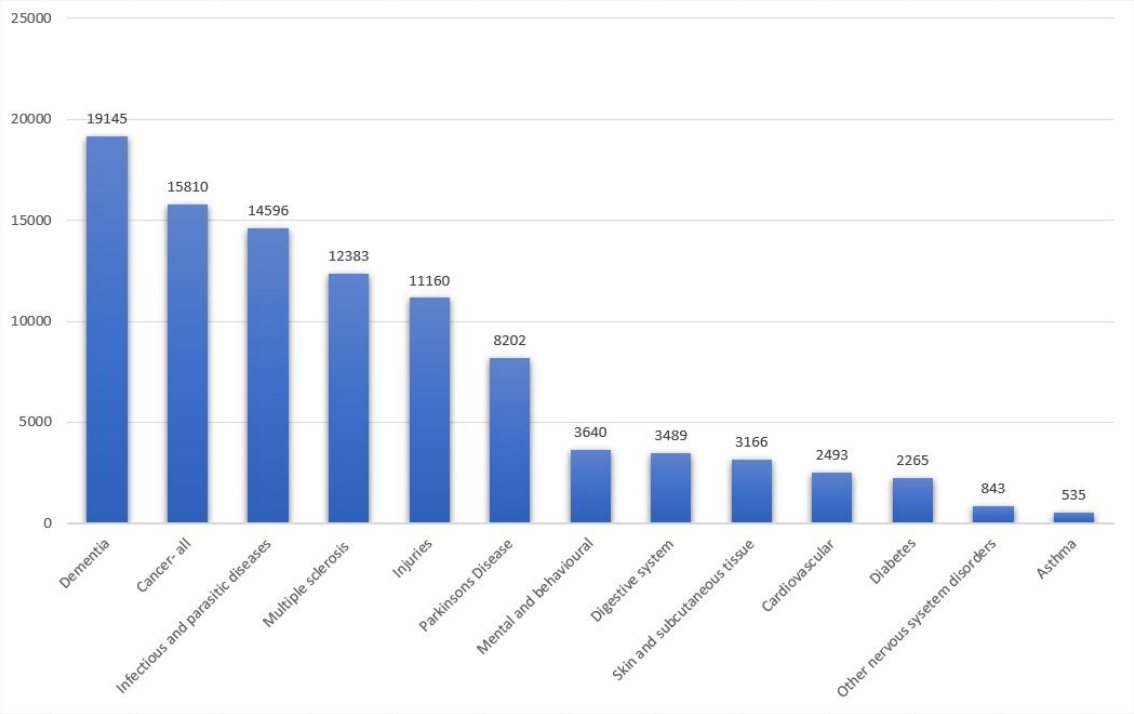


Figure 1.3 : Health system costs (AU$) per person in 2014

Source: Deloitte Access Economics [15]

living with PD for 12 years is equal to the average lifetime financial cost of cancer. While this is lower than many other cancers, it is significantly higher than prostate and breast cancer. Figure 1.3 shows the health costs per person based on several diseases in 2014.

There are other types of health costs (Figure 1.4) that contribute to the living conditions of a patient suffering from PD. These are mentioned below.

- Aged care. Residential aged care placement is often required, particularly in the later stages of PD due to functional impairment, drug complications and comorbidities associated with PD.
- Pharmaceuticals. Drug treatments for PD include drugs such as levodopa, non-subsidized prescription drugs such as ropinirole and over-the-counter drugs.
- Inpatient & outpatient hospital services. Hospital services may be required for the purpose of confirming diagnosis and levodopa responsiveness [17]. Hospital

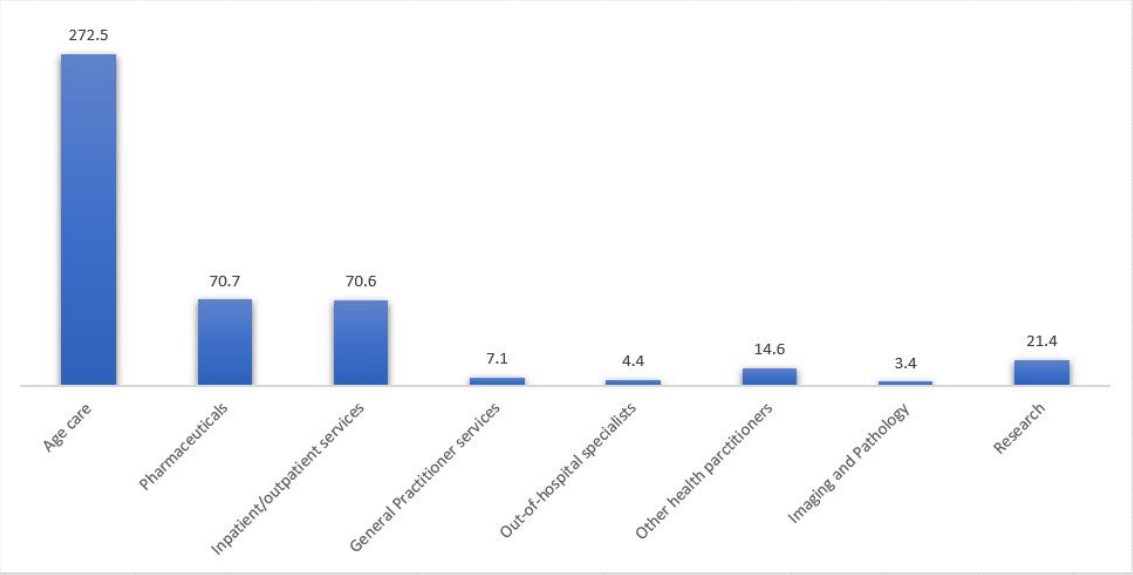


Figure 1.4 : Health costs (AU$) associated with PD, in millions

Source: Deloitte Access Economics [15]

admission may also be required for treatment for falls and other accidents, depression, and autonomic nervous system disorders, such as severe constipation, and other disorders arising from PD or PD medication.

1.3 Challenges in Diagnosing Parkinson's Disease

The precise diagnosis of PD, at its early stages, has enormous challenges for medical doctors as there are currently no laboratory tests that have been proven to be useful in diagnosing PD. Particularly in the early stages, the disease can be challenging to diagnose accurately. Doctors may sometimes request brain scans or laboratory tests in order to rule out other diseases. Blood tests, imaging techniques such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) scan, and SPECT, can be used to help doctors exclude other medical conditions, such as stroke or brain tumors, that produce symptoms similar to those of PD.

Amongst others, one of the methods for PD diagnosis is detecting and analysing voice disorders by using acoustic tools, that record the changes in pressure at lips or inside the vocal tract. Recently, upon signal processing, a group of experts found some features in the voices of the people with PD that can be used as discriminatory measures to differentiate those who have the disease from those who do not. After initial diagnosis, PD treatments are given to help relieve symptoms [18].

Other potential tests to diagnose PD include genetic testing, autonomic function testing, neurophysiological and neuropsychological tests, and neuroimaging [8]. Recommendations made by the European Federation of Neurological Societies/Movement Disorder Society - European Section (EFNS/MDSES) Task Force suggests that these available tests can be useful, although some tests are no longer recommended. Some of these tests, such as neuroimaging, are not routinely used in diagnosing PD. Consequently, it can often take up to two years from disease onset before a diagnosis is made [19]. The presence of other diseases, such as de-

mentia and general ageing, can obscure PD symptoms and reduce the chance of an accurate diagnosis implying that some PD patients may be either misdiagnosed or under-diagnosed. Furthermore, referral differences and accessibility of medical services in regional areas can also impact the chances of being diagnosed.

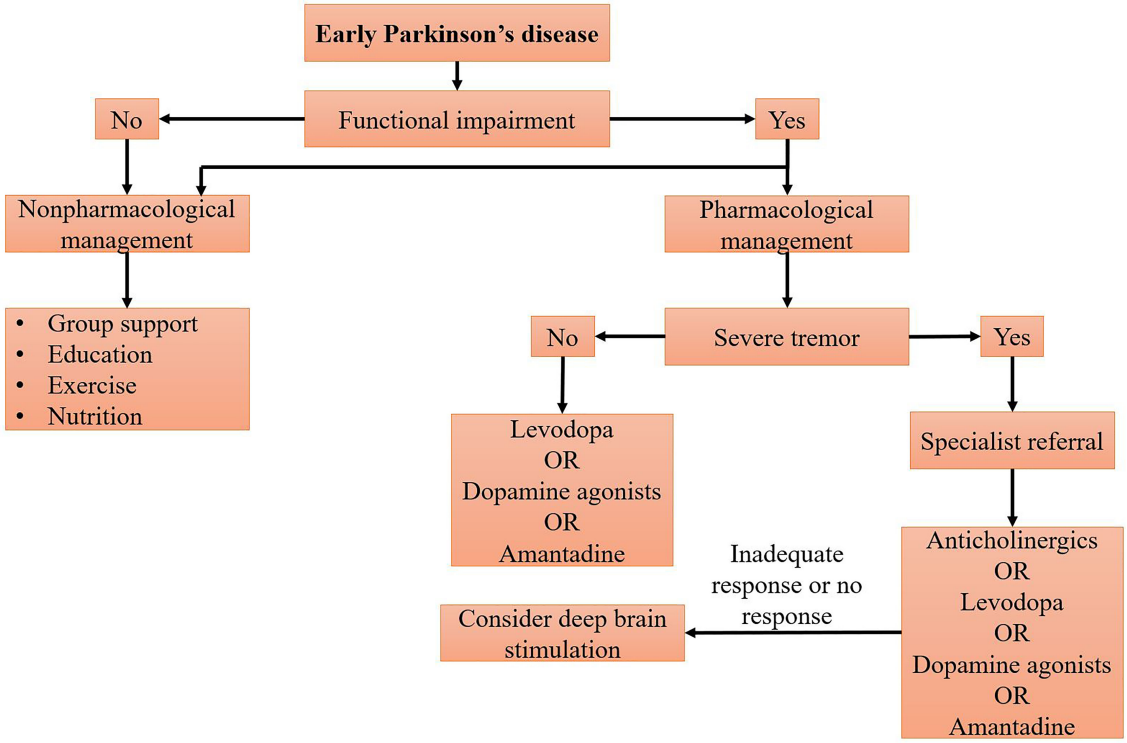


Figure 1.5 : Suggested treatment plan for early PD patient

Source: Therapeutic Guidelines [20]

Currently, there is no cure for PD. Treatment options for PD are generally grouped into drug treatment and non-drug treatment options such as exercise therapy, physiotherapy, speech therapy, and surgery. According to [20], Figure 1.5 and Figure 1.6 illustrate suggested guidelines for PD treatment and management, particularly for early PD and late (or advancing) PD, respectively. As there is no standard treatment for PD, each of these options can effectively manage symptoms of PD. However, various factors that influence appropriate treatment include duration since the onset of PD, age, disease severity, comorbidity, and other individual characteris-

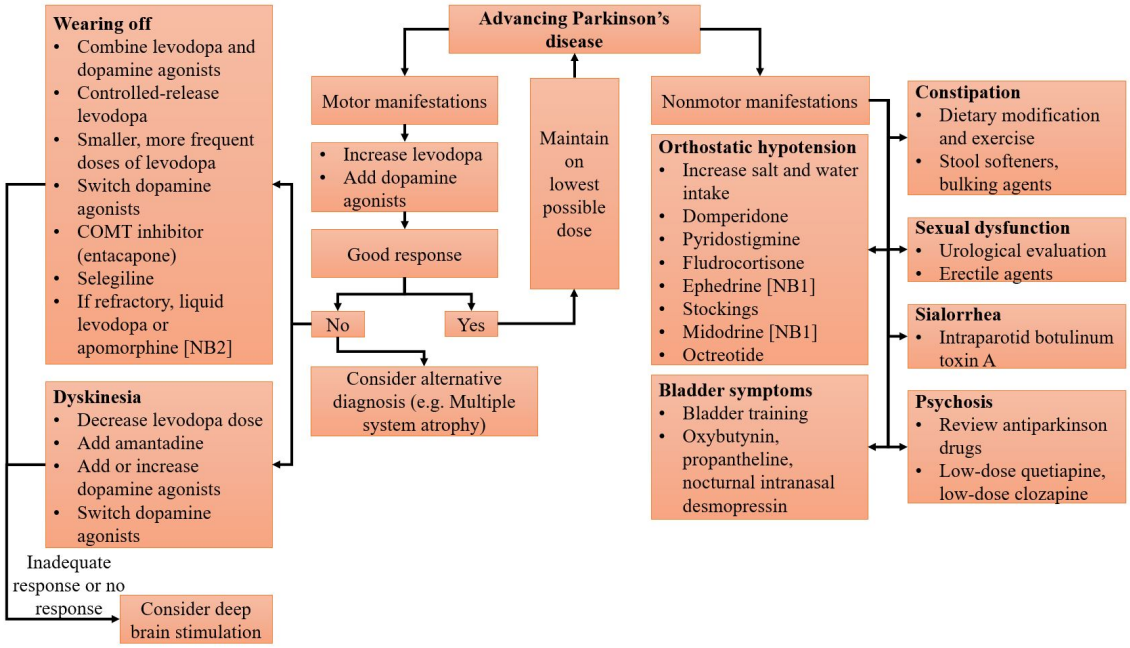


Figure 1.6 : Suggested treatment plan for late PD patient

Source: Therapeutic Guidelines [20]

tics [21]. These factors are particularly important in young people with PD, where misdiagnosis is common. Diagnosing PD patients at early stage is a challenge for clinicians because of the similarities in symptoms with other neurological disorders. Current treatment plans for managing PD symptoms involve prescribing Levodopa. However, as PD progresses, the symptoms rapidly increase, thereby causing medicinal treatment ineffective. Therefore, there is an urgent need to establish objective measures that can be used for the accurate diagnosis of PD. By utilising neuroimaging modalities like MRI, PET and SPECT, diagnosing PD at an early stage can be useful. These options aim to treat both the motor and non-motor symptoms of PD. Therefore, this initiative will require IT experts working together with clinicians and medical scientists using deep-learning tools that have potential medical applications.

1.4 Machine Learning Techniques

Machine learning [22] is a scientific discipline concerned with the design and development of algorithms, and it allows computers to recognise intricate patterns and make intelligent decisions based on data. In combination with image analysis procedures, it allows to develop computer-aided diagnosis (CAD) systems for several neurodegenerative diseases, such as PD [23]. These systems not only process and analyse image data but also can determine if an image belongs to the class of typical images (healthy subjects) or pathological images (patients), performing an automatic diagnosis [24].

In this section, we will divulge on some of the important concepts associated with machine learning. We will touch upon data mining concepts and describe some popular data mining algorithms. Then, a discussion will be provided on deep-learning techniques and its importance in the domain of image classification.

1.4.1 Data Mining Concepts

Data Mining is defined as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. The implementation of classifier systems for medical diagnosis is increasing gradually. Current advances in the field of artificial intelligence (AI) have led to the emergence of expert systems for medical applications. Moreover, in the last few decades, computational tools have been designed to improve the experiences and abilities of doctors and medical specialists in making informed decisions about their patients.

There is no doubt that the evaluation of data taken from patients and the decisions of experts are the most critical factors in diagnosis. However, expert systems and different AI techniques have the potential of being useful supportive tools for identification and diagnosis. Classification systems can help in increasing the accu-

racy and reliability of diagnoses and minimise possible errors, as well as making the diagnoses more time-efficient [25].

Data mining has been defined, by several authors, as the process of analysing large databases to detect meaningful patterns and associations [26, 27]. Data mining is a step within the KDD (Knowledge Discovery in Databases) process that involves using data analysis and discovery algorithms to generate particular patterns (or models) based on the data. Unlike data mining, the traditional method of converting data into knowledge relies on manual analysis and interpretation.

Data mining process uses several techniques from statistics and artificial intelligence in a variety of activities or application areas. The main activities are as follows.

- Classification. Classification involves determining profiles of classes in terms of their features and identifying which of these predefined classes a new item belongs to. For example, given particular classes of patients with different medical treatment responses, classification is used to identify the form of treatment to which a new patient is most likely to respond [28].

- Clustering. Clustering can be used to identify a set of classes where particular items are grouped according to their similar characteristics. For example, based on a patient data set, clustering can be used to identify subgroups of patients with similar treatment schema [29].

- Association. Association involves identifying relationships between items where the presence of one pattern indicates the presence of another pattern. For example, most patients who receive prescriptions for medication A also receive prescriptions for medication B [30].

1.4.2 Popular Classification Techniques

Classification is a pattern-recognition task that has applications in a broad range of fields. It requires the construction of a model that approximates the relationship between input features and output categories [31]. Some of the most popular techniques are discussed here in brief.

The Naïve Bayes classifier is based on applying Bayes' theorem with strong independence assumptions between the features. As one of its main features, the Naïve Bayes classifier is easy to implement because it requires a small amount of training data in order to estimate the parameters, and excellent results can be found in most cases. However, it has class conditional independence, meaning it causes losses of accuracy and dependency [32].

J48 is an efficient implementation of C4.5 [33] tree classifier that produces decision trees. The C4.5 is a classifier using binary trees based on the concept of information entropy computed in training data. In data mining, a decision tree is a predictive model which can be used to represent both classifiers and regression models. When a decision tree is used for classification tasks, it is more appropriately referred to as a classification tree. A classification tree is used to classify an instance to a predefined set based on their attribute values. It is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and the leaf nodes represent classes of class distributions [33]. The complex decision in this model is broken up into a group of several simpler decisions to find out the best solution [34] for the desired classification solution. Classification tree grows recursively by partitioning the training data. It is among the most used data mining algorithms [35].

K-Nearest Neighbor (k-NN) [36] is one of the simplest instance-based learning or lazy learning techniques and it assumes all instances correspond to points in the

n-dimensional feature space. k-NN is a supervised learning algorithm where the result of a new instance query is classified based on the majority of k-NN category. The purpose of this algorithm is to classify a new object based on attributes and training samples. The learner only needs to store the examples, while the classifier does its work by observing the most similar samples of the sample to be classified. The classifiers are only based on memory and do not use any model to fit. In order to classify an instance of a test data into the corresponding categories, k-NN calculates the distance between the test data and each instance of training dataset [37].

1.4.3 Deep-Learning Techniques

Deep-learning is a growing trend in general data analysis and has been one of the ten breakthrough technologies of 2013. Deep-learning is an improvement of artificial neural networks, consisting of more layers that permit higher levels of abstraction and improved predictions from data. To date, it is emerging as the leading machine learning tool in the domains of general imaging and computer vision [38].

In the stream of applying machine learning for data analysis, meaningful extraction of features lies at the heart of its success to accomplish target tasks. Conventionally, meaningful or task-related features were mostly created by human experts based on their knowledge about the target domains, which thus made it challenging for non-experts to exploit machine learning techniques for their studies.

However, deep-learning has relieved such obstacles by combining the feature engineering step into a learning step [39]. That is, instead of extracting features in a hand-designed manner, deep-learning requires only a set of data with minor preprocessing, if necessary, and then discovers the informative representations [40, 41]. Therefore, the burden of feature engineering has shifted from a human-side to a computer-side, thus allowing non-experts in machine learning to use deep-learning for their researches and applications.

In particular, Convolutional Neural Networks (CNNs) have proven to be powerful tools for a broad range of computer vision tasks. Deep CNNs automatically learn mid-level and high-level abstractions obtained from raw data (e.g., images). Recent studies indicate that the generic descriptors extracted from CNNs are incredibly effective in object recognition and localization in natural images. Medical image analysis groups across the world are quickly entering the field and applying CNNs and other deep-learning methodologies to a wide variety of applications.

A typical CNN architecture for image processing consists of a series of convolution layers, interspersed with a series of data reduction or pooling layers. The convolution layers are applied to small patches of the input image. Like the low-level vision processing in the human brain, the convolution layers detect increasingly more relevant image features, for example, lines or circles that may represent straight edges (such as for organ detection) or circles (such as for round objects like colonic polyps), and then higher-order features like local and global shape and texture.

Convolutional layer: The role of the convolutional layer is to detect local features at different positions in the input feature maps with learnable kernels. Specifically, the units of the convolutional layer compute their activations based on a spatially contiguous subset of units in the feature maps of the prior convolutional layer. After each convolutional layer, an activation layer is typically added. After the convolution, the activations for each filter are stacked and passed to the next layer, which is the max-pooling layer.

Activation: The activation function is common to all types of CNNs. It is applied to the output of all operations performed at the network and provides activation of the operations. Many different types of activation functions exist, such as the traditional sigmoid. However, in the context of CNN, the Rectified Linear Unit (ReLU) [41] has gained much popularity. The function itself is a non-saturating

activation function, which is defined as follows.

$$f(x) = x^+ = \max(0, x), \tag{1.1}$$

where x is the input to a neuron.

Max-pooling layer: A pooling layer follows the convolutional layer to down-sample the feature maps of the preceding convolutional layer. Specifically, each feature map in a pooling layer is linked with a feature map in the convolutional layer. Each unit in a feature map of the pooling layer is computed based on a subset of units within a local receptive field from the corresponding convolution feature map. Similar to the convolution layer, the receptive field finds a representative value, e.g., maximum or average, among the units in its field. This layer prevents the subsequent layers from processing non-maximal values, reducing computational load. Furthermore, by reducing the input space and keeping the receptive field of the filters, we can achieve translational invariance.

Dense layer: After a series of convolutional and max-pooling layers, there is always a fully connected layer, also known as dense layer. In this layer, all neurons are connected to all outputs from the last pooling step. The structure of this layer usually mimics a multilayer perceptron (MLP), in which the input layer is the output of the previous max-pooling layer with one or several hidden layers, and an output layer with as many neurons as classes. If the convolution and max-pooling layers together are considered as a sophisticated feature extraction system, the dense layer can be considered as the high-level reasoning part of the CNN.

The output of a CNN is typically one or more probabilities or class labels. The convolution filters are learned from training data. This is desirable because it reduces the necessity of the time-consuming hand-crafting of features that would otherwise be required to pre-process the images with application-specific filters or by calculating computable features. Figure 1.7 illustrates the architecture of a CNN.

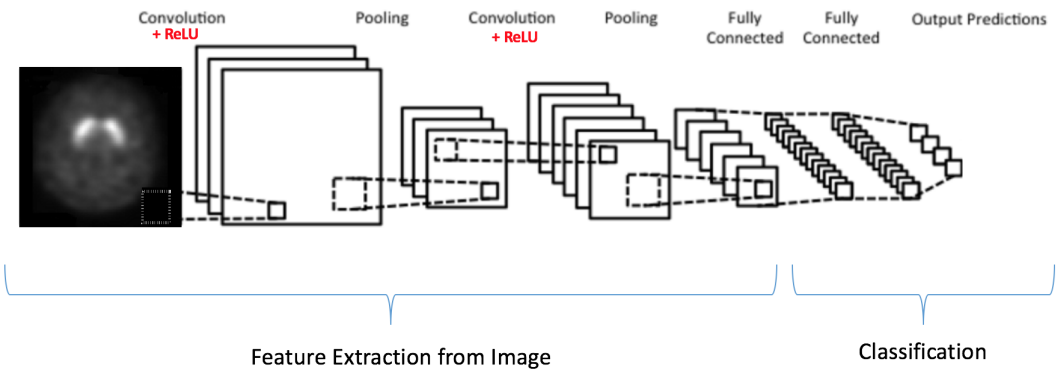


Figure 1.7 : Architecture of a CNN model

1.5 Parkinson Progression Markers Initiative (PPMI)

Parkinson Progression Markers Initiative (PPMI) [42] is one of the most well-known databases utilised for the classification of Parkinson disease. The PPMI repository contains a full set of clinical, imaging and biological data collected as part of an ongoing study along with processed images and bio-specimen analysis results. The PPMI database is a multi-centre international study involving subjects from different countries, adding diversity in the database, thus making it more robust. The repository is updated regularly as a longitudinal follow-up of study subjects continues, and the data generated from different analysis techniques are completed. PPMI is sponsored by the Michael J. Fox Foundation (MJFF) for Parkinson's Research [43].

Since 2002, The Michael J Fox Foundation for Parkinson's Research (MJFF) has been an essential driver for PD biomarker development efforts, and as a result of collaboration and efforts from several researchers, PPMI is the resulting database of these collaborations. For this thesis and all the experiments performed, the data are collected from PPMI.

1.6 Contributions

Accurate diagnosis of PD in early stages is challenging, even for expert medical physicians. The clinical diagnosis is definitive for the advanced stages of PD when the symptoms are fully developed. However, in the early stages, when the symptoms are mild/subtle, an accurate diagnosis becomes challenging. Early and accurate diagnosis of PD patients is crucial for critical reasons such as early management, avoidance of unnecessary medical examinations and therapies and their associated financial costs, side-effects and safety risks.

In this thesis, one of the main contributions is the adoption of deep-learning techniques to address such issues utilising SPECT imaging. An approach is proposed in this thesis, and it can detect early PD patients from non-PD patients using SPECT images. As SPECT images can depict the dopaminergic deficits in the striatum, the proposed approach learns such regions of interest as features necessary for classification of PD from non-PD patients. Experimental results show that our approach has better performance when compared to two benchmark studies.

Another main contribution of this thesis is the multiple stage classification of PD patients based on the progression of PD symptoms. Multiple stage classification is a critical issue in PD because identifying the stages of a PD patient is challenging and time-consuming, which delays the treatment and the chances for a better quality of life. Therefore, to address this issue, a deep-learning approach is proposed, and it utilises SPECT images to learn different features because patients in different stages of PD indicate different forms of dopamine deficit in the striatum. Experimental results show that our approach performs well with high-performance metrics.

Finally, a novel approach is presented to streamline the way to classify PD patients and identify appropriate patients for surgery based on clinical data only. A feature selection process is introduced to select features for PD classification, which

can be used to improve accuracy and to determine if, with feature selection, advanced PD patients can be identified for surgical treatment. Experimental results show that our proposed approach provides a better understanding of features that contribute to reliable and accurate PD classification, indicating that not all features are required for the accurate and efficient classification of PD.

1.7 Organization of this Thesis

The organization of this thesis is as follows. This chapter describes the concepts of PD and the challenges of diagnosing PD. The contributions to this thesis are also described. Chapter 2 provides a review of all previous research works in the classification of PD using different techniques. Chapter 3 describes the deep-learning approach used in this thesis to classify PD patients from non-PD patients. Chapter 4 shows how we can adopt deep-learning to classify patients into multiple stages of PD progression using SPECT images. Chapter 5 presents an approach that demonstrates the effect of feature selection in the classification of PD patients for surgical treatment utilising only clinical data. Finally, Chapter 6 provides the conclusion of this thesis.

Chapter 2

Review of Prior Research Works

In this chapter, a review of previous research studies addressing the classification of Parkinson's disease (PD) using machine learning techniques has been provided. They are categorised based on different attributes utilised for classification.

2.1 Classification of PD Using Vocal Attributes

One of the earliest classifications of PD using machine learning was done by Max Little from the University of Oxford [44]. In his study, he developed a software that learned to detect differences in voice patterns, in order to spot distinctive clues associated with PD. By using machine learning, a large amount of data was collected in order to know if someone had PD or not. A new measure of dysphonia, pitch period entropy (PPE) was introduced, which was robust to many uncontrollable effects, including noisy acoustic environments and normal, healthy variations in voice frequency. He collected sustained phonations from 31 people, 23 with PD, and performed classification using a kernel support vector machine (SVM).

Ramani & Sivagami [1] provided a survey of current techniques of knowledge discovery in databases using data mining techniques that were used in the classification of Parkinson Disease (PD). The diagnosis was based on medical history and neurological examination conducted by interviewing and observing the patient in person using the Unified Parkinson's Disease Rating Scale (UPDRS). The data was collected from a dataset created by Max Little of the University of Oxford. The dataset was composed of a range of biomedical voice measurements from 31 people,

of which 23 were with PD. Various feature selection algorithms were used, in which Fisher filtering was found to be a useful feature ranking system. This filtering was applied to the algorithms for better classification purpose. Based on the filtering algorithm, minimum number of attributes leading to better classification was selected and performed. The Fisher filtering algorithm ranked the input attributes according to the relevance. The main objective was to obtain the minimum error rate with the minimum characteristic features of the Parkinson Dataset. The Random Tree Algorithm classified the dataset accurately and provided 100% accuracy. The Linear Discriminant Analysis (LDA), C4.5 (decision trees), CS-MC4 (decision trees) and k-NN (k-Nearest Neighbor) yielded accuracy results above 90%. k-NN error rate was only 0.0256. Among all, the C-PLS algorithm classified the dataset with the least percentage of 69.74%. The important observation to be made is the features that were selected. The feature relevance analysis showed that the three critical features, namely Pitch Period Entropy, Non-Linear Measure-1 of fundamental frequency and Non-Linear Measure-2 of fundamental frequency, from the dataset were mainly aimed for better classification purpose.

Another study [18] aimed to classify between healthy people and people with PD. The dataset was gathered from the University of California at Irvine (UCI) machine learning repository. The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver Colorado, who recorded the speech signals. The authors applied fuzzy c-means (FCM) clustering and pattern recognition to the database to classify patients with PD. FCM clustering/classification depended on certain assumptions in order to define the subgroups present in a dataset. Then, pattern recognition was applied to identify the structure in data by a comparison with known structures (through clustering). Because of the unbalanced data, FCM clustering resulted in a low classification result of 58.46%. After adding pattern recognition, the success rate

increased to 68.04%. However, it was noted that through FCM and pattern recognition, the positive prediction value was 80.88%. This indicated that the classifier prediction of patients with PD was 80.88%. This was a very satisfactory result as PD was quite complex problems and the data mining tools used were very simple.

Khemphila and Boonjing [25] conducted PD classifications with a reduced number of attributes. They used Information Gain (IG) to determine the attributes of patients. Then, they used Artificial Neural Networks (ANNs) to classify PD in the given dataset. The dataset was taken from the UCI machine learning repository created by Max Little. Their experiment was done using WEKA. They divided the dataset into two sets: training and testing dataset. At first, IG was not used, and classification was done using ANN directly. The accuracy was 91.45% of the training set and 80.76% of the testing set. Then, IG was utilised, and the attributes were sorted according to the importance of features for classification. After implementing IG, the attributes were reduced to 16 and then ANN was implemented again. The accuracy rate was 82.05% for the training set and 83.33% for the testing set. The authors concluded that with IG the results showed a higher accuracy rate for ANN.

Shaikh and Chhabra [45] conducted a study to explore WEKA filters on data mining algorithm Naïve Bayes, which was used to classify PD patients from healthy controls. The datasets were obtained from the University of California Irvine (UCI) machine learning repository. At first, Naïve Bayes classifier was used, and the parameters were noted down. Then, the supervised attribute filter, "Discrete" from WEKA, was used on both data sets, which divided the input values of the datasets to a range of values and its parameters were noted down. Similarly, another filter (unsupervised), "Numeric Transform" was used on both datasets and the results were noted down. Naïve Bayes was used again to compare the performance measures of accuracy, sensitivity and specificity.

An effective and efficient diagnosis system using fuzzy k-nearest neighbor (FKNN) was presented in [46] for the classification of PD. FKNN was an improved version of the k-NN approach, and it incorporated the fuzzy set theory into k-NN. In FKNN, rather than individual classes, the fuzzy memberships of samples were assigned to different categories based on the fuzzy strength parameter. The fuzzy strength parameter was used to determine how heavily the distance was weighted when calculating each neighbor's contribution to the membership samples. The data for the study was gathered from the UCI machine learning repository. The dataset was composed of a range of vocal measurements from 31 people, of which 23 were with PD. The whole approach was comprised of two stages. In the first stage, feature reduction was conducted using Principal Component Analysis (PCA) to eliminate redundant features. In the second stage, FKNN model was trained via 10-fold cross-validation to get the optimal model and then the optimal FKNN model was used to perform classification. The results were compared with SVM classifier based on accuracy, sensitivity, specificity and area under the receiver operating characteristic (ROC) curve (AUC). Experimental results demonstrated that FKNN approach outperformed the SVM approach.

In a study by Gök [47], experimentation on classifying PD from healthy controls was conducted using the dysphonic symptoms (vocal characteristics) of PD. The dataset used was developed by Max Little. The experiment consisted of two phases. The first phase was feature selection (FS). In FS, relevant features were selected from 22 features using linear SVM for discovering informative features. Thus, out of 22, 10 were selected. Then, those ten were ranked again based on their information gain with respect to class. Six classifiers were selected to perform the comparison analysis. The accuracy of the competing algorithms was evaluated by utilising both the selected 10 and all 22 features. The best result was obtained by an ensemble of k-NN on PD dataset using only the ten selected features.

Suganya & Sumathi [48] developed a new data mining algorithm and it was named Artificial Bear Optimization (ABO). The algorithm was designed and implemented to evaluate the performance through the area under the curve (AUC). The dataset was collected from a private healthcare institute. From the dataset, 195 instances were collected for their investigation. These 195 instances had 23 characteristic features that consisted of a wide range of biomedical voice measurements. They utilised voice measurements because majority of reliable studies had shown that PD patients showed vocal deterioration symptoms. They implemented their algorithm and compared it with other algorithms like Subtracting Clustering Features Weighing with Kernel-Based Extreme Learning Machine (SCFW with KELM), FCM, Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO). The performance matrices of accuracy, recall, precision and f-measure were used as evaluation metrics. Based on their results, it was seen that ABO had the highest accuracy (97.5%), recall, precision and f-measure than the other algorithms.

A novel methodology for feature extraction and analysis from human speech was proposed in [49]. The methodology involved speech segmentation done at a pitch cycle level. Formulation of this methodology was based upon the hypothesis that, a cycle-by-cycle analysis can better capture the dysfunctionality of vocal tract musculature due to PD. Speech data from a total of 40 subjects with 18 healthy controls (12 female and 6 male) and 22 PD patients (7 female and 15 male) was used in this study. The original recordings contained two different passages read by all the participants. The word "Pablo" was chosen for analysis since it was repeated three times, providing more data to evaluate. On the whole, there were 115 utterances of this word. Out of all the available word utterances, only the ones with a clear distinction between the vowel segments and silences or syllable transitions were extracted and used for evaluation. The two syllables "Sy1" and "Sy2" in the utterance of the word "Pablo" contained the vowels "a" and "o", respectively.

Instead of dividing the vowel data into stationary segments, it was broken down into smaller segments each spanning a pitch cycle and features were extracted from these segments. Analysis was performed on the two syllables both individually as well as grouped together. The pre-processing of the word utterances was done on MATLAB using a custom developed GUI. Each word utterance was processed to identify the pitch cycle end locations in the voiced portions of each syllable. The boundaries for both syllables were marked manually to avoid the portions containing silences or syllable transitions. Data pertaining to each syllable was extracted using the first and last cycle end locations that fell within their respective boundaries. After the vowel data was extracted, it was normalised to unit amplitude and was further segmented into pitch cycle data segments (PCDS) using the previously identified cycle end locations. On the whole, a total of 115-word utterances were used for analysis after eliminating the unsuitable ones. After pre-processing, from each PCDS, features from temporal, spectral and cepstral domains were extracted. In order to eliminate the features that did not have much difference when the PD and HC groups were compared, within each gender group, the p-value output of a two-sample t-test conducted for each feature variance (FV) set was used. In the first stage of elimination, FV sets with p-values greater than 0.05 were eliminated. Then, an FV subset was created by adding one feature at a time, starting with the feature whose variance has the smallest p-value. With each new addition, clustering performance was evaluated using a k-mean classifier. It was seen that clustering with k-mean classifier using a selected feature set showed excellent performance in classification.

2.2 Classification of PD Using Gait Attributes

Gait analysis plays a vital role in obtaining information on motor deficits in PD. Gait information has been widely used for movement studies in healthy controls

and also in subjects with PD. Analysis of gait parameters is very useful for a better understanding of the mechanisms of movement disorders, and also has the high potential in presenting automatic non-invasive methods based on gait characteristics for diagnosis of PD.

A method was proposed in [50] to classify patients with PD and healthy control subjects using gait analysis via deterministic learning theory. It consisted of two phases: a training phase and a classification phase. The data was provided by PhysioNet bank [51], which contained measures of gait for 93 idiopathic PD patients and 73 healthy controls. In the training phase, gait characteristics were derived from vertical ground reaction forces under the usual and self-selected pace of the subjects. These patterns of PD patients and healthy controls constituted a training set. In the classification phase, a bank of dynamical estimators were constructed for all the training gait patterns. Prior knowledge of gait dynamics represented by the constant Radial Basis Functions (RBF) networks was embedded in the estimators. By comparing the set of estimators with a testing gait pattern of a specific PD patient to be classified (diagnosed), a set of classification errors were generated. The average of the errors were taken as the classification measure between the dynamics of the training gait patterns and the dynamics of the testing PD gait pattern according to the smallest error principle. With RBF network classification along with five-fold cross-validation method, the authors achieved an accuracy of 96.39%.

In [52], freezing of gait (FoG) detection was identified as a two-class classification problem: FoG versus normal locomotion. Similarly, FoG prediction was identified as a three-class classification problem. The authors introduced a new class called the pre-FoG, i.e., the walking periods before FoG. They hypothesised that there was a detectable deterioration of gait on this phase. The steps involved in their approach are:

- Extraction of standard frequency-based features, namely Freezing Index (FI) and total energy in the frequency band 0.5-8 Hz. This is the current standard in the field and serves as a baseline.

- Extraction of various hand-crafted time-domain and statistical features, which are used in pattern recognition problems involving motion or human activity recognition.

- Unsupervised feature learning. This method involves extraction of information from the raw data, without relying on domain-specific knowledge, or the availability of ground truth annotations. They evaluate the use of principal component analysis for extracting a compact representation of the structure of the signals.

The dataset was obtained publicly from the DAPHNet dataset [53], which contained data collected from eight PD patients that experienced regular FoG in daily life. Data were recorded using three 3D accelerometers attached to the shank (above the ankle), the thigh (above the knee) and the lower back of each subject. The subjects completed 20-30 minutes of walking sessions, which comprised of the following tasks:

- Walking back and forth in a straight line, including several 180 degrees turns,

- Random walking with a series of initiated stops and 360 degrees turns, and

- Walking simulating activities of daily life, which included entering and leaving rooms, walking to the kitchen etc.

The authors performed two sets of experiments, one for FoG detection and one for FoG prediction using different feature extraction strategies. The top-ranked features based in Mutual Information (MI) was AAE (Average Acceleration Energy), eigenvalues of dominant directions, range, variance, RMS (Root Mean Square) and

standard deviation. The results indicated that unsupervised feature learning outperformed FI by up to 7.1% and the time-domain and statistical features by up to 8.1%. Regarding FoG detection for FoG prediction, the patterns had different characteristics for each patient. As such, no conclusive result was obtained. However, it was noted that pre-FoG data were discarded from the training set, and the performance on FoG detection increased for all the feature extraction methods.

Mittra and Rustagi [54] carried out a study to identify PD patients by analysing gait data of PD patients. The data was obtained from PhysioNet Gait Analysis Database. The database consisted of data regarding 93 patients with idiopathic PD and 73 healthy control objects. It consisted of Vertical Ground Reaction Force (VGRF) of subjects as they walked for approximately 2 minutes. Every subject had a total of 16 sensors, of which 8 were on the bottom of each foot, to calculate the force in Newton as a function of time. A statistical approach was undertaken to transform the data with techniques such as Minimum, Maximum, Mean, Median, Standard Deviation, Skewness, and Kurtosis. These features, when calculated for every one of the 19 features that existed in the original dataset, gave a total of 133 features. One tuple in the new dataset represented one file of the original dataset, which consisted of 12,118 tuples. Each of these tuples thus provided a representative distribution of the data contained in the file that it pertains to. The distribution could hence show the characteristics of the 12,118 tuples in a single tuple. This data compression technique was used to aid the previously substantial time required during the modelling process. The newly transformed dataset consisted of 133 features and 310 tuples. A total of 5 different classification algorithms, such as K-Nearest Neighbors (k-NN), Logistic Regression, SVM (Linear Kernel, RBF Kernel, Poly Kernel), Decision Trees, and Random Forests, were used in an attempt to achieve an accuracy as high as possible. It was observed that k-NN achieved the highest accuracy.

Spatial-temporal gait features were utilised in [55] for the classification of PD. Spatial-temporal gait characteristics of an individual were affected by ageing, and physical differences between subjects, including height, body mass and gender, might increase gait variability and limit the degree to which pathological trends might be reliably discerned. The aim of this study was two-fold. The first was to use a multiple regression normalisation strategy that accounted for several characteristics to identify differences in spatial-temporal gait features between PD patients and non-PD patients. The second aim was to use machine learning strategies to classify PD after normalisation. The data was gathered from an existing database, which included gait data of 23 PD patients and 26 non-PD patients. Gait data was normalised using two approaches: dimensionless (DS) approach and multiple regression (MR) approach. Five machine learning strategies were employed to classify PD gait and they were kernel Fisher discriminant (KFD), Naïve Bayes, k-nn, SVM and Random forest (RF). These machine learning strategies were applied using raw spatial-temporal gait features, spatial-temporal gait features normalised using the DS approach and spatial-temporal gait features normalised using the MR approach. It was observed that SVM achieved maximum classification accuracy using on raw data. However, RF yielded the maximum classification accuracy using normalised DS and MR data.

According to [56], gait analysis involved estimation and evaluation of biomechanical features associated with walking. Analysis of gait and posture was one of the components of the clinical assessment of PD. In this study, a new set of features were collected from 20 subjects (10 PD and 10 non-PD) using the accelerometer sensor. The sensor provided data on several gait features such as average stride time, root mean square (RMS) of body acceleration, maximum and minimum acceleration, symmetry, stride-to-stride variability, variability of signal per stride and velocity. These features were then analysed to determine the most influential features using

maximum information gain minimum correlation (MIGMC). The next step included developing a feature set in which features were maximally dissimilar to each other. This step was done by pairwise Pearson correlation analysis. The threshold of 80% was considered to determine highly correlated features. Among all pairwise correlations, the one with the highest value and its associated features were chosen. Five machine learning methods, SVM, RF, Naïve Bayes, AdaBoost and Bagging, were selected to compare across different feature sets. Similarity network analysis was performed to validate the optimal feature set. It was observed that standardisation could improve all classifiers' performance. In addition, the feature set obtained using MIGMC provided the highest classification performance. It was shown that the results from Similarity Network analysis were consistent with the results from the classification task, emphasising on the importance of choosing an optimal set of gait features to help objective assessment and automatic diagnosis of PD.

In [57], the authors used a Hidden Markov Model (HMM) with Gaussian mixtures to classify PD patients from healthy subjects. HMM was a state machine with two layers consisting of state and observation layers in which a Markovian process controlled the selection of the state in each time. The data was gathered from the PhysioNet bank. The results showed that the HMM classifier performed better than the results obtained from least squared support vector machines (LS-SVM). The HMM method could effectively separate the gait data in terms of stride interval obtained from healthy control and PD patients with an accuracy of 90.3%.

A study to detect freezing of gait in PD patients was carried out in [58] using Logistic regression modelling. The data was gathered from the UCI machine learning repository. This data set was obtained by processing the signals obtained in three dimensions by the acceleration sensor placed in the ankle. The motions were sampled at 64 Hz. The signals from the sensor were sent to the computer for further processing via Bluetooth. PD patients were asked to perform their daily activities.

A two-stage method was proposed to detect the freezing of gait (FoG). First, the acceleration signals were measured using the acceleration sensor placed in the ankle of the patients. Second, the FFT (Fast Fourier Transform) algorithm was applied to these signals. Then, the frequency spectrum coefficients were extracted. To compose the feature set, five different parameters, including variance, maximum amplitude, minimum amplitude, maximum energy, and minimum energy, were calculated and then used as the feature set for each category. After composing the feature set, the logistic regression classifier was used to classify the FoG and Non-FoG cases. Logistic regression was a predictive analysis method. Logistic regression classifier was used to explain the relationship between a dependent binary variable and one or more nominal ranges or proportional arguments and to interpret the data. The logistic regression classifier was trained and tested using 16 samples. 10-different performance measures were used to evaluate the performance of the classifier,. A comparative study with four different classifiers was provided. It was seen that the proposed logistic regression model performed the best.

2.3 Classification of PD Using Striatal Binding Ratio (SBR)

Prashanth, et al. [59] performed a study on early PD classification using Single Photon Emission Computed Tomography (SPECT) imaging. The data was gathered from the PPMI database. They used the Striatal Binding Ratio (SBR) values of the four striatal regions (left and right caudate, left and right putamen) calculated from the SPECT images. The authors gathered a total of 674 SBR values (181 normal and 493 PD). They used Support Vector Machines (SVM) and multivariate binomial logistic regression (MLR) to develop automatic classification and prediction/prognostic models for early PD. SVMs performed hard classification, whereas MLR performed soft classification enabling the development of prediction/prognostic model for PD risk estimation. 10-fold cross-validation was per-

formed to evaluate the classification accuracy. In their study, they found that the accuracy of SVM with radial basis function provided the highest accuracy. Only four features were utilized without the need for any feature selection techniques.

Hirschauer et al. [60] focused on the diagnosis of PD using Enhanced Probabilistic Neural Network (EPNN). A comprehensive computer model was presented using motor, non-motor and neuroimaging features from the PPMI dataset. The model was tested for differentiating PD patients from those with scans without evidence of dopaminergic deficit (SWEDD). The results were compared with four other commonly used algorithms: the probabilistic neural network (PNN), SVM, k-nearest neighbor (k-NN) and classification tree (CT). Based on the results, EPNN had the highest accuracy.

Prashanth et al. [61] worked on the detection of early PD through multimodal features and machine learning. The authors used non-motor features of rapid-eye-movement (REM) sleep behavior disorder (RBD) and olfactory loss, along with other significant features such as cerebrospinal fluid (CSF) and dopaminergic imaging markers from 183 healthy normal and 401 early PD subjects. These data were gathered from PPMI and the classifiers used were Naïve Bayes, SVM, Boosted trees and Random Forests. The dataset was divided in a way so that 70% was used for training and the rest 30% was used for testing. Based on the results, SVM provided the highest accuracy.

Oliviera et al. [11] conducted a study to assess the potential of a set of SBR features extracted from SPECT images to detect and confirm dopaminergic degeneration, and thus assist in the clinical decision for the diagnoses of PD. Seven features were extracted from each brain hemisphere and tested on 652 images gathered from the PPMI database. The discriminative potential of the extracted features to detect PD were evaluated both individually and in groups. Three different classifiers

were used, and they were SVM, k-NN and logistic regression. Leave-one-out cross-validation (LOOCV) was used to assess classification accuracy. It was observed that SVM outperformed the other classifiers.

In [62], the objective was to develop a fully-automated computational solution for computer-aided diagnosis in PD using SPECT. The images were gathered from PPMI and pre-processed using automated template-based registration followed by computation of the binding potential at a voxel level. Voxels represented point values on a regular grid in the three-dimensional space. Then, the binding potential images was used for classification, based on the voxel-as-feature approach and using the SVM classifier, i.e., each voxel in a volume of interest (VOI) resulted in a feature for the SVM classifier. Two different reference VOIs, cortex reference VOI and occipital reference VOI, were used to compute the binding potential. Cortex reference VOI also included the striatal VOIs, which depicted the striatal region. Three approaches were tested for building the feature vectors. In approach one, only the voxels of the striatal VOI with lower mean binding potential were used to assemble the feature vector. In approach two, all voxels of striatal VOIs were used. For these two approaches, when necessary, the mean binding potential of the images were mirrored, relatively to the sagittal plane, to guarantee that the striatal VOI with lower mean binding potential was always on the same side. In approach three, all voxels of striatal VOIs were used without mirroring. Leave-one-out cross-validation (LOOCV) was used to evaluate classification accuracy. It was noted that the SVM classification, with the third approach, achieved the highest accuracy.

A comprehensive analysis of SPECT images was carried out in [63], using voxel-based logistic lasso model. In this study, the authors believed that region-based analysis had its limitations, and as such, other areas of the brain were included for a better comprehensive analysis. Therefore, regions such as thalamus and globus pallidus were included in their analysis. Their approach included all voxels from

these regions in the data and allowed an algorithm to decide which voxels are more informative. Logistic lasso model was used as the machine learning method for classification. The logistic lasso model used was voxel-based. It worked by using a linear combination of a sparse set of voxels to calculate the probabilities of belonging to the classes. Voxels in the sparse set were chosen purely based on the training data. Those voxels useful for classification were likely to be a subset of all voxels that were analyzed and possibly affected by PD. SPECT scans from 658 (210 controls and 448 PD) subjects were collected from the PPMI database. The SBR was calculated for all voxels, which were represented as features. The logistic lasso model was applied to the SPECT images in four different ways. First, all of the images were used as training images, with 10-fold cross-validation to determine the training accuracy. Next, the images were divided into three equal-sized groups, each group containing the same fraction of control and PD images as the original set. Then, holding back one group at a time as a test set, the remaining two groups were merged to form a training set. This was called the split-data case. Finally, all images were summed up along the z-axis to create 2-D images and were referred as 2-D split-data case. It was noted that 3-D voxel-wise logistic analysis provided accurate classification. The analysis showed that sub-regional voxels in all regions were informative for classification.

In [64], an automated classification of PD using SPECT scan images was presented. The authors used the Single Value Decomposition (SVD) technique to reduce the training set of image data into vectors in feature space, called D space. The automatic classification technique used the distribution of the training data in D space to define classification boundaries. Other patients could be mapped into D space, and their classification could be automatically given. The approach was tested using 116 patients for whom the diagnosis of either Parkinsonian syndrome or non-Parkinsonian syndrome was confirmed from post imaging follow-up. Naïve

Bayes (NB) classifier was used to perform the classification. A leave-one-out cross-validation was performed to train and test the approach. It was observed that NB combined with SVD achieved the accuracy of 94.86%

In [65], SVM was utilised to develop a computer-aided diagnosis model to classify PD patients through SPECT imaging. The dataset was gathered from a private health institute with a total of 80 subjects (41 PD and 39 non-PD). The authors utilised Principal Component Analysis (PCA) and Independent Component Analysis (ICA) to extract features. PCA was a statistical approach which aimed to reduce the dimensions of observation space. The reduction was obtained by creating new linear combinations, called principal components, of variables characterising the objects. ICA was another statistical approach that represented a multidimensional random vector as a linear combination of non-gaussian random variables (called independent components) to be as independent as possible. These two techniques were used to extract features, and SVM was utilised to classify the images. It was noted that the accuracy of the approach was 95%.

2.4 Classification of PD Using Imaging Modalities

Electroencephalogram (EEG) signals were utilised in [66], to classify PD patients using CNNs. EEG signals could easily identify the functions of the cortical and subcortical parts of the brain. EEG signals were complex and non-linear in nature, and hence, many linear feature extraction approaches were unable to accurately characterise these signals. When EEG signal displays complexity, aggravation of the PD was observed. This was due to the presence of non-linear components in the EEG signals. The data was provided by a hospital which included EEG signals of 20 PD patients. The authors implemented a 13-layer CNN to learn and classify two classes (PD and normal). A 10-fold cross-validation was performed to evaluate the network. Performance metrics such as accuracy, sensitivity and specificity were

calculated to analyse the network's efficiency.

Martinez et al. [67] utilised a deep-learning network to classify PD patients from healthy controls subjects using SPECT images. The data was collected from the PPMI database. A total 301 subjects' SPECT images were collected. In their training process, a bounding box was drawn around the regions of interest (ROI), which was gathered by calculating the intensity threshold. Once this was applied, the rest of the image was obviated. A 10-fold cross-validation was used for evaluation, from which metrics like accuracy, sensitivity and specificity were extracted.

Choi et al. [68] utilised deep neural network, called PD Net, for the classification of PD from healthy controls. In their study, a deep-learning based SPECT interpretation system was used to refine diagnosis of PD. The system was trained on SPECT images obtained from PPMI and tested on another dataset obtained from Seoul National University Hospital (SNUH). PD Net showed a high accuracy that was comparable with the experts' evaluation referring quantification results. The system achieved an accuracy of 96%.

2.5 Conclusion

The purpose of this review was to view previous studies in the classification of PD using various attributes. It is clear from the research reviewed that PD classification is ongoing research and a serious medical issue. Along with this, it is also clear that machine learning technology has played a significant role in contributing to the field of PD classification. This field of inquiry is very important because at its center is a concern, i.e., helping PD patients in the early diagnosis of PD. In doing so, we can see the importance of this research growing as it will help achieve a better quality of life for PD patients.

Chapter 3

Classification of Parkinson's Disease using Deep-Learning

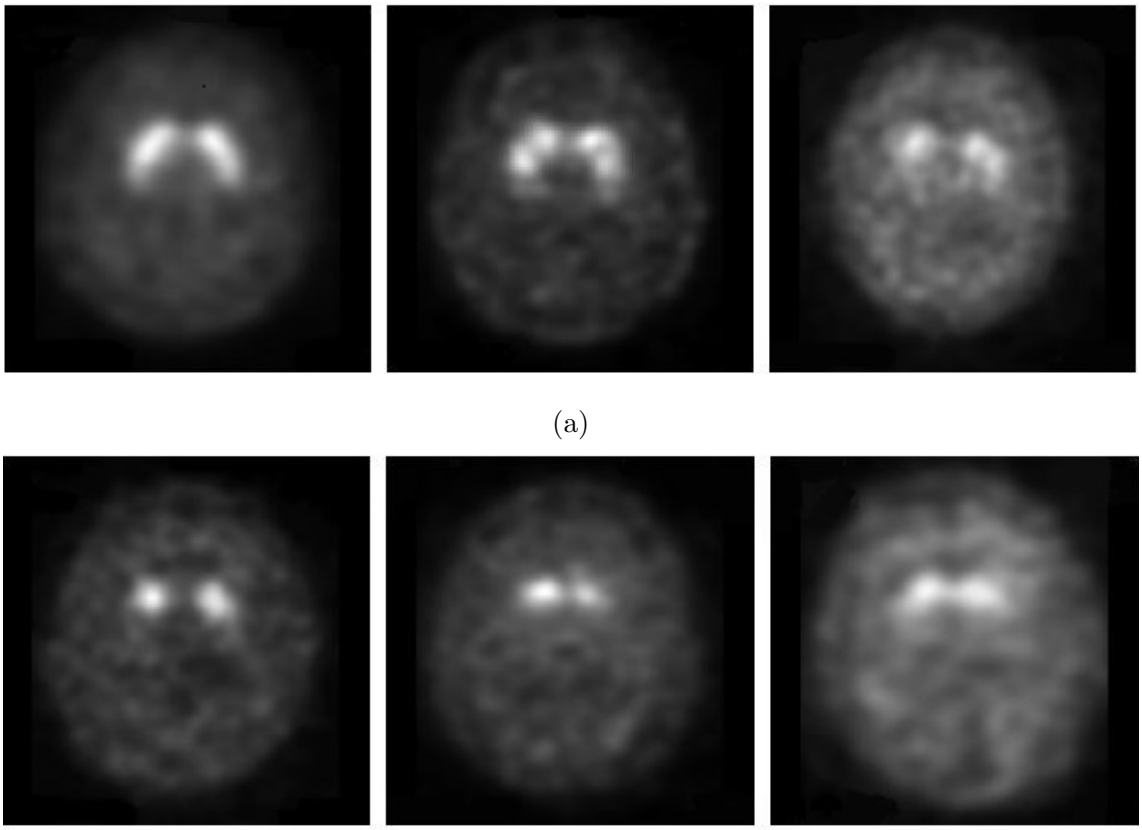
Parkinson's disease (PD) is a severe chronic, progressive and neuro-degenerative illness that affects people worldwide. PD results in a substantial reduction of dopamine content in the striatum along with corresponding losses of dopamine transporters (DATs). DAT loss in the presynaptic terminals is a key feature of PD identification. The progressive degeneration of dopaminergic neurons can be assessed by using radioligands in imaging-based approaches [11]. In this chapter, we develop a deep-learning model to classify PD patients from healthy controls using Single Photon Emission Computed Tomography (SPECT). Early and accurate diagnosis of PD is essential for early management and proper prognosis and for initiating neuro-protective therapies.

3.1 Introduction

Parkinson's disease (PD) is characterised by the malfunction and death of vital neurons located in the brain. These vital neurons produce dopamine, a chemical that sends messages to the part of the brain that controls movement and coordination [69, 59, 70, 71]. The major symptoms of PD are tremor, rigidity, postural instability and slow movement [2]. As PD progresses, the amount of dopamine produced in the brain further decreases.

Recent neuroimaging techniques, such as dopaminergic imaging using SPECT with 123I-Ioflupane (DaTSCAN), have demonstrated the detection of even early

stages of the disease [11, 12, 13]. Dopaminergic imaging discriminates patients with PD by identifying presynaptic dopaminergic deficits in the caudate and putamen with high sensitivity and specificity [14]. Normal SPECT scans are characterised by intense and symmetric DAT binding in the caudate nucleus and putamen on both hemispheres that appear as two "comma-shaped" regions (see Figure 3.1a). Any asymmetry or distortion of this shape implies an abnormal finding (see Figure 3.1b).



(a)

(b)

Figure 3.1 : SPECT scans showing examples of (a) Healthy control and (b) PD patient

The routine assessment of SPECT images is only done through visual evaluation by experienced nuclear medicine physicians. In most cases, a simple visual inspection is sufficient to confirm that there is a reduction of dopamine in the basal ganglia.

However, in many cases, the diagnosis may not be obvious, especially in early PD patients [62]. In addition, significant variations exist between individual physicians because of differences in experience and training. By adopting machine learning techniques, these problems in the diagnosis of PD patients can be easily reduced. Indeed, previous studies have demonstrated that diagnostic tools based on machine learning and deep-learning techniques could assist clinicians in the early diagnosis, treatment planning and monitoring of PD progression [59, 72].

When applying machine learning to data analysis, meaningful feature extraction or feature representation lies at the heart of its success in accomplishing the target tasks. Conventionally, meaningful or task-related features were mostly designed by human experts based on their knowledge about the target domains. Therefore, it is challenging for non-experts to exploit machine learning techniques for their studies.

Deep-learning [39] has reduced such obstacles by absorbing the feature engineering step into a learning phase. That is, instead of extracting features in hand designed manner, deep-learning requires only a set of data with minor pre-processing, if necessary, and then discovers the informative representations by self-learning [40, 41].

In medical imaging, Convolutional Neural Networks (CNNs) have proven to be powerful tools for a broad range of computer vision tasks [38]. For example, Electroencephalogram (EEG) signals were utilised in [66], to classify PD patients using CNNs. EEG signals could easily identify the functions of the cortical and subcortical parts of the brain. EEG signals were complex and non-linear in nature, and hence, many linear feature extraction approaches were unable to accurately characterise these signals. When EEG signal displays complexity, aggravation of the PD was observed. This was due to the presence of non-linear components in the EEG signals. The data was provided by a hospital which included EEG signals of 20 PD

patients. The authors implemented a 13-layer CNN to learn and classify two classes (PD and normal). A 10-fold cross-validation was performed to evaluate the network. Performance metrics such as accuracy, sensitivity and specificity were calculated to analyse the network's efficiency.

However, other published works have utilised SPECT images for classification of PD with different approaches. In [64], an automated classification of PD using SPECT scan images was presented. The authors used the Single Value Decomposition (SVD) technique to reduce the training set of image data into vectors in feature space (D space). The automatic classification technique used the distribution of the training data in D space to define classification boundaries. Other patients could be mapped into D space, and their classification could be automatically given. The approach was tested using 116 patients for whom the diagnosis of either Parkinsonian syndrome or non-Parkinsonian syndrome was confirmed from post imaging follow-up. Naïve Bayes (NB) classifier was used to perform the classification. A leave-one-out cross-validation was performed to train and test the approach. It was observed that NB combined with SVD achieved the accuracy of 94.86%

In [65], SVM was utilised to develop a computer-aided diagnosis model to classify PD patients through SPECT imaging. The dataset was gathered from a private health institute with a total of 80 subjects (41 PD and 39 non-PD). The authors utilised Principal Component Analysis (PCA) and Independent Component Analysis (ICA) to extract features. PCA was a statistical approach which aimed to reduce the dimensions of observation space. The reduction was obtained by creating new linear combinations, called principal components, of variables characterising the objects. ICA was another statistical approach that represented a multidimensional random vector as a linear combination of non-gaussian random variables (called independent components) to be as independent as possible. These two techniques were used to extract features, and SVM was utilised to classify the images. It was

noted that the accuracy of the approach was 95%.

The main objective of this chapter is to classify PD patients from healthy controls using SPECT images by adopting CNNs. This research was carried out to determine if a smaller and more compact neural network could outperform the state-of-the-art studies as mentioned in [67, 68]. To achieve our objective, we designed our proposed approach with minimal architectural complexity in order to address the practicality of utilising our network in a clinical setting. We compared our experimental results with [67, 68]. From our experimental results, our network outperforms the state-of-the-art works by large margins in all evaluation metics.

3.2 Proposed Network Approach

The proposed network consists of three convolutional layers, each with 128 nodes and two dense layers. This architecture is constructed because of the high accuracy of the images during training (Table 3.1). SPECT images for this research are gathered from the PPMI database.

The images are collected from PPMI. PPMI is a longitudinal and multinational study to assess the progression of clinical features, imaging and biological markers in PD patients and the control group. It involves subjects from different countries, adding diversity to the database, thus making it more robust. PPMI is sponsored by the Michael J. Fox Foundation [43]. All of the PD subjects are at different stages of the disease. PD subjects have confirmations, from the imaging core laboratory, that the screening of dopamine transporter SPECT scan is consistent with dopamine transporter deficit.

SPECT scans last for 30 to 45 minutes. The raw data is transferred to the imaging core laboratory and then iteratively reconstructed, using a hybrid ordered subsets' expectation maximization (OSEM) algorithm in a HERMES workstation

(Hermes Medical Solutions, Sweden). The reconstructed images are then transferred to PMOD software (PMOD Technologies, Zurich, Switzerland) for subsequent processing and attenuation correction. The final pre-processed images were saved in DICOM format.

The PPMI database consists a total of 1,359 SPECT images of PD patients that are categorized into different stages of PD. For this research study, they are all combined into one group, i.e. PD patients. The healthy control group consisits of 341 images only. In order to balance the dataset, data augmentation is perfromed on the healthy controls only (see Section 3.3.1). After data augmentation, the total number of SPECT images for the healthy control group increased to 1,364. After gathering the images, the following steps are proposed for the process of classification.

3.2.1 Network Structure

Our network is a modified version of AlexNet [73]. AlexNet is one of the first deep neural networks that has been utilized for image classification tasks. The network consists of five convolutional layers and three dense layers with different kernel sizes. The first layer has a kernel size of 11×11. The second layer has a kernel size of 5×5 and layers three to five each has a kernel size of 3×3. However, one of the disadvantages of this approach is the duplication of data. This occurs because of the overlapping blocks of pixels, which leads to an increase of memory consumption for processing the image.

In contrast, our network has only three convolutional layers and two dense layers, each with a kernel size of 3×3. This is because smaller-sized kernels of fewer layers lead to lower costs than larger-sized kernels with more layers. In addition, more layers and multiple larger-sized kernels lead to more in-depth network architecture and more complex feature learning which is time-consuming and expensive. Therefore, by reducing the kernel size, our network gains the advantage of learning the fea-

tures of the image faster. Furthermore, as shown in Figure 3.4, our network is also more robust and performs better than AlexNet with excellent scores in accuracy, sensitivity and specificity.

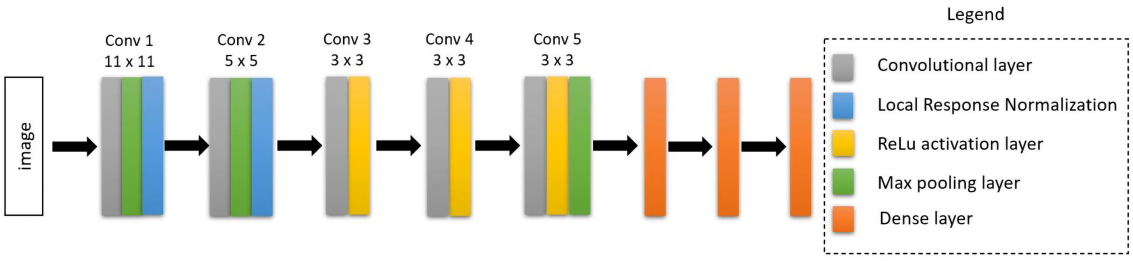
In AlexNet, Local Response Normalization (LRN) is added to the first two convolutional layers and then the activation layer, Rectified Linear Unit (ReLU) [41], is added to the next three layers. The ReLU function enables faster training of CNN, since the calculation of its derivative has a lower computational cost, without losing any of its generalization ability [73]. The function itself is a non-saturating activation function and is defined by

$$f(x) = x^+ = \max(0, x), \tag{3.1}$$

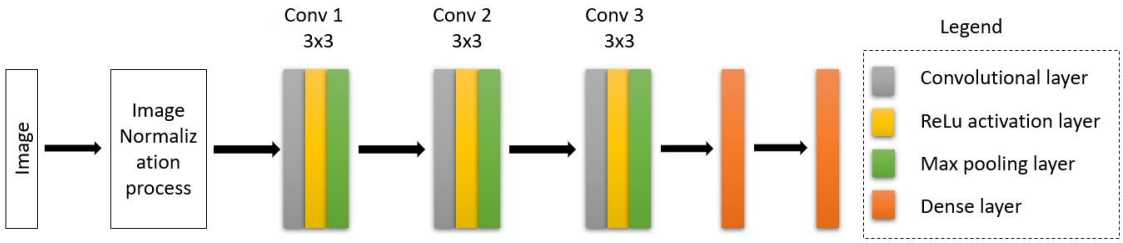
where x is the input to a neuron.

In AlexNet, such processes are used to amplify the features in the image, such as enhancing the brightness of the image, for better classification of the image. However, in our network, overall image normalization is used because LRN leads to increased memory consumption and computation time. The normalization process is carried out during the pre-processing stage, and the network is trained on the image that is generated. This process allows more distinguishable features to be visible during training. Our image normalization process is described below. In addition, our network only utilizes the ReLU activation function in all three layers. Our image normalisation process is described in Section 3.2.2. Figures 3.2a and 3.2b illustrate the architectures of AlexNet and our network.

Our network model is optimized using the Adam optimizer [74] with the default learning rate of 10^{-3}. The learning rate is decayed exponentially with a decay of 10^{-8}. These standard parameters of Adam have been proven for neural networks to be computationally efficient, of little memory usage and well suited for problems



(a)



(b)

Figure 3.2 : Network differences of (a) AlexNet and (b) Our network

with big data. Hence these parameters are utilized in our network. The input images utilized to train our network are resized to 224×224. Finally, the dense layers are activated using the sigmoid function. Since this research study in this paper is focused on the classification between PD patients and healthy controls, i.e., a binary classification problem, binary cross-entropy (BCE) [75] is utilized as the loss function. Binary cross-entropy is defined by:

$$BCE = -\sum_{i=1}^{I}\sum_{j=1}^{J} t_{i,j} log(s_{i,j}), \tag{3.2}$$

where $I = 2$ represents the number of classes, J is the number of the training images, $t_{i,j}$ is the binary indicator of value of either 1 or 0 depending on whether or not the class label i is the correct label of sample j, and $s_{i,j}$ is the predicted probability that sample j belongs to class i.

In order to tackle the issue of overfitting, a technique called dropout [39] is used. Dropout works by "turning off" some neurons with a probability "$1-p$", and using

only the reduced network. After this step, the "off" neurons are turned on again with their last weight matrix. This procedure is repeated in every training iteration. At testing time, all neurons are active, so their outputs are weighted by a factor of "p", as the approximation of results using all possible 2^n networks. In this work, we use a dropout probability of 0.1 (i.e., $p = 0.9$) for every convolutional layer and 0.5 for the final dense layers.

Finally, our layer structure also differs considerably from AlexNet. In AlexNet, there are different numbers of nodes in different layers. For example, each of layers one and two has 96 nodes, each of layers three and four has 256 nodes, each of layers five and six has 384 nodes, and each of the last two layers has 256 nodes. In our architecture, we have 128 nodes in each of our layers. To understand the impact of different network architectures on accuracy and provide a reason for why our network has three convolutional layers and two dense layers, and why we have 128 nodes in each layer, we test different combinations of layers in a network architecture. Table 3.1 compares the results of training accuracies achieved with several different combinations, where CL stands for convolutional layer and DL stands for the dense layer. Hence, as seen in Table 3.1, it can be concluded that the architecture with 128 nodes, three convolutional layers and two dense layers performs the best because it provides the highest training accuracy during training.

3.2.2 Image Normalisation

In image processing, normalisation is a process that changes the range of pixel values. Neural networks process input images using small weight values, and images with large pixel values can disrupt or slow down the training process. As such, it is a good practice to normalise the pixel values so that each pixel value is within the range of the difference from the mean value of the same pixel in all images in a dataset. In this process, the mean values of the pixels on images of the whole

Table 3.1 : Training accuracy results of different architectures

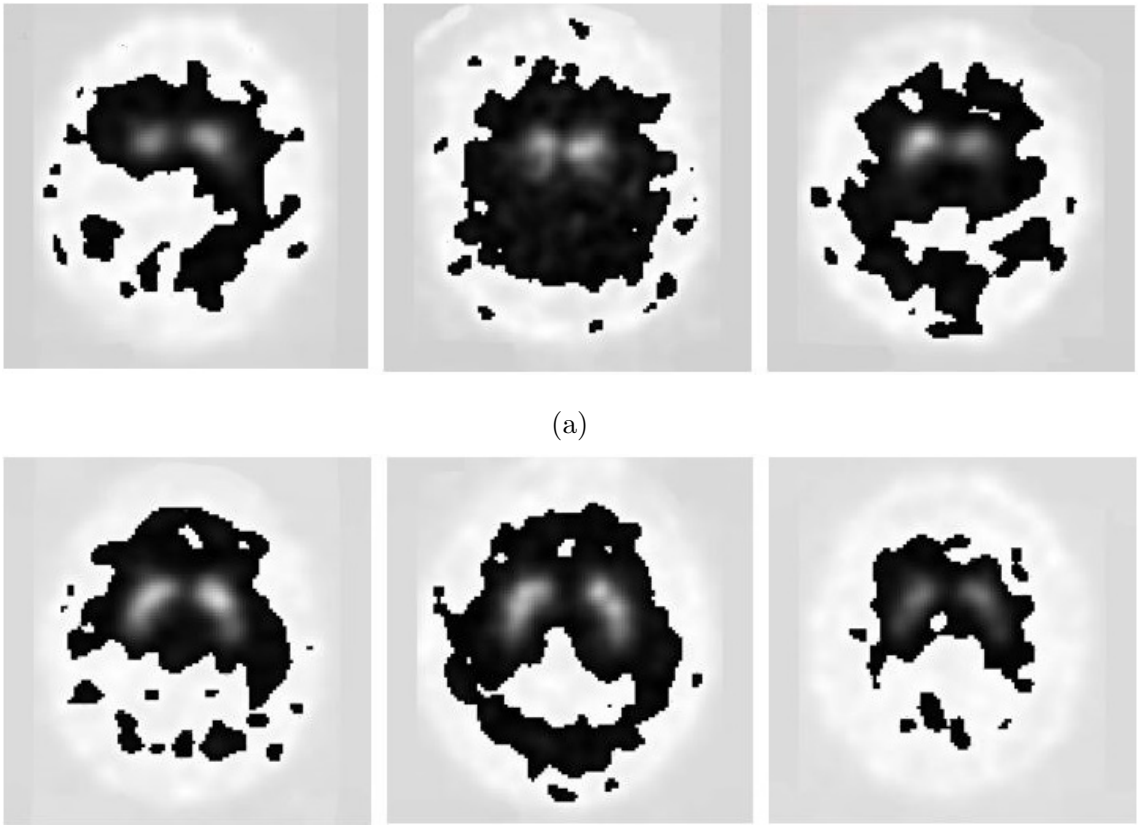
# of Nodes	Number of CLs: 2		Number of CLs: 3		Number of CLs: 4	
	Number of DLs: 1	Number of DLs: 2	Number of DLs: 1	Number of DLs: 2	Number of DLs: 1	Number of DLs: 2
32	97.65%	98.17%	98.03%	97.98%	97.69%	97.77%
64	98.11%	96.87%	98.10%	97.64%	97.93%	97.53%
128	97.97%	97.61%	98.05%	**99.47%**	97.35%	97.87%

dataset are generated first. Then, each image is normalized by subtracting the value of a pixel from the mean of the corresponding pixel values of all images in the dataset. This process reduces the values of each image, and makes the average of pixel intensity values of the new images to zero, centering the images to get stable image gradients. The normalized image is achieved by the following equation:

$$N_i = O_i - (\frac{1}{n} * \sum_{j=1}^{n} O_j), \tag{3.3}$$

where n is the total number of images in a dataset, O_i represents the i-th image in the dataset and N_i is the corresponding normalized image, for $i = 1, 2, \cdots, n$.

The main reason for this process is because, while training the network, multiplication of weights and addition to biases are performed on the initial inputs, in order to cause activations that are then back-propagated to the gradients to train the network model. We want the features of the image to have a similar range so that the gradients do not spray out. Thereby, we achieve one global learning rate multiplier. Moreover, the image normalization can improve network convergence speed without having the adverse effects of pooling and down-sampling, and also reduce training time [76]. Figures 3.3a and 3.3b show the normalised images of a PD and a healthy control, respectively.



(a)

(b)

Figure 3.3 : SPECT scans showing examples of (a) Normalized PD patient and (b) Normalized healthy control

3.3 Experimental Setup

The main aim of the experiment is to classify PD patients from healthy controls using SPECT images and CNNs. Because this research involves binary classification, two classes were identified, i.e., class 0 for healthy control group and class 1 for PD patients. The experiments were performed using Keras model on top of TensorFlow [77]. The Google Brain team created this open-source software solution for machine learning applications on textual data sets. The framework supported running the training operation of the network on graphical processing units (GPUs) or traditional microprocessor-based computer processing units (CPUs). This platform also

supported several machine learning algorithms with the same optimizer. Keras was an open-source neural network library written in Python. It was designed to enable fast experimentation with deep neural networks and focused on being user-friendly, modular, and extensible [78].

3.3.1 Data Augmentation

At the time of this experiment, the number of SPECT images for the healthy control group were only 341 images. With very few images for a particular class, the primary issue was the skewness of the data, which caused an imbalance in the dataset. Therefore, the dataset was not ideal as it would lead to inaccurate results, leading to wrong interpretations.

In order to address this issue, the process of data augmentation was applied to the healthy control group. Data augmentation was a strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data. Several data augmentation techniques such as cropping, padding and horizontal flipping were commonly used to train large neural networks. However, most approaches, which have been used in training neural networks, only used basic types of augmentation [79].

For augmenting our image data, the horizontal flipping technique was utilised. Horizontal filling augmentation was the standard and widely used technique to implement and has been proven useful on other datasets [80]. We have also utilised colour augmentation, which was defined as altering the color properties of the image. For colour augmentation, we have used brightness adjustment along with hue and saturation adjustment. After data augmentation, a total of 1,364 SPECT images for the healthy control group were gathered.

3.3.2 Training and Testing the Network

A total of 2,723 SPECT images were collected to train and test our network. The healthy control group consisted of 1,364 images (341 original images and 1023 augmented images), and PD patients consisted of 1,359 original images. As data augmentation was performed on the healthy control group, all augmented images were utilized for training the network only. 10-fold cross-validation was performed to assess the performance of the model, and it partitioned the entire dataset into 10 subsets. In each fold, the testing subset was selected from the 10 subsets and the remaining nine subsets were used for training. This process was repeated 10 times.

In performing the 10-fold cross-validation, the entire dataset was randomly split by two methods. In **Method 1**, all the images from the PD and healthy controls group (including augmented images) were combined into one dataset. Then, the entire dataset was split randomly into 10 parts of equal size. For **Method 2**, the dataset for three groups of images, i.e., PD, healthy controls with only augmented images and healthy controls with only original and non-augmented images were first randomly split into 10 parts. Then, the corresponding parts from the three groups were combined together to form one equal-sized subset. This method provided 10 equal-sized subsets which was utilized for 10-fold cross-validation.

Note that, in the 10-fold cross-validation process, all augmented data that were originally placed in the testing subsets. However,we show not only the results with the augmented data but also the results without the augmented data.

3.3.3 Evaluation Metrics

The performance measures of this model were evaluated using 10-fold cross-validation. Similar to benchmark studies, we used accuracy, sensitivity and speci-

ficity to measure and compare the classification performance.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{3.4}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{3.5}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{3.6}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives. Note that, in this paper, each PD sample is regarded as a positive sample and each sample of healthy control (HC) is regarded as negative sample.

3.4 Results

The images were passed through the network model and trained sufficiently. As 10-fold cross-validation was performed with two methods, the results of the two methods are presented below. It was observed that, our model quickly converged and achieved a training accuracy of 95.83% with method 1. By performing the 10-fold cross-validation with method 2, our network achieved an excellent training accuracy of 99.47%. Tables 3.2 and 3.3 provides detailed information on the results.

Accuracy: After performing the 10-fold cross-validation by utilizing method 1, it was observed that our proposed network, with augmented images, showed an accuracy result of 95.34%, and without augmented images, achieved an accuracy of 97.47%, respectively. By applying method 2, it was observed that our network, with augmented images, achieved an accuracy result of 99.34%, and without augmented images, achieved an accuracy result of 99.18%, respectively.

These results are by far the highest accuracy results reported in this area of research. From Table 3.3, it can be observed that out of 1,359 PD patients, 1,346 PD patients were correctly classified as PD patient. Similarly, out of 1,364 healthy controls, 1,359 subjects were correctly classified in the healthy control group. However,

out of the original 341 healthy controls, only 1 was misclassified as a PD subject. These results are noteworthy because it proves that our network model can classify PD patients from control group with the highest accuracy.

Sensitivity: The sensitivity results achieved by our proposed network are the highest in comparison to those in [67] and [68]. It can be noted that, from Table 3.3, out of 1,359 PD patients, only 13 PD patients were misclassified. All of the other true positives were accurately and correctly classified as PD patients. From Table 3.4, it can be observed that, our network achieved an excellent overall sensitivity result of 98.75% for method 1, irrespective of whether augmented images or non-augmented images were used. Similarly, our network achieved an overall sensitivity of 99.04%

Table 3.2 : Detailed performance results of the 10-fold using Method 1

	Results of PD group			**Results of healthy control (HC) with augmented images**			**Results of healthy control (HC) without augmented images**			
Fold	**PD**	**Classified as PD**	**Misclassified as HC**	**HC**	**Classified as HC**	**Misclassified as PD**	**HC**	**Classified as HC**	**Mislassified as PD**	**Testing time (sec)**
1	141	138	3	132	124	8	33	31	2	10.23
2	148	146	2	124	112	12	31	28	3	10.37
3	121	120	1	152	136	16	38	34	4	10.25
4	136	135	1	136	134	2	34	34	0	9.86
5	135	135	0	136	124	12	34	31	3	9.45
6	117	112	5	156	147	9	39	37	2	11.17
7	152	152	0	120	119	1	30	30	0	10.35
8	112	110	2	160	152	8	40	38	2	10.12
9	125	125	0	148	128	20	37	32	5	10.54
10	172	169	3	100	78	22	25	20	5	10.17
	1359	**1342**	**17**	**1364**	**1254**	**110**	**341**	**315**	**26**	**102.51**

Table 3.3 : Detailed performance results of 10-fold with Method 2

	Results of PD group			**Results of healthy control (HC) with augmented images**			**Results of healthy control (HC) without augmented images**			
Fold	**PD**	**Classified as PD**	**Misclassified as HC**	**HC**	**Classified as HC**	**Misclassified as PD**	**HC**	**Classified as HC**	**Misclassified as PD**	**Testing time (sec)**
1	136	135	1	136	136	0	34	34	0	10.66
2	136	134	2	137	137	0	34	34	0	10.25
3	136	136	0	136	134	2	34	34	0	9.53
4	136	136	0	136	136	0	34	34	0	10.69
5	136	134	2	137	136	1	34	34	0	9.76
6	136	135	1	136	136	0	34	34	0	9.65
7	136	132	4	136	136	0	34	34	0	9.65
8	135	133	2	137	137	0	35	35	0	10.35
9	136	136	0	136	135	1	34	33	1	10.22
10	136	135	1	137	136	1	34	34	0	10.10
	1359	**1346**	**13**	**1364**	**1359**	**5**	**341**	**340**	**1**	**100.86**

by applying method 2. These results indicated our network's reliability in correctly diagnosing PD patients when compared with the benchmark studies.

Specificity: Our network also achieved the highest specificity result with both augmented and original images. From Table 3.4, it can be seen that with augmented images, the overall specificity score with in method 1 was 91.94%, and with method 2, the specificity score was 99.63%. Similarly, without augmented images, our network achieved an overall score of 92.38% with method 1 and 99.71% with method 2. These results demonstrated that our network can be relied upon to differentiate healthy

Table 3.4 : Performance metrics of the network

	Performance results with Method 1							Performance results with Method 2						
		With augmented images			Without augmented images				With augmented images			Without augmented images		
Fold	**Tr.Accuracy**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Tr.Accuracy**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
1	96.23%	95.97%	97.87%	93.94%	97.13%	97.87%	93.94%	99.71%	99.63%	99.26%	100%	99.41%	99.26%	100%
2	95.97%	94.85%	98.65%	90.32%	97.21%	98.65%	90.32%	99.45%	99.27%	98.53%	100%	98.82%	98.53%	100%
3	94.22%	93.77%	99.17%	89.47%	96.86%	99.17%	89.47%	99.53%	99.26%	100%	98.53%	100%	100%	100%
4	99.16%	98.90%	99.26%	98.53%	99.41%	99.26%	100%	99.79%	100%	100%	100%	100%	100%	100%
5	95.98%	95.57%	100%	91.18%	98.22%	100%	91.18%	98.79%	98.90%	98.53%	99.27%	98.82%	98.53%	100%
6	95.25%	94.87%	95.73%	94.23%	95.51%	95.73%	94.87%	99.72%	99.63%	99.26%	100%	99.41%	99.26%	100%
7	99.07%	99.63%	100%	99.17%	100%	100%	100%	98.55%	98.53%	97.06%	100%	97.65%	97.06%	100%
8	96.91%	96.32%	98.21%	95.00%	97.37%	98.21%	95.00%	99.45%	99.26%	98.52%	100%	98.82%	98.52%	100%
9	93.36%	92.67%	100%	86.49%	96.91%	100%	86.49%	99.77%	99.63%	100%	99.26%	99.41%	100%	97.06%
10	92.13%	90.81%	98.26%	78.00%	95.94%	98.26%	80.00%	99.76%	99.27%	99.26%	99.27%	99.41%	99.26%	100%
	95.83%	**95.34%**	**98.75%**	**91.94%**	**97.47%**	**98.75%**	**92.38%**	**99.47%**	**99.34%**	**99.04%**	**99.63%**	**99.18%**	**99.04%**	**99.71%**

controls from PD patients when compared with the benchmark studies. Table 3.4 presents the results of all performance metrics in each fold using augmented and original images in both methods.

From the above experiments, it can be concluded that method 2 performs better in all cases because the proportions of training and testing data from all three groups (i.e., PD, HC with only augmented data and HC without augmented data) were more evenly distributed into the 10 parts for the 10-fold validation process.

3.5 Discussion and Conclusion

In medical imaging analysis, CNNs have proven to be powerful tools for a broad range of computer vision tasks [38]. Image classification is one of the first areas in which deep-learning techniques make significant contributions to medical image analysis. In medical image classification, multiple images are considered as inputs,

with a single diagnostic result as output. Most interpretations of medical images are performed by physicians. However, image interpretation and pattern recognition by humans have their limitations because of human factors such as subjectivity, significant variations across interpreters and fatigue.

Many diagnostic tasks require an initial search process to detect abnormalities or particular patterns, and to quantify measurements and changes over time. Computerized tools, specifically image analysis and machine learning, are the key enablers to improve diagnosis, by facilitating identification of the findings that require treatment to support the expert's workflow. Among these tools, deep-learning has quickly been proved to be the state-of-the-art foundation, leading to improved accuracy. Therefore, we have adopted deep-learning and SPECT imaging in tackling the challenging task of diagnosing early PD patients from healthy control patients.

In this chapter, we have adopted CNNs and developed a deep-learning based network to classify PD patients from the healthy controls, using SPECT images. Our network has performed exceptionally well with excellent results in accuracy, sensitivity and specificity. The importance of this study lies in the fact that, with such results, the challenge of diagnosing PD patients in their early stages can be diminished. Therefore, this network can be of aid to clinicians in diagnosing PD patients effectively.

We first compare the detection accuracies of using AlexNet and the proposed network without using the augmented data in testing. It is seen that, with AlexNet, the accuracy is only 63.68% in compariosn to our network's 99.18%. The sensistivity and specificity results achieved by AlexNet are also low in comparison with our network. Figure 3.4 illustrates the performance difference between both networks. In regards to computational efficiency, it can be noted that our network, when compared to AlexNet, perform much better with faster training and testing times.

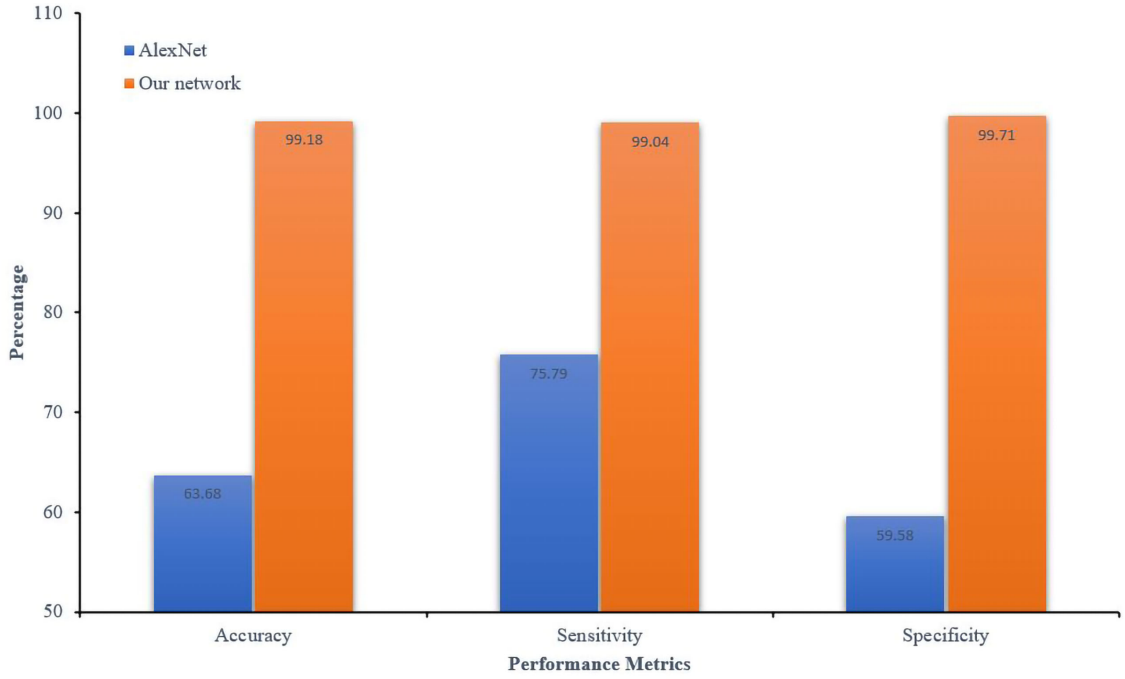


Figure 3.4 : Performance comparison of our network with AlexNet

The detailed performance information for both networks are shown in Table 3.5.

We then compare our approach with the benchmark studies of [67] and [68]. Firstly, in terms of network architecture, the network architecture mentioned in [67] uses two convolutional layers and one dense layer that is activated by a softmax function. However, the kernel size of all convolutional layers in their network is 5×5. For our proposed network, three convolutional layers and two dense layers are utilised, and a sigmoid function activates our dense layers. We utilised the sigmoid function because sigmoind function is better suited for a binary classification [81]. The kernel size of our convolutional layers is 3×3. Secondly, regarding pre-processing the images, in [67], the images are cropped to focus only on the regions of interest (the caudate and the putamen regions) in the images. Then, the network is trained on those cropped images. However, in our approach, entire images are taken into consideration, and the network is trained accordingly. This allows our network to be more robust and efficient in classifying PD patients from healthy controls because

Table 3.5 : Computational cost comparison with AlexNet and our network

	AlexNet		Our network (Method 1)		Our network (Method 2)	
Fold	**Training time(min)**	**Testing time(sec)**	**Training time(min)**	**Testing time(sec)**	**Training time(min)**	**Testing time(sec)**
1	23.67	12.54	20.89	10.23	20.18	10.66
2	23.48	11.68	19.86	10.37	20.45	10.25
3	22.97	10.79	20.39	10.25	19.64	9.53
4	23.99	11.43	20.67	9.86	19.55	10.69
5	22.59	10.17	19.35	9.45	20.21	9.76
6	23.75	11.43	19.12	11.17	19.32	9.65
7	22.89	12.65	19.96	10.35	20.35	9.65
8	23.38	11.81	20.44	10.12	20.16	10.35
9	23.67	10.27	20.38	10.54	19.65	10.22
10	22.19	11.53	20.19	10.17	20.49	10.10
	232.58	**114.3**	**201.25**	**102.51**	**200**	**100.86**

classification is performed based on the whole SPECT image of the brain. Finally, the performance metrics, in comparison to [67], indicate that our network performs better in terms of accuracy, sensitivity and specificity.

In comparison with [68], their network, PD Net, was developed with three convolutional layers with the kernel size of the first layer being 7×7, followed by 5×5 for the second layer, and 3×3 for the last layer. However, our network is designed with three convolutional layers with a kernel size of 3×3. This architecture is utilised because of the high accuracy achieved during training the network (Table 3.1). In terms of performance evaluation, PD Net achieved an accuracy of 96% and sensitivity of 94.2%. Our network outperforms [68] with an accuracy of 99.18% and sensitivity of 99.04%. Figure 3.5 shows the comparison results with the two networks.

Another significant advantage is the minimal complexity of the network, with

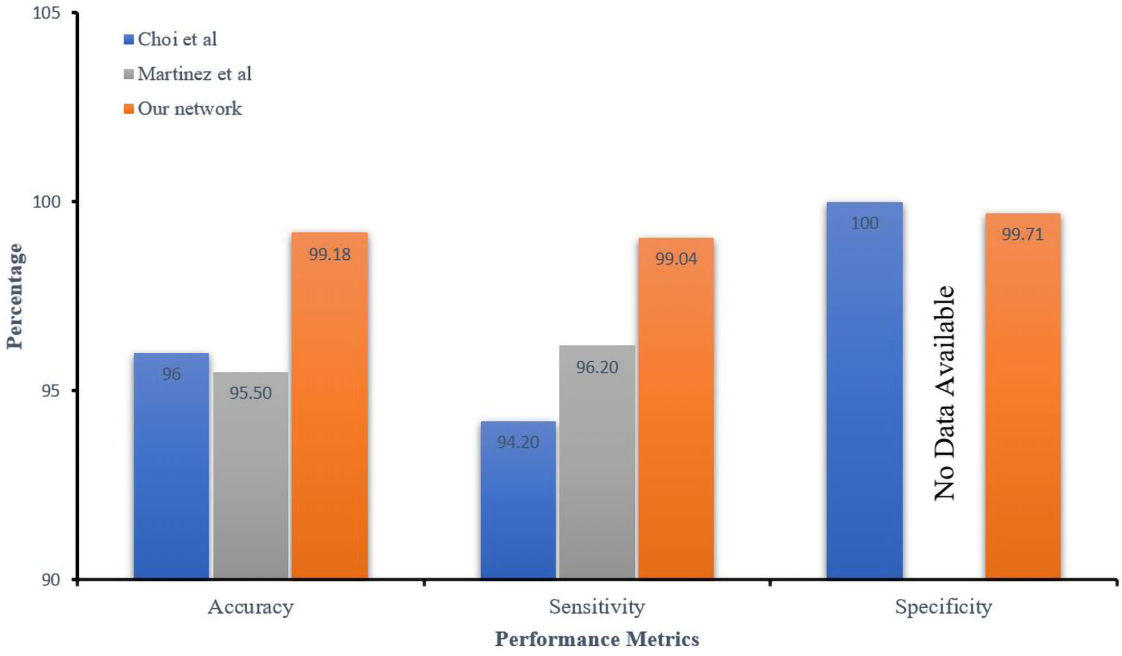


Figure 3.5 : Network performance comparison with benchmark studies

its associated high processing speed. Majority of deep-learning networks are very complex and challenging to implement on a large scale. However, the proposed network in this paper, with only three convolutional layers and a kernel size of 3×3, results in a remarkably reliable and accurate result with an accuracy of 99.18% (Figure 3.5). This is the highest accuracy achieved so far in this area of research. In addition, as shown in Table 3.5, the CNN model takes only 100.86 seconds to accurately classify the entire dataset of 2,723 images, indicating that our network classifies each testing image at an extremely high speed of 0.037 seconds per image. We cannot compare the time with the two state-of-the-art approaches because they were performed on different datasets and they did not share their codes.

Sensitivity and Specificity also show excellent results of 99.04% and 99.71% (Figure 3.5), respectively. Because complicated image feature selection is not required in the classification of SPECT images, this network shows potential in its practicality in a clinical setting. Due to its low network complexity, implementing this network is

feasible on a large scale, as it can be run on any commercially available off-the-shelf machinery. This is important because, for physicians, time is crucial for diagnosis and treatment of PD. They have to be quick and precise in their diagnosis so that they can plan a treatment regimmen for the PD patient. Implementing a time-consuming network for such diagnosis is disadvantageous because it would cause delays in diagnosis or treatment, resulting in poor management of PD symptoms.

Finally, the sample size used for this network is the highest with a total number of 2,723 SPECT images (1,359 images for PD patients and 1,364 for healthy control). With such a large sample size, our network can learn as many features as possible, enabling itself to be more robust and efficient. This, in turn, facilitates the model to distinguish PD patients from the healthy controls accurately.

SPECT scans are efficient tools in diagnosing early PD as they illustrate the dopaminergic deficits in the brain. Experienced nuclear physicians usually evaluate a SPECT scan. However, significant variations exist between individual physicians because of human factors such as lack of practical experience and training, which leads to inconsistent diagnosis among physicians resulting in misdiagnosis in early PD. Our approach eliminates the human factor and learns the features from ROIs in the SPECT images that are necessary for classification. Our experimental results show that, in comparison to benchmark studies, our network's performance is much better with the highest accuracy reported. Our network is also fast because of the minimal architectural complexity allowing for fast and accurate classification. With these results, it can be noted that our network model has a great potential in automatic PD classification, and could improve the diagnosis of PD by being a great asset clinically, to assist clinicians making a precise diagnosis of patients with PD, especially during the pre-motor stage.

However, major limitations have been identified in our study, and they are the

concerns of image resolution of a SPECT scan and the availability of using such technology. SPECT imaging technology is not available in most hospitals because it is nuclear in nature. This requires very demanding support to maintain the materials for the scan. Even if this support is available, doctors do not specifically ask for it because most PD patients, when diagnosed, are in the advanced stages of PD. Another limitation of SPECT is that it is difficult to obtain a reliable quantification. Furthermore, the low resolution of a SPECT image limits the visualization of the basal ganglia in PD [82]. Despite these limitations, SPECT imaging is a powerful tool for PD diagnosis because it has the power to distinguish early PD patients from healthy patients. This is why SPECT imaging has been used in this paper. It can be seen that our network has performed exceptionally well.

In comparison with other published research studies, our CNN-based model outperforms them in terms of performance metrics, confirming that the innovative deep-learning approach that we developed, has its advantages and is competitive. Table 3.6 provides the details of the comparison with other published research studies.

In conclusion, we have processed SPECT images of PD and healthy controls and classified them accordingly. The classification model has been developed using CNN. The performance measures have shown high scores for accuracy (99.34%), sensitivity (99.04%), and specificity (99.63%). Medical practice aspires to diagnose patients at the earliest of clinical signs, in order to monitor disease progression, and to rapidly find optimal treatment regimens. However, for PD, it is a challenging issue because there are patients who exhibit many overlapping clinical indications. Therefore, misdiagnosis in early PD is common. In the present study, we have developed an effective model to address the problem, by accurate identification of features of degenerative PD at its early stage. This enables early diagnosis, which is crucial for effective PD patient management. Based on previous studies, we can confidently claim that our model outperforms not only the previous studies, but

Table 3.6 : Comparison table with previous studies

Research Works	**Dataset Used**	**Sample Size**	**Classifier Used**	**Accuracy**	**Sensitivity**	**Specificity**
Towey et al. [64]	Own dataset	116	Naïve Bayes	94.8%	93.7%	97.3%
Prashanth et al.[59]	PPMI	493	SVM	96.14%	96.55%	95.03%
Oliveira et al.[11]	PPMI	652	SVM	97.9%	98%	97.6%
Prashanth et al.[14]	PPMI	584	SVM	97.29%	97.37%	97.18%
Oliveira and Castelo-Branco [62]	PPMI	445	SVM	97.86%	97.75%	98.09%
Martinez et al. [67]	PPMI	301	CNN	95.5%	96.2%	N/A
Choi et al. [68]	PPMI	431	PD Net	96%	94.2%	100%
This study	**PPMI**	**2723**	**CNN**	**99.34%**	**99.04%**	**99.63%**

also any experienced physician. Since it is simple to use, the potential for this CNN model to be used clinically, in day-to-day PD diagnosis, is enormous. It could also be used by clinicians to observe the deterioration and progression of PD conditions quantitatively.

Chapter 4

Classification of Parkinson's Disease (PD) into Multiple Stages of Progression

PD occurs mostly in age 50 and above, and is difficult to identify in the earlier stages. The onset of PD is characterised by cardinal motor impairments like tremor, rigidity and bradykinesia. However, as the PD progresses, the symptoms increase enormously, varying significantly at different stages of progression. Therefore, the diagnosis of the patient's PD stage is critical for improving the quality of patient's life. In this chapter, we address this issue by developing a deep-learning model, namely a deep convolutional neural network for PD stage classification (PDStageNet), that is highly capable of identifying and classifying PD patients into 5 clinical stages of PD progression based on SPECT images.

4.1 Introduction to PD Stages

PD impacts people in many different ways. Not all patients will experience all of the symptoms of PD, and if they do, they will not necessarily experience them in quite the same order, nor at the same level of intensity. Even so, there are typical patterns of progression in PD that are defined in stages. PD is classified into five different stages described below.

- Stage One. During this initial stage, the patient has mild symptoms that generally do not interfere with daily activities. Tremor and other movement symptoms occur on one side of the body only. Friends and family may notice changes in posture, walking and facial expressions.

- Stage Two. In this stage, the symptoms begin to advance. Tremor and other movement symptoms affect both sides of the body. Walking problems and poor posture may become apparent. In this stage, the patient is still able to live alone, but completing day-to-day tasks becomes more difficult and may take longer.

- Stage Three. This stage is considered as the mid-stage in the progression of the disease. Loss of balance and slowness of movements and falls are hallmarks of this phase. Although the patient is still fully independent, symptoms significantly impair activities of daily living such as dressing and eating.

- Stage Four. During this stage, symptoms are severe, limiting the bodily movements of the patient. It is possible to stand without assistance, but movement may require a walker. The patient needs help with activities of daily living and is unable to live alone.

- Stage Five. This stage is the most advanced and debilitating stage of PD. Stiffness in the legs may make it impossible to stand or walk. The patient either requires a wheelchair or is bedridden. Around-the-clock nursing care is required for all activities. The patient may experience hallucinations and delusions [83].

According to a study conducted by Deloitte Access Economics Australia [15], it was reported that in 2014, the PD estimates by stages were as follows. Figure 4.1 illustrates the increments in PD incidence in Australia.

- 55,900 patients in the initial stages of PD (Stages 1 to 3) compared with 44,300 in 2005.

- 9,100 patients in the intermediate stage of PD (Stage 4) compared with 7,100 in 2005.

- 4,200 patients in the end stage of PD (Stage 5) compared with 3,300 in 2005.

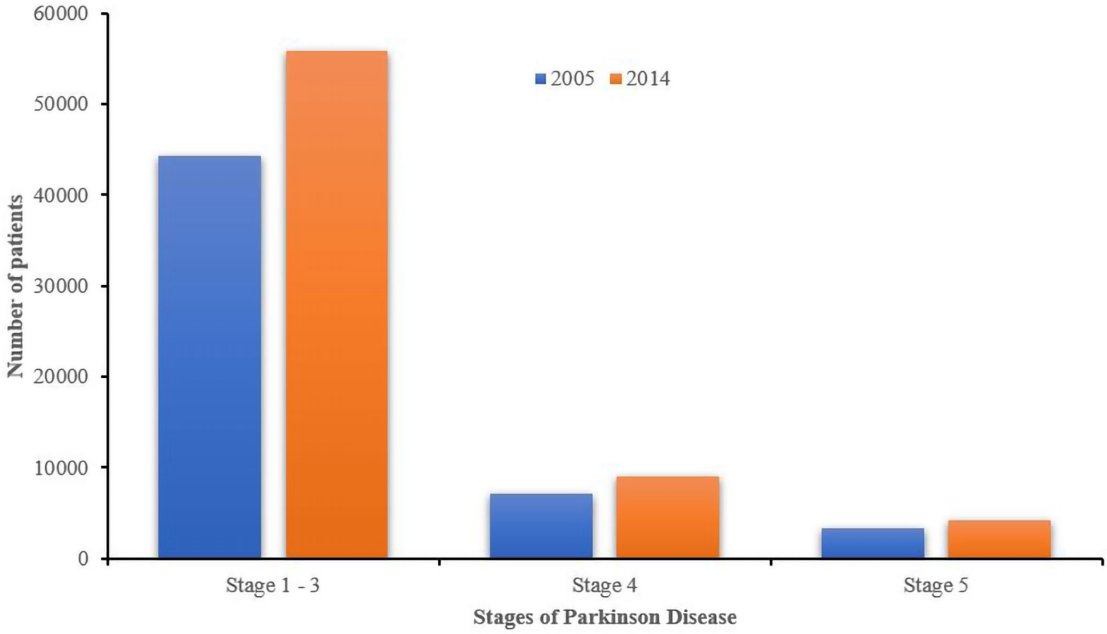


Figure 4.1 : Incidence of PD by disease stage

The clinical diagnosis is definitive for the advanced stages of the disease when the symptoms are fully developed. However, in the early stages, when the symptoms are mild/incomplete or subtle, an accurate diagnosis becomes difficult. Early and accurate diagnosis of PD patients in their respective stages of PD is crucial for reasons such as early management, avoidance of unnecessary medical examinations and therapies and their associated financial costs, side-effects and safety risks [13].

In this chapter, we adopt CNNs to classify PD patients, using SPECT images, into their respective stages of PD progression. This is the first research study where CNNs were adopted for multiple stage classification of PD by utilising SPECT scans. However, there are other published works that performs similar classifications. The differences lie either in the methodology or in the type of data.

In [84], statistical approaches like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were undertaken to extract vocal features nec-

essary for the multiple stage classification of PD. SVM was then used to perform the classification. The vocal data was gathered from the University of California-Irvine (UCI) repository, which included 31 subjects consisting of 8 healthy controls and 23 PD patients. Four classifiers, Support Vector Machine (SVM), Adaptive Boosting (AdaBoost), k-Nearest Neighbor (k-NN) and Adaptive Resonance Theory-Kohonen Neural Network (ART-KNN), were then used to perform the classification and then compared based on accuracy results. It was noted that SVM performed the best.

In [85], Recurrent Neural Network (RNN) was utilised to learn clinical similarities of PD patients. Classification was carried out using Linear regression (LR) and SVM. RNN was a feed-forward neural network that computed a fixed sequence of learned non-linear transformations to convert an input pattern into an output pattern. This enabled the network to perform sequential prediction. In this study, the data was collected from the PPMI database. The data comprised of features from motor symptoms, cognitive functioning, psychotic symptoms, sleep problems and depressive symptoms. It also included Hoehn and Yahr (HY) scale scores for multiple class classification. After feature learning, LR and SVM were utilised to perform classification. It was seen that SVM combined with RNN performed better.

4.2 Proposed Network Approach

The complete layout of our proposed approach, PDStageNet, is illustrated in Figure 4.2. We utilised SPECT images obtained from PPMI database for our experiments. All of the images were pre-processed images that can be used for experiments.

To categorise the images into different stages of PD, we utilised the Hoehn and Yahr (HY) scale rating available with the clinical data of the PPMI database. The HY scale had been published in 1967 and has been used to describe the five clinical stages of PD in an indicidual. It has been a widely used clinical rating scale, which

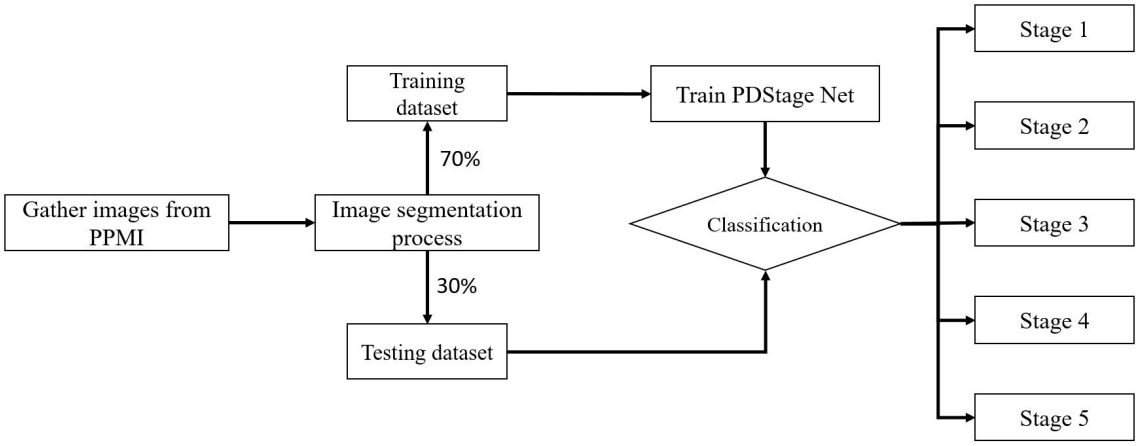


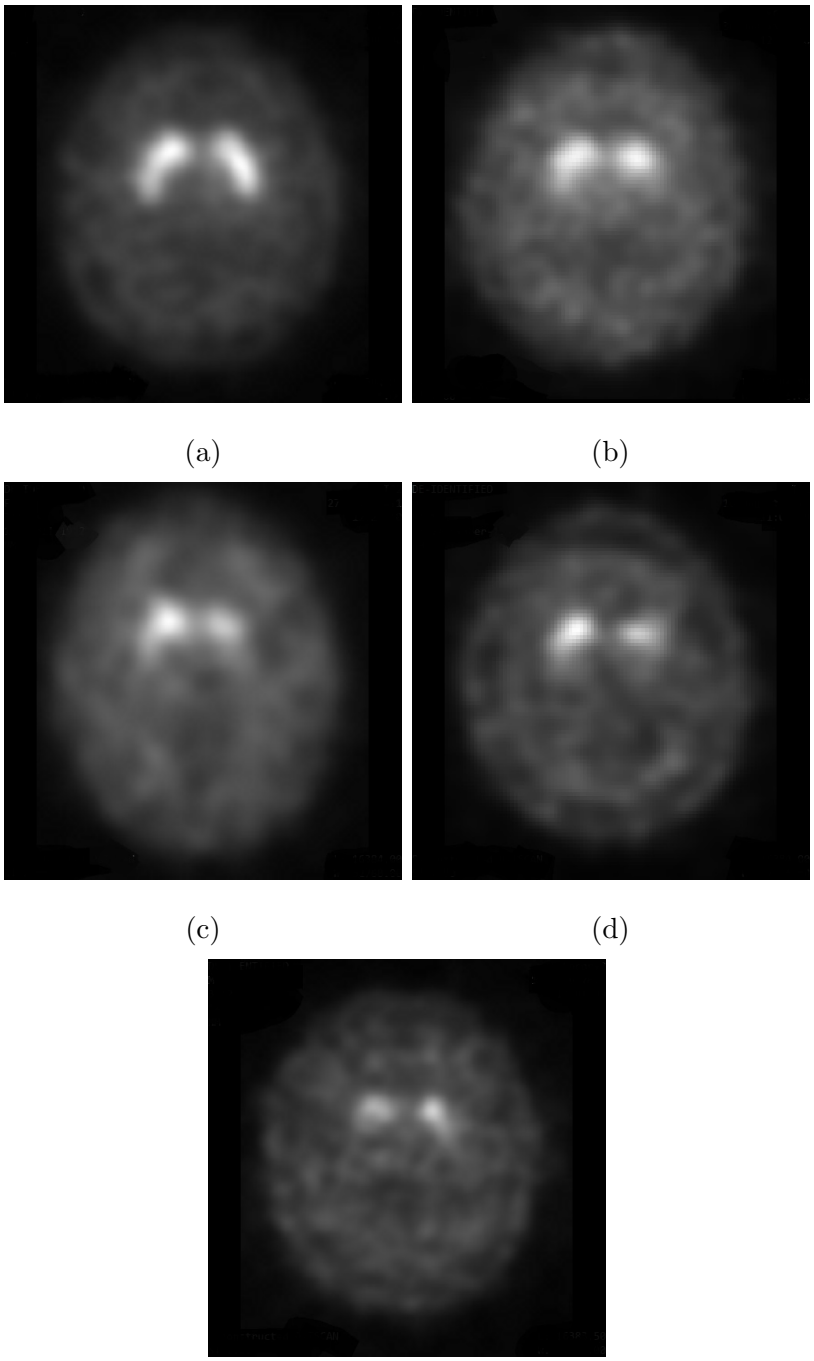
Figure 4.2 : Framework including trainnig and testing for PD stage classification

defined broad categories of motor function in PD. The scale captured typical patterns of progressive motor impairment, which can be applied whether or not patients were receiving therapy. Progression in HY stages has been found to correlate with motor decline, deterioration in the quality of life, and neuroimaging studies of dopaminergic loss [86].

In the database, the patient's clinical data was identified with a unique ID, which was matched with its respective SPECT image. This allowed the images to be labelled according to the HY scale. For example, if the HY scale showed a value of 1, the corresponding SPECT image was categorized into stage 1. This process was repeated until all the images were categorized accordingly, providing us with a total of 1,319 images, of which 217 images were in Stage 1, 967 images in Stage 2, 101 images in Stage 3, 24 images in Stage 4 and 10 images in Stage 5. Figure 4.3 illustrates SPECT scans for all 5 stages. Table 4.1 provides a description on the severity of PD stages based on the HY scale.

4.2.1 Image Segmentation Process

An image binarisation technique is utilised to segment all SPECT images availabled in the dataset. This technique enhances the quality of the two regions of



(a) (b)

(c) (d)

(e)

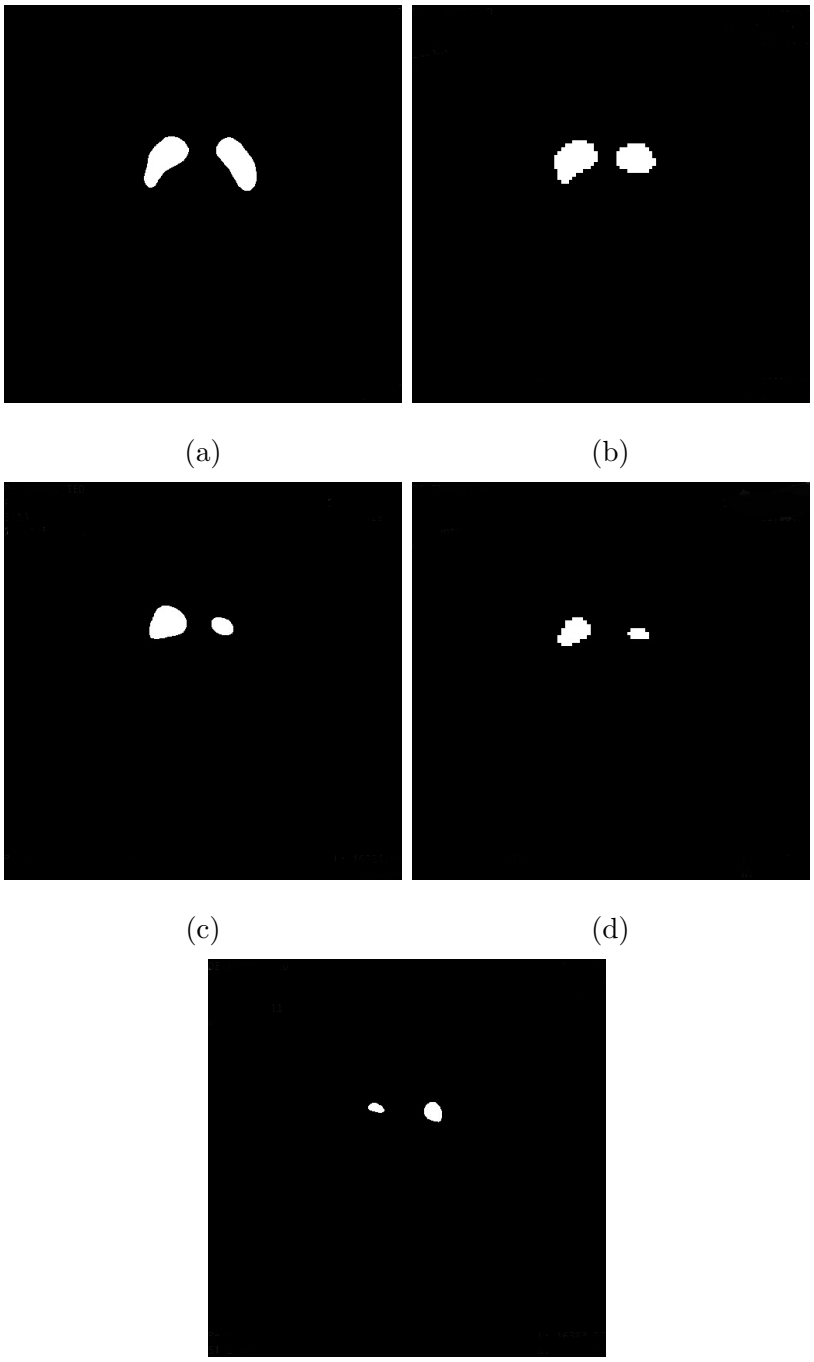
Figure 4.3 : SPECT images of PD in (a) Stage 1, (b) Stage 2, (c) Stage 3, (d) Stage 4, and (e) Stage 5

Table 4.1 : PD Stage characteristics based on HY scale

PD Stage	**HY scale rating**	**Characteristics**	**Number of images**
1	1	Unilateral involvment only. Tremor in one limb, minimal changes in posture, locomotion and facial expression.	217
2	2	Bilateral involvement without impairement in balance. Posture instability and gait	967
3	3	Significant slowing of bodily movements and difficulties to perform daily activities	101
4	4	Severe symptoms. Limited walking, rigidity and bradykinesia. Individual is unable to live alone and requires help in performing daily activities	24
5	5	Cachetic stage. Individual is restricted to bed or a wheelchair unless aided. Can suffer from delusions and halluciantions	10

interests (ROIs) located in the 'substantia nigra' of the brain. These ROIs depict the reduction of DAT availability in a PD patient, and they are important and necessary features for PD stage classification.

In this process, all SPECT images are passed through the binarisation process where a mask is applied on every image. Then, the threshold value is adjusted and matched accurately to the original SPECT image. As different stages of PD progression have different shapes and sizes, the threshold values also differ. Therefore, by adjusting the image threshold values, we are able to generate accurately segemented images of all SPECT images. Finalyy, by utilising the patient's SPECT information and the corresponding HY rating values, provided in the clinical dataset, all of the segmented images are accurately labelled to the correct stages of PD progression. Figure 4.4 illustrates the segmented images of the original images in Figure 4.3.



(a) (b)

(c) (d)

(e)

Figure 4.4 : Segmented SPECT images of PD in (a) Stage 1, (b) Stage 2, (c) Stage 3, (d) Stage 4, and (e) Stage 5

4.2.2 Network Architecture

Our network, PDStageNet, is a smaller and more compact version of the VGG-16 network [87]. VGG-16 network is 16 layers deep and can classify images into 1,000 generic object categories, such as mouse, keyboard, pencil etc. As a result, VGG-16 has learned rich feature representations for a wide range of images. This is very beneficial for our network as the new learned features will have more weights in them, allowing for more robust classifications. The major difference between VGG-16 and our network lies in the number of layers. VGG-16 has 16 layers, so it is exceptionally slow to train the entire network. The entire network architecture weights are large in terms of disk/bandwidth. Due to its depth and number of fully-connected nodes, deploying VGG-16 is a time-consuming and not feasible to use in a medical domain.

In comparison, our network has only 5 convolutional layers, where the first convolutional layer is equipped with max-pooling layer along with the Rectification Linear Unit (ReLU) activation layer [41]. Then, the next two convolutional layers are stacked, and ReLU and the max-pooling layers follow them. This similar combination is repeated for the final two convolutional layers. Furthermore, our network has one dense layer along with a sigmoid layer in comparison to VGG-16's three dense layers. As our network architecture is smaller in size, the implementation of our network is much easier and less time-consuming. Therefore, by addressing the issue of practical implementation in the medical domain, our network can be utilised as a computer-aided diagnostic system, where it can assist in the diagnosis of the stages of PD with much less time possible and with high accuracy. Below we describe our network structure layer by layer.

- Our initial convolutional layer has 64 filters with a 3×3 kernel. ReLU activation is utilised by batch normalisation. Our pooling layer uses a 3×3 pool

size to reduce the spatial dimensions quickly from 100×100 to 33×33 (we are using $100 \times 100 \times 3$ input images to train our network).

- Then, we stack two convolutional layers together (prior to reducing spatial dimensions of volume). This allows our network to learn a richer set of features. We do not increase our filter size, but we decrease our pooling size from 3×3 to 2×2 to ensure that we do not reduce the dimensions too quickly.

- For the last two convolutional layers, they are also stacked together but with increased filter size (from 64 to 128). The deeper we go in the network, the smaller the spatial dimensions are, and the more features we learn. However, our pooling size remains the same.

- Finally, we add our dense layer along with the sigmoid activation layer as this is a multi-class classification.

Another notable difference of the proposed network from the VGG-16 network is the image segmentation process. In our network, the image segmentation process is carried out during the pre-processing stage. By performing this process, the network is trained on the segmented images generated, along with the original image, so that more distinguishing features can be learned during the training phase. Figures 4.5a and 4.5b illustrate the architectures of VGG-16 and the proposed network, respectively.

4.3 Experimental Setup

The main aim of this chapter is to classify PD patients into different stages of PD from SPECT images by adopting CNNs. In this experiment, five classes were identified, each for its corresponding PD stage. The experiments were performed using Keras library on top of TensorFlow [77]. The Google Brain team created this

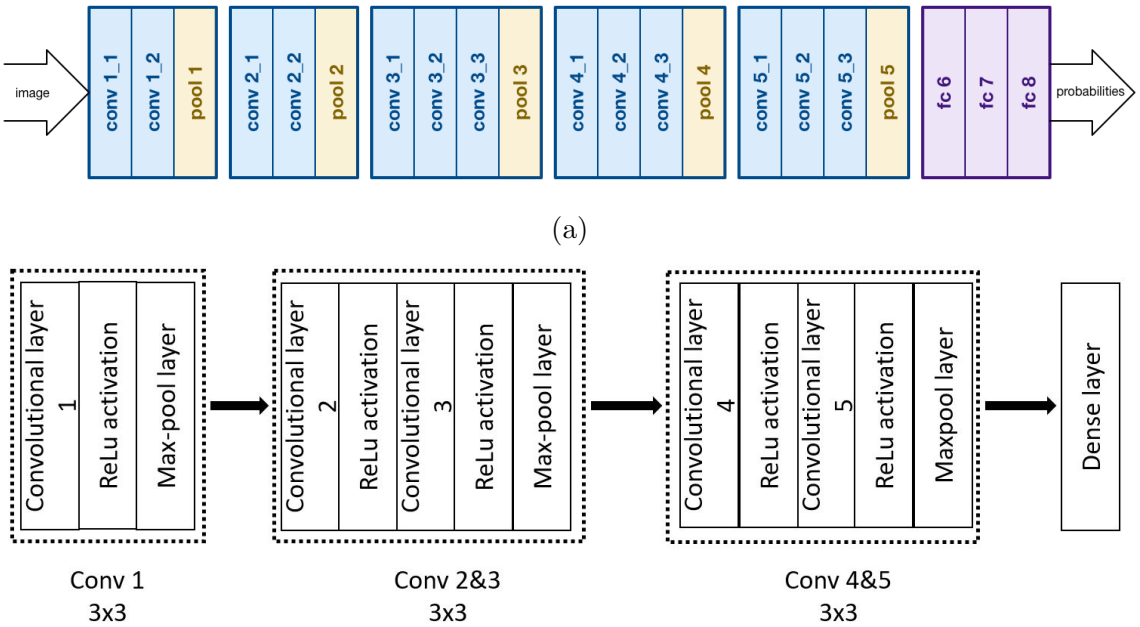


Figure 4.5 : Network Architectures of (a) VGG-16 and (b) PDStageNet

open-source software solution for machine learning applications on textual data sets. The framework supported running the training operation of the network on graphical processing units (GPUs) or traditional microprocessor-based central processing units (CPUs). This platform also supported several machine learning algorithms with the same optimizer. Keras is an open-source neural network library written in Python. It is designed to enable fast experimentation with deep neural networks and focused on being user-friendly, modular and extensible [78]. Keras library was used to develop the network model necessary for our experiments. All of the important packages required for developing the network model such as max-pooling, activation, flatten, dense were all imported from Keras.

4.3.1 Training the network

A total of 1,319 SPECT images were analysed in this chapter. A random split of 70-30 was carried out to divide the database in training and testing datasets.

This provided 929 images for training the network, and 390 images for testing. Our CNN-based deep-learning model, PDStageNet, was trained on the segemented SPECT images. The input images were first resized to 100×100. Then, they were converted into a Keras compatible array followed by appending the image to a data list. The labels were extracted from the image file path and appended to the label list. LabelBinarizer was use to binarise the labels. LabelBinarizer was a utility class that helped create a label indicator matrix from a list of multi-class labels. LabelBinarize utilised a onve-vs-all learning technique, so, in training, it learned one binary classifier per class. In doing so, the multi-class labels were converted to binary laebsl (belonging or not belonging to the class). The LabelBinarizer indentified the total number of classes in the dataset by reading every image label from the training dataset. Then, for every image that passed through PDStageNet, their corresponding labels were extracted and transformed into an array of binary labels which represented whether the image belonged to that particular label or not.

```
>>> from sklearn import preprocessing
>>> lb = preprocessing.LabelBinarizer()
>>> lb.fit([1,2,5,4,3,1,5,4,2,2])
LabelBinarizer()
>>> lb.classes
array([1,2,3,4,5])
>>> lb.transform([1,4,3,5,2])
array([1,0,0,0,0],
       [0,0,0,1,0],
       [0,0,1,0,0],
       [0,0,0,0,1],
       [0,1,0,0,0])
```

Figure 4.6 : Example of LabelBinarization process

As an example, the LabelBinarization process is shown in Figure 4.6. In this process, random images from different classes, with class labels 1, 2, 5, 4, 3, 1, 5, 4, 2 and 2, repsectively are loaded. The LabelBinarizer would read every image class label, and produce an array listing the labels of the classes of the images loaded. In this example, the array is [1,2,3,4,5], representing all five classes. Then for new input images, their corresponding labels are extracted. Finally, using the total number of classes as the array dimension, the LabelBinarizer transforms each extracted label into a binary array with components equal to 1 or 0 depending on whether the corresponding image belongs to the class of the label or not. In the example shown in the figure, the input images have the labels of 1, 4, 3, 5 and 2. Therefore, the LabelBinarizer then transforms the labels to the arrays of [1,0,0,0,0], [0,0,0,1,0], [0,0,1,0,0], [0,0,0,0,1] and [0,1,0,0,0], respectively. The output of this process is a file that contains all learned vectors of multi-class labels in a binarised format.

Our network model was optimised using the Adam optimizer [74] with default learning rate of 10^{-3}. The learning rate was decayed exponentially with a deacy rate of 10^{-8}. These standard parameters of Adam have been proven for neural networks to be computationally efficient, of little memory usage and well suited wor problems with big data. Hence these parameters were utilised in our network. As our research was a multi-class classification, the loss fucntion utilised was the categorical cross-entropy. Categroical cross-entropy(CCE) can be defined by:

$$CCE = -\sum_{c=1}^{M} y_{s,c} log(p_{s,c}), \tag{4.1}$$

where M is the total number of classes, $y_{s,c}$ is the binary indicator of value either 1 or 0 depending on whether or not the class label c is the correct label of sample s, and $p_{s,c}$ is the predicted probability that sample s belongs to class c. Finally, PDStageNet was trained over 1000 epochs with a batch size of 32. The pseudocode for training is provided.

Algorithm 1: Training PDStageNet

Parse the following command line arguments

Path to input dataset

Path to trained output model

Path to output label binarizer

Set number of epochs, learning rate, batch size and image dimensions

epochs = 1000;

InLr = 1e – 3;

BS = 32;

Imgdims = 100, 100, 3;

Initialize data list and labels list to hold preprocessed images and labels

for *every image in dataset* **do**

Read the image

Resize image according to Imgdims;

Convert resized images to Keras compatible array;

Append the images array to the data list;

Extract class labels from image file path and append to labels list;

end

Convert data list and label list into Numpy array

Initialize Label Binarizer;

Assign a unique value to each label and update label list

Initialize training stage

Input data list and labels list

Input optimizer as Adam(Lr=InLr, decay =1e – 8)

Train model with data list and label list;

Use categorical cross-entropy as loss function, Adam, batch size, epochs and calculate training accuracy

Save trained model in Keras

4.3.2 Evaluation metrics

As this study involves multi-class classification, the accuracy of the model was calculated differently compared to binary classification. For our model, the overall accuracy is calculated and is formulated by:

$$Accuracy = \frac{TP_{S1} + TP_{S2} + TP_{S3} + TP_{S4} + TP_{S5}}{N_{TI}}, \tag{4.2}$$

where TP_{S1} is the number of true positives for stage 1, TP_{S2} is the number of true positives for stage 2, TP_{S3} is the number of true positives for stage 3, TP_{S4} is the number of true positives for stage 4, TP_{S5} is the number of true positives for stage 5 and N_{TI} is the total number of images in the testing dataset.

In order to further understand the classification results of individual stages, precision is calculated. Precision is the measure of accuracy, provided that a specific class has been predicted. Finally, sensitivity and specificity are calculated as well in order to grasp our network model's potential to correctly classify particular images in the dataset. The above-mentioned metrics are defined as follows.

$$Precision_s = \frac{TP_s}{TP_s + FP_s}, \tag{4.3}$$

$$Sensitivity_s = \frac{TP_s}{TP_s + FN_s}, \tag{4.4}$$

$$Specificity_s = \frac{TN_s}{TN_s + FP_s}, \tag{4.5}$$

where s represents a specific stage and TP_s, TN_s, FP_s and FN_s are the number of true positives, true negatives, false positives and false negatives of the corresponding stage.

It is critical to note that, the definitions of TN_s, FP_s and FN_s are different, as this is multiple stage classification. They are described below.

- In a confusion matrix, the total number of TN_s, for class s, is the sum of all columns and rows excluding that class' column and row. For example, in stage

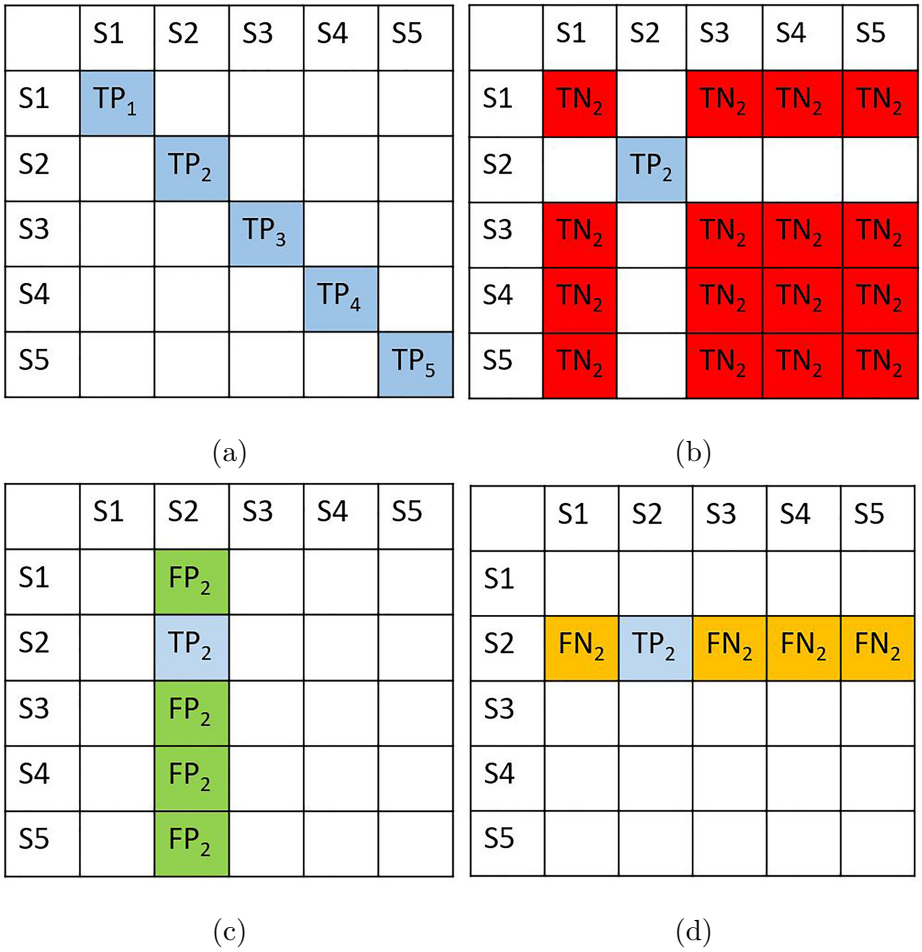


Figure 4.7 : Examples of (a) TP, (b) TN, (c) FP, and (d) FN for a multi-class classification

2, the total number of TN_2 is the sum of all columns and rows except that of stage 2 are calculated, as shown in Figure 4.7b.

- Similarly, the total number of FP_s, for class s, in a confusion matrix, is the sum of all values in the corresponding column, excluding the TP_s. Considering stage 2 as an example, the total number of FP_2 is the sum of all values in the corresponding column, excluding TP_2, as depicted in Figure 4.7c.

- Finally, the total number of FN_s for class s is the sum of all values in the corresponding row, excluding the TP_s. If we take stage 2 as an example,

the total number of FN_2, is the sum of all values in the corresponding row, excluding TP_2, as shown in Figure 4.7d.

4.3.3 Testing the network

For testing PDStageNet, 390 SPECT images were utilised. All of the images were passed through the PDStageNet model. At first, every image was resized to 100×100 to match the image dimensions of PDStageNet. Then, the resized images were converted into a NumPy array. Next, PDStageNet was initialised for testing. Then, label binariser file was read by PDStageNet. During the testing process, for every testing image array, PDStageNet generated output predictions based on the features learned. These predictions were then compared with the binarised labels. The output predictions that mathced the labels with the highest probability resulted in classifcation. The pseudocode for the testing phase is provided.

4.4 Results

Based on the results after testing the network, as shown in Table 4.2, it can be deduced that our proposed network, PDStageNet, achieved an overall high accuracy of 96.67%, which implied that, out of 390 images used for testing, PDStageNet accurately classified 377 images into their respective stages.

Figure 4.8 presents the confusion matrix of the experimental results, which provides a better understanding of sensitivity, specificity and precision results. Regarding sensitivity and specificity results, our proposed network achieved an overall result of 96.31% for sensitivity and 99.25% for specificity. Individual stage sensitivity, specificity and precision results are calculated separately, and the results are shown in Table 4.3.

With such high accuracy, sensitivity and specificity, our proposed network has demonstrated to be a powerful tool in the classification of PD patients into their

Algorithm 2: Testing PDStageNet

Parse the following command line arguments

Path to trained PDSageNet

Path to label binarizer file

Path to testing dataset

Load testing dataset;

for *every image in testing dataset* **do**

  Resize image (100×100);

  Convert resized image to Keras compatible array;

  Save as a numpy array;

end

Call trained PDStageNet using Keras;

Open and read label binarizer file;

Initialize testing of images;

for *every image array* **do**

  Generate output predictions of testing images based on features learned by PDStageNet

  Utilize binarized labels and compare output predictions;

  Print classification probability on the image based on matched output predictions and labels

end

respective stages. This can be beneficial in clinical practice, where our network can be utilised in diagnosing PD patients in a time-efficient manner, so that the physicians can plan out the patients' treatment regimen and begin managing theur symptoms immediately.

Table 4.2 : Results of the network model

Stages	**# of testing images**	**Correctly classified**	**Misclassified images**
Stage 1	62	62	0
Stage 2	288	276	12
Stage 3	30	30	0
Stage 4	7	6	1
Stage 5	3	3	0
Total	390	377	13

4.5 Discussion and Conclusion

Accurate diagnosis of PD in the early stages and differential diagnosis of PD at any stage are challenging and critical medical conditions, because there are patients

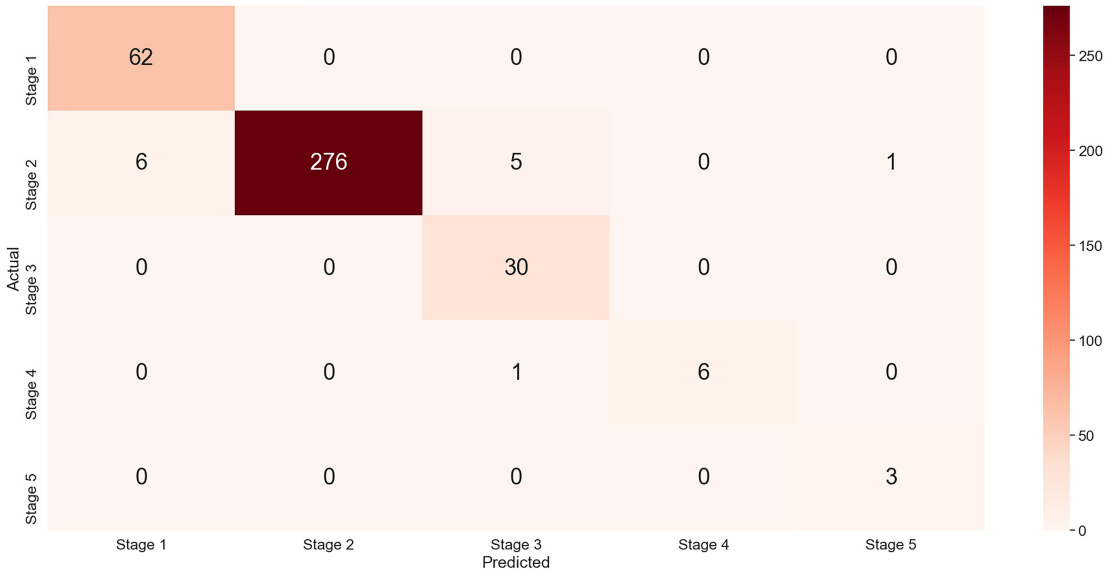


Figure 4.8 : Confusion matrix of the study

Table 4.3 : Individual stage metrics of the network model

Stages	**Sensitivity**	**Specificity**	**Precision**
Stage 1	100%	98.17%	91.18%
Stage 2	95.83%	100%	100%
Stage 3	100%	98.33%	83.33%
Stage 4	85.71%	100%	100%
Stage 5	100%	99.74%	75%

who share, while in different stages of PD progression, many non-classical, and overlapping clinical symptoms. Because there are no conclusive clinical tests, patients with early PD may not meet the clinical diagnostic criteria, leading to delays in their treatment until they are in the advanced stages of PD.

According to John Hopkins Medicine [88], in 2011, the U.S. Food and Drug Administration approved SPECT scan for diagnosing PD in the early stages. SPECT images are powerful tools because they can clearly show the progression of PD by depicting the presynaptic dopaminergic deficits in the striatum. These images are usually analysed through a visual evaluation by experienced nuclear medicine physicians. In most cases, a simple visual inspection is sufficient to confirm or reject the reduction of dopamine in the basal ganglia. However, the diagnosis may not be obvious, especially for early-stage PD patients. Besides, significant variations exist between individual physicians due to the differences in experience and training. Therefore misdiagnosis in early PD is common. By adopting machine learning techniques, these issues in diagnosing PD patients can be easily reduced.

In the current study, we have adopted a CNN-based a deep-learning model, PDStageNet, to accurately classify PD patients into various stages of PD using

SPECT images. Our network model idnetifies the shape and size of the region as essential characteristics and learns them as features important for multiple stage classification. Our network model has been trained on 929 SPECT images and tested on 390 images. The results indicate that our network has achieved an overall high accuracy of 96.67%. With our network model's escellent performance, the challenging task of diagnosing PD patients into their respective stages becomes easier for clinicians, thereby reducing the time required to plan treatment regimens. While the accuracy of our network model is useful, sensitivity and specificity are more appropriate metrics, since it only penalizes false negatives and positives.

Sensitivity and specificity of our network model have also been used to measure the performance of our network model. The overall sensitivity of our network model is 96.31%. In case of specificity, our network has also performed well, achieving an overall specificity score of 99.25%. This is noteworthy because these results demonstrate that our network model is more dependable and reliable in accurately diagnosing PD patients into different stages of PD progression.

As this is a multiple-stage classification, the performance of our network to accurately classify the stage of PD progression is critical. In that aspect, the performance metrics for different stage classification are calculated, and the results are shown in Table 4.3. It can be seen that our network performs well, where each stage demonstrates high scores for precision, sensitivity and specificity. As sensitivity and specificity focus on the false positive and negatives, it can be seen that our network has the highest sensitivity scores for stages 1, 3 and 5, which indicates that our network can accurately classify all the images from these stages. Regarding specificity, it can be noted that stages 2 and 4 have achieved the highest specificity results. This because of the fewer number of false positives corresponding to those classes in our network. Additionally, not only did our network classify the images accurately, it classified them with a very high probability level. All correctly classified images

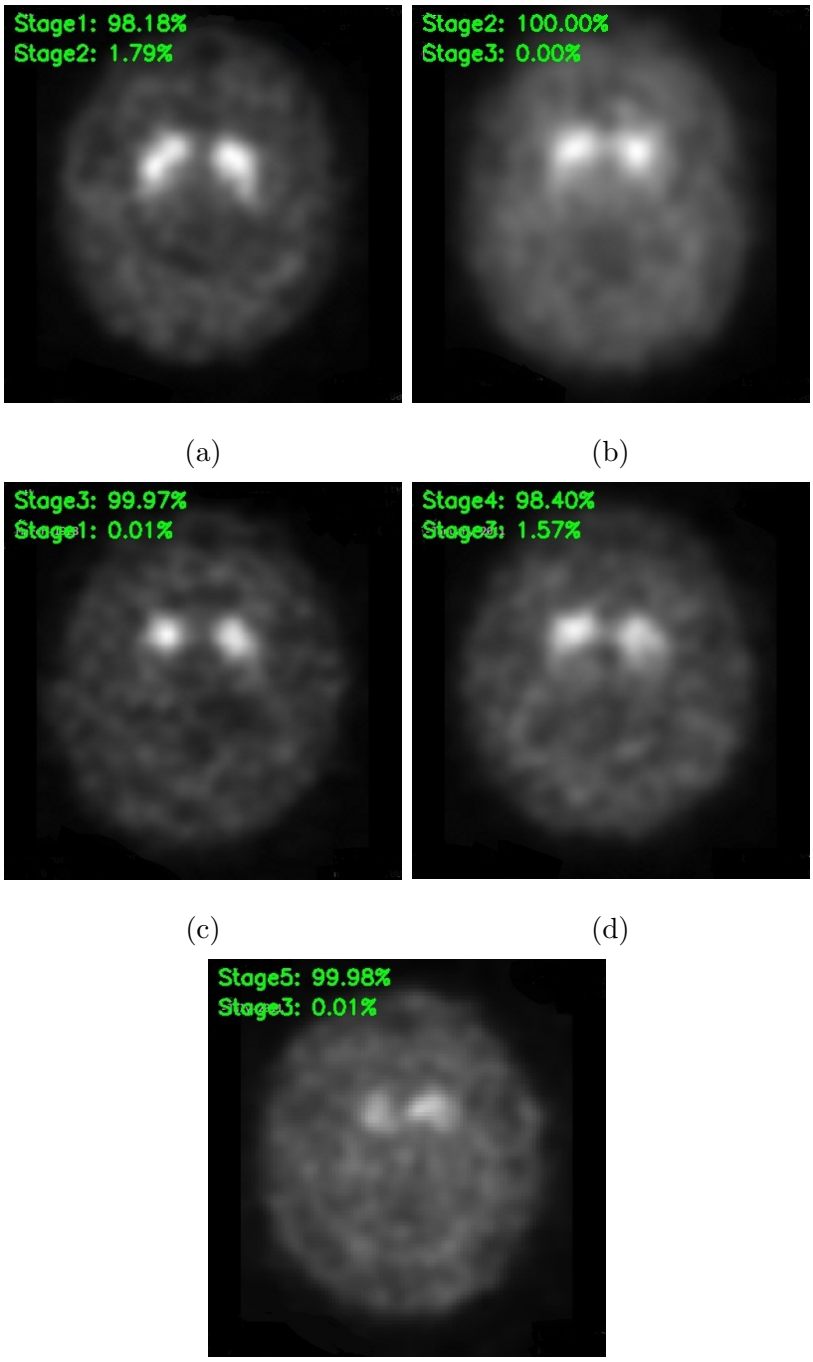


Figure 4.9 : Example of high confidence classification of PD in (a) Stage 1 (b) Stage 2 (c) Stage 3 (d) Stage 4 and (e) Stage 5

have been classified with a probability level of 98% and above, which indicates that our network model has sufficiently learned enough features from the SPECT images to accurately distinguish patients in different stages of PD progression. Figure 4.9 provides examples of such results.

However, our network model has some limitations. Firstly, an unbalanced dataset was used during the training stage. At the time of this study, PD patients in stage 2 had the highest number of SPECT images, whereas stage 5 had the lowest number of SPECT images. This discrepancy resulted in a high number of misclassifications in stage 2. Secondly, the higher number of misclassifications in stage 2 was due to similar shapes and sizes of the ROIs. It was observed that the ROIs of late stage 1 and early stage 2 were quite similar, causing the misclassifications. Similarly, the patients bordering in late stage 2 and those in early stage 3 shared the same characteristics. Thirdly, another limitation noted in our study was the low probability classification level for some SPECT images. For example, in Figure 4.10, the images belong to stage 2. However, it can be noticed that the shapes of the ROIs are similar to that of a stage 1 image. Although they have been correctly classified as stage 2, we can see that the classification probability level is close to 50%. Despite these limitations, our network has performed well, in this particular dataset, with an overall high classification accuracy of 96.67%, sensitivity of 96.31%, specificity of 99.25% and precision of 89.90%. These results confirm that the deep-learning approach developed has its advantages, and is effective and is outweighing the limitations mentioned above.

4.5.1 Comparison Analysis with Prior Works

This is the first research study where CNNs were adopted for multiple stage classification of PD by utilizing SPECT images. However, there are other published works that perform similar classifications. The differences lie either in the methodol-

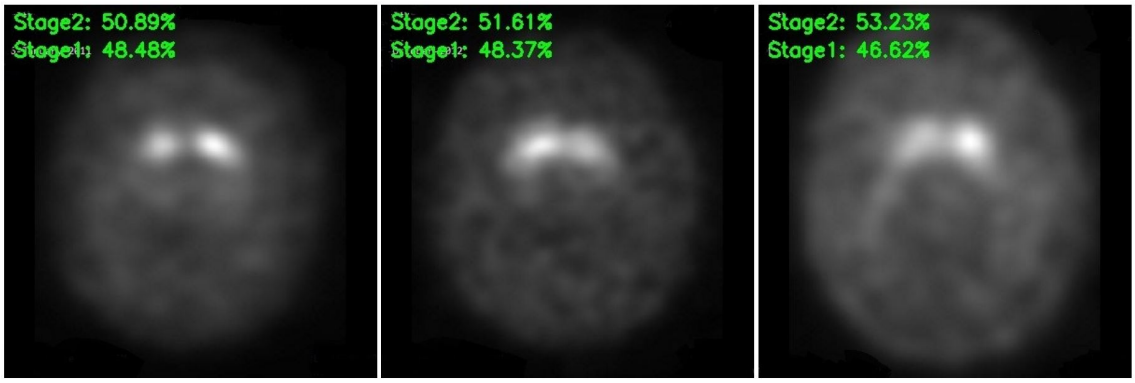


Figure 4.10 : Examples of low confidence classification

ogy or in the type of data. Table 4.4 provides a detailed comparison analysis between our study and similar published works. It can be concluded that our approach has outperformed all other approaches.

In conclusion, the precise diagnosis of PD patients in their early or at any stage is a challenging and essentail medical issue, due to many unclassical and overlapping common clinical manifestations shared among patients. Therefore, misdiagnosis of early PD is common. In this study, to address this issue, we have developed PDStageNet for the multiple stage classification of PD using SPECT images. From a total of 1,319 images, 929 images have been used for training the network model, and 390 images have been used for testing. Our network model has performed well, achieving an overall accuracy of 96.67%, accurately classifying 377 images. Our network has also achieved high sensitivity and specificity results, with 96.31% for sensitivity and 99.25% for specificity. This area of research is novel because there has not been any research studies reported in this area. The potential for our CNN model, to be used clinically, is significantly high because it could assist clinicians in their diagnosis. The implementation of such a network model would save time during PD diagnosis, allowing more time for treatment and management of PD.

Table 4.4 : A comparative analysis of the present study and the published studies

Research works	**Dataset used**	**Methodology**	**Accuracy**
Shahbaba & Neal [89]	Oxford Parkinson's Disease detection dataset	Dirichlet process mixtures	87.70%
Das [90]	Oxford Parkinson's Disease detection dataset	ANN	92.90%
Guo et al. [91]	Oxford Parkinson's Disease detection dataset	GP-EM	93.1%
Li et al. [92]	Multiple datasets	Fuzzy-based non-linear transformation + SVM	93.47%
Åström & Koker [93]	Oxford Parkinson's Disease detection dataset	Parallel ANN	91.20%
Spadoto et al. [94]	Oxford Parkinson's Disease detection dataset	PSO + OPF Harmony search + OPF Gravitational search + OPF	84.01%
Sakar & Kursun [95]	Oxford Parkinson's Disease detection dataset	Mutual information based feature selection + SVM	92.75%
Ozcift et al. [96]	Oxford Parkinson's Disease detection dataset	CFS-RF	87.13%
Chen et al. [46]	Oxford Parkinson's Disease detection dataset	PCA-FKNN	96.07%
Che, C. et al. [85]	PPMI	RNN-SVM	75%
Caesarendra, W. et al. [84]	Oxford Parkinson's Disease database	PCA-SVM	79.17%
This study	**PPMI**	**CNN**	**96.67%**

Chapter 5

Classification of Parkinson's Disease Patients for Surgical Treatment

Approximately ten million people worldwide are currently living with PD [71]. Many researchers believe that the disease results from an interaction between genetic and environmental factors that leads to progressive degeneration of neurons in susceptible regions of the brain. Currently, there are no laboratory tests that have diagnostic values for PD. For advanced PD, if medications do not help manage the symptoms, surgical treatment is the only alternative. In this chapter, we develop a model to help streamline the patient selection process for surgical treatment using only clinical data.

5.1 Introduction

Parkinsons disease (PD) is a disorder of the nervous system. It results from damage to the nerve cells in a region of the brain that produces dopamine, a chemical that is vital for the smooth control of muscles and movement.

In the early stages, many diagnostic methods, such as blood tests, brain imaging techniques such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET scan), and Single Photon Emission Computed Tomography (SPECT), are used to exclude other medical conditions such as stroke or brain tumors that imitate symptoms of PD [97]. Amongst others, one of the methods for the diagnosis of PD is detecting and analysing voice disorders by using acoustic tools that record the changes in pressure at lips or inside the vocal tract. It has been found [18] that

some features in the voices of the patients with PD can be used as discriminatory measures to differentiate PD by adopting data mining techniques.

Selection of appropriate patients for surgical treatment, in clinical practice, is based on many factors such as the age of the patient, disease stage, disease duration, comorbidities, and responsiveness to levodopa medication. It is usually required to form an interdisciplinary team consisting of neurologist, neurosurgeon, psychiatrist, neuropsychologist, rehabilitation specialist, and sometimes a social worker, to discuss these factors and then decide if the patient is suitable for surgical treatment [98, 99]. Therefore, there is an urgent need to streamline the process of selecting appropriate PD patients for surgical treatments. In this study, we propose an effective model to classify PD and select suitable patients for surgery using data mining algorithms and feature selection based on information gain with PD patients.

Data Mining is defined as the nontrivial extraction of implicit, previously unknown, and potentially useful information from generic data. The use of classifier systems in disease diagnosis is increasing. Technological advances in the field of Artificial Intelligence (AI) have led to the emergence of expert systems and Decision Support Systems (DSS) for medical applications. Moreover, in the last few decades, computational tools have been designed to improve the experiences and abilities of doctors and medical specialists in taking decisions regarding diagnosis and treatment about their patients. However, expert systems and different AI techniques for classification have the potential of being valuable supportive tools. Classification systems can help in increasing the accuracy and reliability of diagnoses and minimising possible errors, as well as making the diagnoses more time-efficient [25].

This area of research is important because advanced PD patients who do not respond to drug treatments require surgical therapy to control PD symptoms. However, the selection of appropriate patients for surgery is complicated, costly and

involves several discussions by an interdisciplinary team. This is time-consuming, resulting in significant delays in the treatment. Using data mining algorithms and feature selection, this research introduces a streamlined model to classify PD and identify appropriate patients for surgery. This will provide the team insight as to which PD patient to select for surgery based on their values after the feature selection process.

5.2 Proposed Network Architecture

This study utilises real patients' clinical, medical and surgical data to develop a practical model for PD classification and for selection of suitable PD patients for surgical treatment. Our proposed approach is outlined in Figure 5.1.

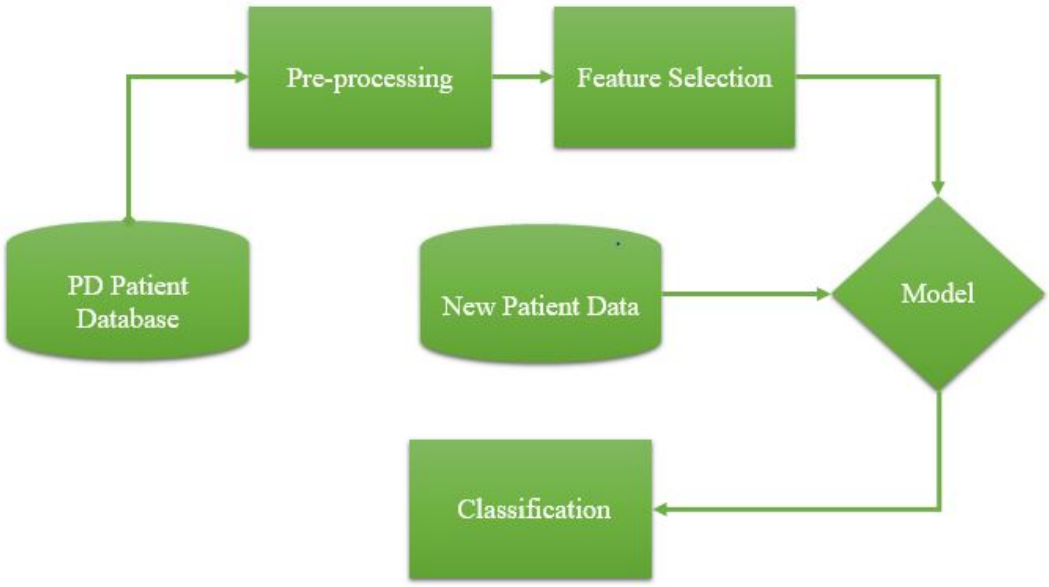


Figure 5.1 : Proposed approach architecture.

For this study, a new dataset was developed by gathering different types of data from PPMI such as medication, health, surgical, PD symptoms, MRI scan results and post medication data. This provided us with 1,080 patients, of which 40 patients

opted for surgery, with 40 attributes for each patient. The dataset was further divided into a training dataset (70%) and a testing dataset (30%). The training dataset consisted of records of 758 patients, and the testing dataset consisted of records of 322 patients each with 40 attributes, where 23 out of 322 patients were of surgical patients.

5.2.1 Pre-processing

During the pre-processing stage, the dataset was analysed to see if it had any missing values or redundant values. To address the issue of missing values, WEKA filter "Remove Missing Values" was used [100]. For redundant data, the latest record of the data registered in the database was used erasing old data. We then used two datasets for our experiments. Experiment 1 had all 40 attributes. For Experiment 2, the dataset, after feature selection, was split-up into 10 datasets each consisting of 10% of the attributes.

5.2.2 Feature Selection

Feature selection is the process of selecting a subset of relevant features (e.g., variables, predictors) for use in model construction. Feature selection techniques are used to avoid overfitting and improve model performance. They can be used to provide faster and more cost-effective models. They are also used to gain more in-depth insight into the underlying processes that generated the data [101]. Thus, with this model, the selection of appropriate patients for surgical treatment should be highly accurate and effective.

The objective is to reduce the attributes in order to identify the most important feature that contributes to the classification. In our work, we select the filter methods because they are moderately robust against the overfitting problem. We use the Information Gain (IG) technique as this technique computes the information gain

of a feature with respect to class. In order to define Information Gain, Entropy must be defined first. Entropy is the measure of disorder of uncertainity in a dataset. It can be deifined by:

$$H(S) = -\sum_{i=1}^{c} p_i log_2 p_i, \tag{5.1}$$

where p_i is the proportion of samples that belongs to class c.

Information Gain can be defined as the change in entropy by splitting a dataset according to a given value of a random feature [102]. IG can be calculated by:

$$IG(S, a) = H(S) - H(S|a), \tag{5.2}$$

where $IG(S, a)$ is the information gain for the dataset S with respect to feature a, $H(S)$ is the entropy for the dataset before any change and $H(S|a)$ is the entropy for the dataset with respect to the feature a.

The Ranker filter is also used along with IG. Ranker helps in ranking the features based on their information gain with respect to class. The features are ranked in decreasing order where the feature with highest IG is ranked first and the feature with the lowest IG is ranked last.

5.2.3 Model

The developed model aims to provide an accurate classification of PD patients and effectively assess if the patients require surgical treatment. When a new patient arrives, all of his/her information is stored in the database. By this stage, several features would have been extracted through the feature selection technique. Therefore, the data collected from new patients consist of those features only. Their features are applied to the model to know whether they are applicable for surgery. Figure 5.2 illustrates how the model works. Figure 5.3 provides the list of features of the dataset.

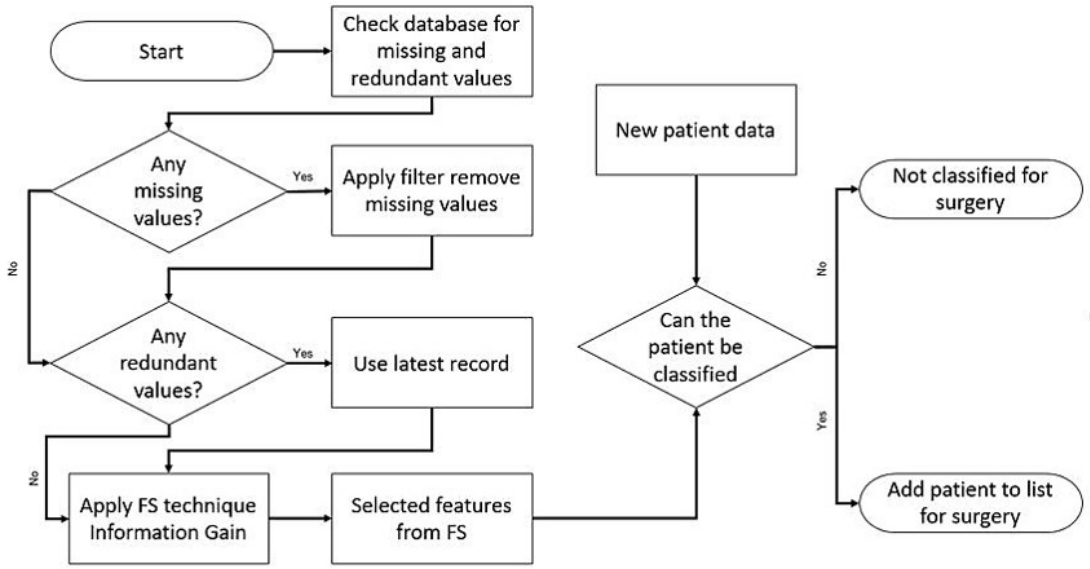


Figure 5.2 : Flowchart of the model

5.2.4 Classification

In this step, the new patient will be classified into whether the patient is suitable to undergo a surgery or not. This would help the specialists to undertake proactive

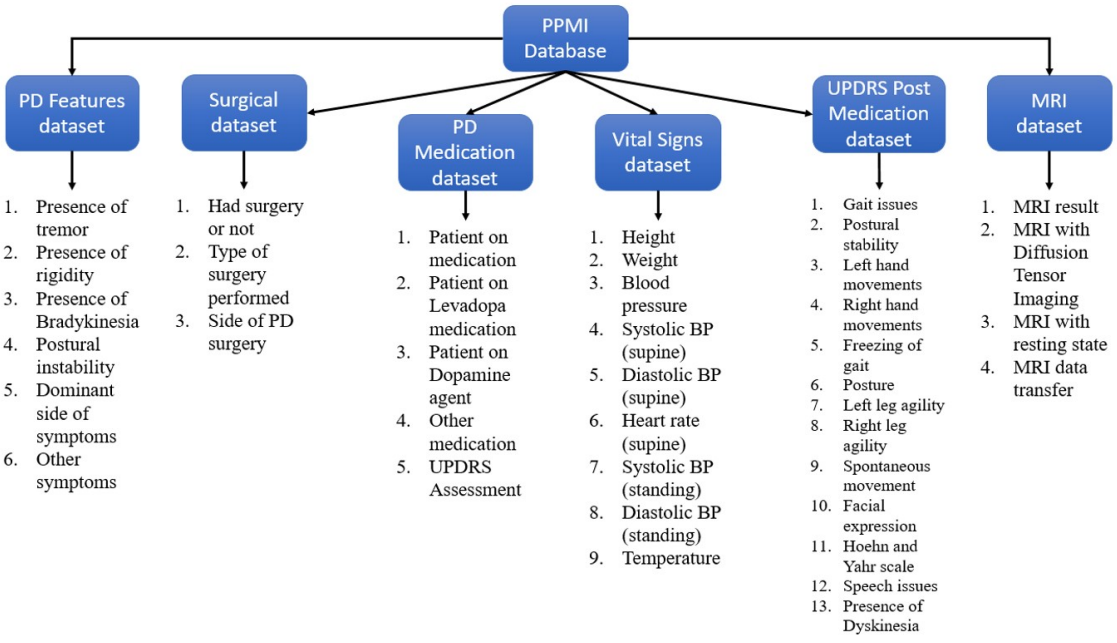


Figure 5.3 : List of features

steps so that the patient can get the right treatment. In this chapter, we utilised Naïve Bayes, Decision tree, Support Vector Machines and Multilayer Perceptron for our experiments. The details of our experiments are provided in Section 5.3.

5.3 Experiments and Results

All of our experiments were performed using the WEKA platform. The data was gathered from PPMI. For our study, we gathered data from various datasets in PPMI presenting us with 1,080 subjects with 40 attributes.

5.3.1 Experiment 1: General Classification of PD Patients for Surgery Using Different Classifiers

This experiment aims to develop a model to provide an accurate classification of a PD patient so that it can suggest if the patient is indicated for a surgical option. Since we are trying to classify which patients require surgery, there are two classes for this research, i.e., class 0 for those that do not need surgery and class 1 that requires surgery.

In this experiment, Naïve Bayes, Support Vector Machine (SVM), J48 (Decision Tree), and Multilayer Perceptron (MLP) that is a form of a neural network algorithm from WEKA, are used to construct the model [32]. In this experiment, all 40 features are used for the classification of PD patients. Table 5.1 presents classification results.

For performance measurement, a confusion matrix is obtained to estimate four measures: Accuracy, Precision, F-measure and Sensitivity. They are calculated as follows.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{5.3}$$

$$Precision = \frac{TP}{TP + FP}, \tag{5.4}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{5.5}$$

Table 5.1 : General Classification of PD Patients Using Different Classifiers

Learning Machine	**Parameters Used**	**Model Accuracy**	**Precision**	**Sensitivity (weighted average)**	**F-Measure**
Naïve Bayes	Simple Estimator function	93%	0.862	0.929	0.894
Decision Tree (J48)	Binary splits, Confidence factor = 0.25, Size of tree = 10	94.7%	0.95	0.95	0.932
Multilayer Perceptron	Number of layers = 3, Number of epochs = 500 and Sigmoid function	98.13%	0.982	0.981	0.980
Support Vector Machine	RBF Kernel	93%	0.862	0.929	0.894

$$F - measure = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}, \tag{5.6}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

For this experiment, the weighted average value of sensitivity is used because it provides an overall performance of the model by taking into account the results of both classes. As a result, MLP has the highest accuracy of 98.13% followed by J48 at 94.7%, SVM and Naïve Bayes at 93% each.

5.3.2 Experiment 2: Classification of PD Patients for Surgery after Using Feature Selection

The aim of this experiment is to study the effect of feature selection in the accuracy of the classification. Feature selection is one of the dimensionality reduction techniques for reducing the attribute space of a feature set. More precisely, it determines how many features should be enough to give moderate accuracy.

For feature selection, we utilised IG filter from the WEKA platform. This filter acted as an attribute evaluator as it evaluated attributes according to their information gain. This process also used the "Ranker" filter to rank the features based on their information gain with respect to class. For this experiment, we used the same dataset as in Experiment 1. Ten datasets were then built depending on the number of selected features. The first dataset contained only 10% of the total attributes. Then, each time, the feature selections were increased by 10%. Therefore, dataset 1 contained 10% of all attributes, dataset 2 contained 20%, dataset 3 contained 30%, ..., and dataset 10 contained 100% of all attributes.

We chose MLP as the classifier because it had the highest accuracy result from Experiment 1. The MLP model was made up of three hidden layers with 500 nodes in each layer. The default activation function, i.e., Approximate Sigmoid function, was used in our model. Each feature-reduced dataset was used for a 10-fold cross-validation for evaluation. It was observed that, with only 60% of the attributes, our model achieved the highest accuracy result. This proved that not all features might be necessary to attain a highly accurate classification.

The sensitivity is also measured for the test set. For this experiment, the results of sensitivity are provided with respect to surgery class only. These results are listed so that we can measure how the proposed algorithm can correctly classify PD patients for surgery.

Table 5.2 : Classification of PD Patients for Surgery Using Feature Selection

% of Features selected	# of Features selected	Correctly classified for surgery	Incorrectly classified for surgery	Correctly classified for non-surgery	Incorrectly classified for non-surgery	Sensitivity (with respect to surgery)	Precision (95% confidence interval)	Surgery vs Non-surgery accuracy
10%	4	0	23	299	0	0	0.92±0.034	92.85%
20%	8	5	18	299	0	0.217	0.94±0.031	94.4%
30%	12	11	12	299	0	0.478	0.96±0.024	96.6%
40%	16	14	9	299	0	0.609	0.97±0.021	97.2%
50%	20	16	7	298	1	0.696	0.97±0.018	97.5%
60%	24	17	6	299	0	0.739	0.98±0.018	98.13%
70%	28	15	8	299	0	0.652	0.97±0.015	97.5%
80%	32	17	6	299	0	0.739	0.98±0.015	98.13%
90%	36	17	6	299	0	0.739	0.98±0.015	98.13%
100%	40	17	6	299	0	0.739	0.98±0.015	98.13%

A 95% confidence interval is also measured. The primary purpose of the confidence interval is to provide a range of values for an estimated parameter rather than a single point value. All of these results are presented in Table 5.2. The confidence interval is calculated as follows.

$$CI = const * \sqrt{error * \frac{1 - error}{n}}, \tag{5.7}$$

where *const* is a constant value (which is 1.96 in our case) corresponding to the probability (i.e., 95%), *error* is the classification error and n is the sample size.

5.4 Discussion and Conclusion

In the medical domain, an unmet medical need in the management of PD is how to simplify the complicated process of selecting suitable PD patients for surgical treatment, a breakthrough in recent years. Therefore, our objective is to develop a practical classification model that can accurately identify PD patients for surgery

based on data mining algorithms combined with feature selection using clinical data. The model will enable physicians to, reliably and accurately, classify PD patients using available clinical parameters without the time consuming and costly medical team meetings and discussions.

In the current study, we have developed a novel model to streamline the way to classify PD and to identify appropriate PD patients for surgery. The model is remarkably reliable and accurate, with an accuracy of 98.13%.

From Experiment 2, it can be noted that not all features/attributes contribute to the high accuracy of classification. Using only 24 (60%) attributes, the classification accuracy of PD patients for surgery is similar to the result of classification using all the features. Figure 5.4 indicates of how much contribution each feature has made to the classification accuracy. The values of sensitivity results differ in the two experiments. The weighted average of sensitivity, in Experiment 1, takes into account the sensitivity values of both classes (0 and 1), multiplies them with the instances classified into their respective classes and divides the total with the total number of instances. However, for Experiment 2, the values of sensitivity concerning only the surgery class are provided so that we can measure how the proposed algorithm can correctly classify PD patients for surgery.

Furthermore, each of the 24 features does not make an equal contribution to the classification of PD patients. The top features that contribute more are presented in Table 5.3. This has been generated by using the correlation filter from WEKA. The correlation filter evaluates the worth of a feature by measuring the Pearson's correlation between the attribute and the class [103].

Understanding the importance of the top features in their contribution to classification is important because it provides us with an idea of how the features interact with the class for classification. For example, the Hoehn and Yahr scale (HY) is a

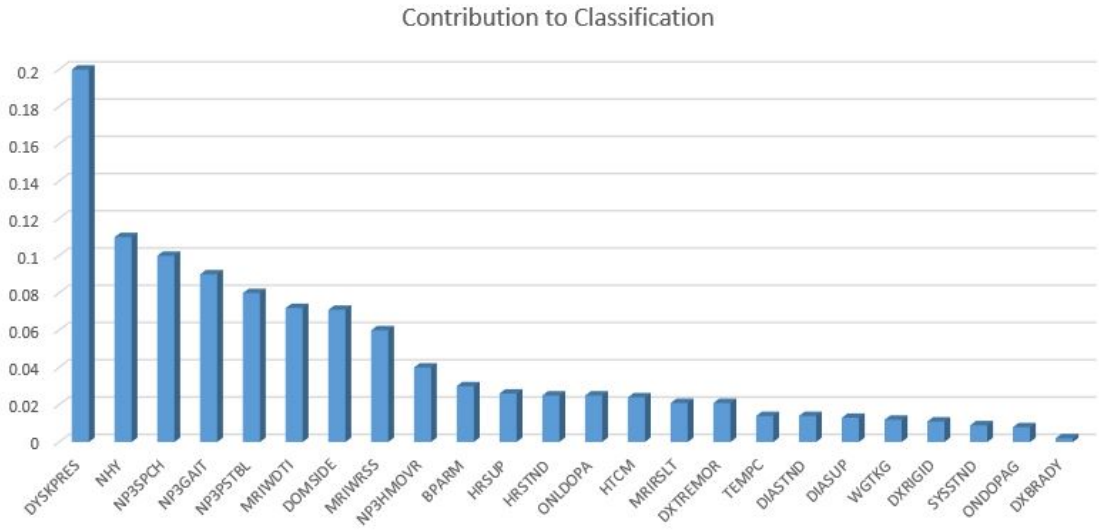


Figure 5.4 : Contribution of features to the classification of PD patients for surgery

widely used clinical rating scale, which defines broad categories of motor function in PD. Among its advantages, it is simple and easily applied. It captures typical patterns of progressive motor impairment, and it can be applied to verify whether or not patients are receiving therapy. Progression in HY stages has been found to correlate with motor decline, deterioration in the quality of life, and neuroimaging

Table 5.3 : Top seven features' details

Attribute code	**Category**	**Description**
DYSKPRES	Motor Assessments	Presence of dyskinesia
NHY	Motor Assessments	Hoehn and Yahr scale
NP3SPCH	Motor Assessments	Speech issues after medication
NP3GAIT	Motor Assessments	Gait disorder after medication
NP3PSTBL	Motor Assessments	Postural stability after medication
MRIWDTI	MRI Result	MRI with Diffusion Tensor Imaging (DTI)
DOMSIDE	PD Features	Dominant side of PD

studies of dopaminergic loss [86].

As shown in Table 5.2, among the 23 patients that have undergone surgery, the proposed model with 60% of the attributes, has correctly classified 17 of them. Out of those 17 patients, nine patients can be correctly classified by the seven attributes listed in Table 5.3. This was achieved by splitting the training dataset into two parts, i.e., one part included the top 7 features and the remaining part had the rest of the features. We then utilised MLP as the classifer and evaluated on the testing dataset. Figure 5.5 shows a comparison of classification accuracies using the top seven features and the rest of the features.

5.4.1 Comparison of Current Study with Other Research Works

In this section, we compare the performance among the approaches proposed in the chapter and the related works, and demonstrate the comparison results in Table 5.4.

In this study, data on subjects that have been diagnosed into various stages of PD are gathered. Based on the data, we have classified whether the subject needs to undergo surgery. To the best of our knowledge, we are the first to use data mining techniques to streamline the process of selecting suitable PD patients for surgery with potentially critical clinical applications.

In this chapter, we propose a streamlined model for the classification of PD patients for surgical treatment. Majority of the other related existing approaches focus on only pure classification and diagnosis of PD patients during the early stages. We can compare our experimental results with other published approaches because our experiments does also provide the classification accuracy results of diagnosing PD patients from healthy controls.

The significant contribution from the present study is that our experiments have

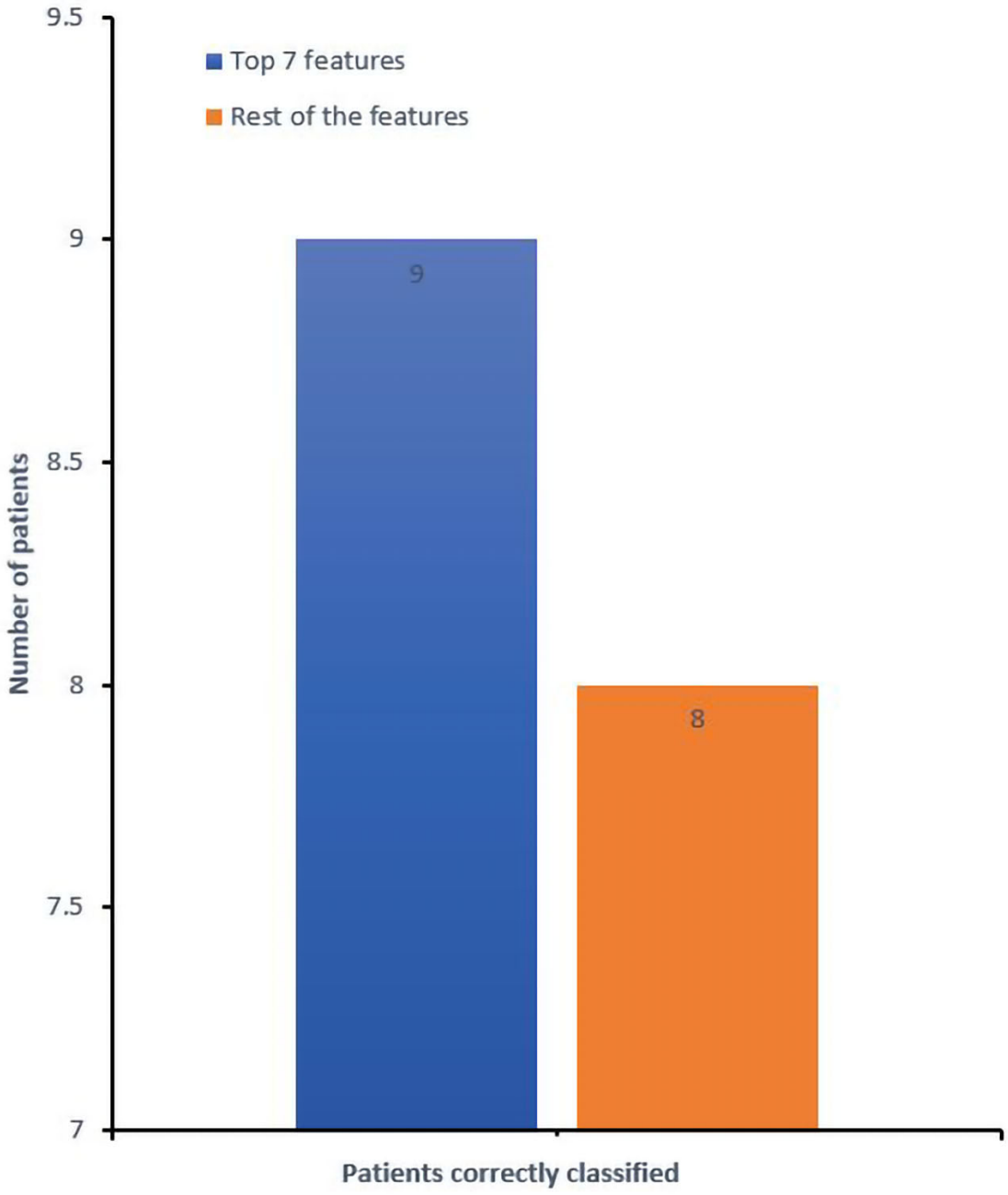


Figure 5.5 : Classification results using top 7 features and the rest of the features, respectively

given high performance in this area of research. Furthermore, most of the related works have a limitation of either having a small dataset or small sample size as detailed in Table 5.4.

In conclusion, we have proposed and developed a data mining and feature selection-based model for accurate PD classification and selection of suitable PD patients for surgery. Potentially, this model could fill the unmet medical need of streamlining the complicated process of selecting suitable PD patients for surgery.

In Experiment 1, we have developed a novel PD classification model with several classifiers. After comparing several alternative classifiers, we have found that MLP consistently outperforms the others in most experiments with the highest PD classification accuracy of 98.13%. Experiment 2 has identified the most important attributes required for such classification by using feature selection. One of the key findings is that, using only 60% of the attributes, MLP classification with IG has produced a remarkably high accuracy (98.13%) indicating that a smaller number of clinical parameters is sufficient for reliable and accurate diagnosis of PD.

The results from our experiments have demonstrated that the developed model can be a useful tool in clinical practice for accurate classification of PD and selection of appropriate PD patients for surgery. Our model has also provided a better understanding of features that contribute to reliable and accurate PD classification, indicating that not all features are required for the accurate and efficient PD classification.

Table 5.4 : Summary and comparisons of previous related works on PD classification

Research works	**Dataset used**	**Attributes in dataset**	**Sample size**	**Classifier used**	**Accuracy**	**Feature selection**	**Accuracy after feature selection**
Little et al. [44]	Developed own dataset	Vocal attributes	31	SVM	91.4%	No	N/A
Rustempasic and Can [18]	Max Little dataset	Vocal attributes	31	Fuzzy C-Means clustering	80.88%	No	N/A
Khemphila and Boonjing [25]	Max Little dataset	Vocal attributes	31	Artificial Neural Network (ANN)	80.76%	Yes	83.33%
Shaikh and Chabra [45]	Max Little dataset	Vocal attributes	31	Naïve Bayes	69.23%	Yes	78.46%
Gok [47]	Max Little dataset	Vocal attributes	31	k-Nearest Neighbor (k-NN)	N/A	Yes	98.46%
Prashanth et al. [59]	PPMI	Striatal binding ratio (SBR) values	493	SVM	96.1%	No	N/A
Hirschauer et al. [60]	PPMI	SBR and clinical data	666	EPNN	92.5%	No	N/A
Prashanth et al. [61]	PPMI	Clinical data	584	SVM	96.4%	No	N/A
Suganya and Sumanthi [48]	Private dataset	Vocal attributes	195	ABO	97.5%	No	N/A
Luukka[104]	Max Little dataset	Vocal attributes	195	Fuzzy entropy measures + Similarity	79.22%	Yes	85.03%
Current research	PPMI	Clinical data	1080	MLP	98.13%	Yes	98.13%

Chapter 6

Conclusion and Future Work

Precise diagnosis of PD in their early stages and differential diagnosis of PD at any stage are challenging and important medical problems, because there have been patients who exhibit many non-classical, and overlapping common, clinical indications. Therefore, misdiagnosis of early PD is common. In this thesis, a deep-learning approach has been proposed to discriminate PD patients from non-PD patients utilizing SPECT images. Furthermore, we have proposed another approach to classify PD patients based on their stages of PD progression. Finally, a classification model has been proposed to streamline the process of identifying PD patients for surgical treatment using clinical data.

6.1 Classification of PD Using Deep-Learning

In this thesis, a deep-learning model has been developed to classify PD patients from non-PD patients utilizing SPECT images. The network has been trained on SPECT images of PD and non-PD patients, and has classified them accordingly. Experimental results show a remarkable high accuracy of 99.45%, sensitivity of 98.93% and specificity of 100%. By accurate identification of features of degenerative PD at its early stage, our approach has addressed the challenging issue of early classification of PD. The evaluation results of our network's performance have shown its importance in early PD diagnosis, which is crucial for effective PD patient management. Based on our experimental results, we can confidently claim that our model outperforms the benchmark studies by a large margin. Since the network's architectural complexity is low, the potential for this CNN model to be used clinically,

in day-to-day PD diagnosis, is high. It could be used by clinicians, quantitatively, to observe the deterioration and progress of PD conditions.

6.2 Multiple Stages Classification of PD

This thesis has proposed an effective CNN-based model to classify PD patients into multiple stages of PD progression using SPECT images. A total number of 1,319 SPECT images has been gathered, where 929 images have been used for training the network model, and the remaining 390 images have been used for testing. Our network model has performed well, achieving an overall high accuracy of 96.67%, accurately classifying 377 images. Sensitivity and specificity have also been calculated with a high sensitivity score of 96.31% and specificity score of 99.25%. To the best of our knowledge, this is the first study reported that classifies PD patients into multiple stages of PD progression. The significance of this study lies in the fact that, as diagnosing PD patients in the early stages is challenging for clinicians, with the help of our network model's high-performance results, diagnosing PD patients into their respective stages becomes easier for clinicians. The potential of our CNN model, to be used clinically in the day-to-day diagnosis of PD, is significantly high because it can benefit clinicians in their diagnosis. The implementation of such a network model would save time during PD diagnosis, allowing more time for treatment and management of PD symptoms.

6.3 Classification of PD Patients for Surgical Treatment

In this thesis, we have proposed a feature selection based model for accurate PD classification and selection of suitable PD patients for surgery using clinical data. Potentially, this model could fill the unmet medical need of streamlining the complicated process of selecting suitable PD patients for surgery.

In Experiment 1, we have developed a novel PD classification model with several

classifiers. After comparing several alternative classifiers, it has been noted that Multilayer Perceptron (MLP) has consistently outperformed the others, achieving the highest PD classification accuracy of 98.13%. Experiment 2 identifies the essential features required for such classification by using feature selection. One of the key findings reported is that using only 60% of the attributes, MLP classification with IG has produced a remarkably high accuracy of 98.13%. This infers that, with a smaller number of clinical parameters, our proposed model can be reliable and accurate in the diagnosis of PD.

The results from our experiments have also demonstrated that the proposed model can be a useful tool in clinical practice for accurate selection of appropriate PD patients for surgery. With a feature selection process, our proposed approach provides a better understanding of features that contribute to reliable and accurate PD classification, indicating that not all features are necessary for the accurate and efficient classification of PD.

6.4 Future Work

Adopting deep-learning algorithms for diagnosis and classification of PD shows a very high potential for practical use. Diagnosing PD, in its early stages, is challenging due to many common clinical manifestations in patients with other neurodegenerative diseases. However, with its challenges, the road for such technology to fulfil its full potential in PD diagnosis requires several plans. According to [105], several key steps can be undertaken to fulfil such plans. These are clinical validation, open platform standardisation and data sharing.

Clinical validation is required for regulatory approval, much like the measurement of, for example, blood pressure. Such approval requires that the information provided is an accurate parameter of a clinically relevant feature of the disease and that there is confirmed evidence that this parameter has a relevant response within

some clinical application and numerical target ranges exist in which the parameter measures adequate treatment response. Fundamentally, for machine learning-based models to be reliable and accurate, the information must be of critical value in relation to its corresponding disease. Such crucial data can only be validated by clinicians/medical specialists.

Platform standardisation requires that the software and algorithms are made publicly available for general and widespread use because any technology lives or dies by the scale of its adoption and dissemination on the broader community.

Finally, data sharing is a crucial aspect for such research area. By sharing data, we can further refine algorithms and provide increasingly extensive and rigorous clinical validation evidence. Incentives encouraging data sharing should be established at institutional and disciplinary levels.

Bibliography

[1] R. G. Ramani and G. Sivagami, "Parkinson disease classification using data mining algorithms," *International Journal of Computer Applications*, vol. 32, 2011.

[2] D. P, V. S. Jebakumari, and D. Shanthi, "Diagnosis and classification of parkinsons disease using data mining techniques," *International Journal of Advanced Research Trends in Engineering and Technology*, vol. 3, pp. 86–90, 2016.

[3] E. R. Dorsey, A. Elbaz, E. Nichols, F. Abd-Allah, A. Abdelalim, J. C. Adsuar, M. G. Ansha, C. Brayne, J.-Y. J. Choi, D. Collado-Mateo *et al.*, "Global, regional, and national burden of parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016," *The Lancet Neurology*, vol. 17, no. 11, pp. 939–953, 2018.

[4] M. Delenclos, D. R. Jones, P. J. McLean, and R. J. Uitti, "Biomarkers in parkinson's disease: Advances and strategies," *Parkinsonism & related disorders*, vol. 22, pp. S106–S110, 2016.

[5] E. Garnett, G. Firnau, and C. Nahmias, "Dopamine visualized in the basal ganglia of living man," *Nature*, vol. 305, no. 5930, pp. 137–138, 1983.

[6] A. J. Stoessl, S. Lehericy, and A. P. Strafella, "Imaging insights into basal ganglia function, parkinson's disease, and dystonia," *The Lancet*, vol. 384, no. 9942, pp. 532–544, 2014.

[7] R. de la Fuente-Fernández, "Imaging of dopamine in pd and implications for motor and neuropsychiatric manifestations of pd," *Frontiers in neurology*, vol. 4, p. 90, 2013.

[8] A. Berardelli, G. Wenning, A. Antonini, D. Berg, B. Bloem, V. Bonifati, D. Brooks, D. Burn, C. Colosimo, A. Fanciulli *et al.*, "Efns/mds-es recommendations for the diagnosis of parkinson's disease," *European Journal of Neurology*, vol. 20, no. 1, pp. 16–34, 2013.

[9] G. Treglia, E. Cason, A. Stefanelli, F. Cocciolillo, D. Di Giuda, G. Fagioli, and A. Giordano, "Mibg scintigraphy in differential diagnosis of parkinsonism: a meta-analysis," *Clinical autonomic research*, vol. 22, no. 1, pp. 43–55, 2012.

[10] C. Scherfler, G. Göbel, C. Müller, M. Nocker, G. K. Wenning, M. Schocke, W. Poewe, and K. Seppi, "Diagnostic potential of automated subcortical volume segmentation in atypical parkinsonism," *Neurology*, vol. 86, no. 13, pp. 1242–1249, 2016.

[11] F. P. Oliveira, D. B. Faria, D. C. Costa, M. Castelo-Branco, and J. M. R. Tavares, "Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of parkinson's disease based on [123 i] fp-cit spect images," *European journal of nuclear medicine and molecular imaging*, vol. 45, no. 6, pp. 1052–1062, 2018.

[12] T. Booth, M. Nathan, A. Waldman, A.-M. Quigley, A. Schapira, and J. Buscombe, "The role of functional dopamine-transporter spect imaging in parkinsonian syndromes, part 1," *American Journal of Neuroradiology*, vol. 36, no. 2, pp. 229–235, 2015.

[13] J. L. Cummings, C. Henchcliffe, S. Schaier, T. Simuni, A. Waxman, and P. Kemp, "The role of dopaminergic imaging in patients with symptoms of

dopaminergic system neurodegeneration," *Brain*, vol. 134, no. 11, pp. 3146–3166, 2011.

[14] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, "High-accuracy classification of parkinson's disease through shape analysis and surface fitting in 123i-ioflupane spect imaging," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 794–802, 2017.

[15] D. A. Economics, "Living with parkinsons disease–an updated economic analysis 2014," 2015.

[16] A. I. of Heaalth and W. (AIHW), "Health system expenditure on disease and injury in australia 2000-01, second edition," 2010.

[17] J. A. Temlett and P. D. Thompson, "Reasons for admission to hospital for parkinsons disease," *Internal medicine journal*, vol. 36, no. 8, pp. 524–526, 2006.

[18] I. Rustempasic and M. Can, "Diagnosis of parkinson's disease using fuzzy c-means clustering and pattern recognition," *Southeast Europe Journal of Soft Computing*, vol. 2, no. 1, 2013.

[19] M. A. Hely, J. G. Morris, W. G. Reid, and R. Trafficante, "Sydney multicenter study of parkinson's disease: Non-l-dopa–responsive problems dominate at 15 years," *Movement disorders: official journal of the Movement Disorder Society*, vol. 20, no. 2, pp. 190–199, 2005.

[20] J. Tiller, "Therapeutic guidelines: Neurology (version 3)," *Melbourne: Therapeutic Guidelines Ltd*, vol. 163, 2007.

[21] L. M. Rubenstein, E. A. Chrischilles, and M. D. Voelker, "The impact of parkinsons disease on health status, health expenditures, and productivity," *Pharmacoeconomics*, vol. 12, no. 4, pp. 486–498, 1997.

[22] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks*, vol. 10, no. 5, pp. 988–999, 1999.

[23] J. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez, "Gmm based spect image classification for the diagnosis of alzheimers disease," *Applied Soft Computing*, vol. 11, no. 2, pp. 2313–2325, 2011.

[24] F. Segovia, J. Górriz, J. Ramírez, I. Alvarez, J. Jiménez-Hoyuela, and S. Ortega, "Improved parkinsonism diagnosis using a partial least squares based approach," *Medical physics*, vol. 39, pp. 4395–4403, 2012.

[25] A. Khemphila and V. Boonjing, "Parkinsons disease classification using neural network and feature selection," *World Academy of Science, Engineering and Technology*, vol. 64, 2012.

[26] M. J. Berry and G. Linoff, *Mastering data mining*. Wiley New York, 2000.

[27] G. Shmueli, N. R. Patel, and P. C. Bruce, *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. John Wiley and Sons, 2011.

[28] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[29] M. Verma, M. Srivastava, N. Chack, A. K. Diswar, and N. Gupta, "A comparative study of various clustering algorithms in data mining," *International Journal of Engineering Research and Applications (IJERA)*, vol. 2, 2012.

[30] D. T. Larose and C. D. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.

[31] S. Bouktif, E. M. Hanna, N. Zaki, and E. A. Khousa, "Ant colony optimization algorithm for interpretable bayesian classifiers combination: application to

medical predictions," *PloS one*, vol. 9, 2014.

[32] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.

[33] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[34] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[35] C. Nasa and S. Suman, "Evaluation of different classification techniques for web data," *International journal of computer applications*, vol. 52, no. 9, pp. 34–40, 2012.

[36] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine learning*, vol. 6, no. 1, pp. 37–66, 1991.

[37] T. M. Mitchell *et al.*, "Machine learning. 1997," *Burr Ridge, IL: McGraw Hill*, vol. 45, no. 37, pp. 870–877, 1997.

[38] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[39] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[40] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[41] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[42] P. P. M. Initiative, "Parkinson's progression markers initiative," http://www.ppmi-info.org/, 2017.

[43] "The michael j fox foundation for parkinson's research," https://www.michaeljfox.org/, 2017.

[44] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, 2009.

[45] T. A. Shaikh and A. Chhabra, "Effect of weka filters on the performance of the navie bayes data mining algorithm on arrhythmia and parkinson's datasets," 2014.

[46] H.-L. Chen, C.-C. Huang, X.-G. Yu, X. Xu, X. Sun, G. Wang, and S.-J. Wang, "An efficient diagnosis system for detection of parkinsons disease using fuzzy k-nearest neighbor approach," *Expert systems with applications*, vol. 40, no. 1, pp. 263–271, 2013.

[47] M. Gök, "An ensemble of k-nearest neighbours algorithm for detection of parkinson's disease," *International Journal of Systems Science*, vol. 46, no. 6, pp. 1108–1112, 2015.

[48] P. Suganya and C. Sumathi, "A novel metaheuristic data mining algorithm for the detection and classification of parkinson disease," *Indian Journal of Science and Technology*, vol. 8, no. 14, 2015.

[49] S. B. Appakaya and R. Sankar, "Classification of parkinsons disease using pitch synchronous speech analysis," in *2018 40th Annual International Conference*

of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018, pp. 1420–1423.

[50] W. Zeng, F. Liu, Q. Wang, Y. Wang, L. Ma, and Y. Zhang, "Parkinson's disease classification using gait analysis via deterministic learning," *Neuroscience letters*, vol. 633, pp. 268–278, 2016.

[51] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[52] S. Mazilu, A. Calatroni, E. Gazit, D. Roggen, J. M. Hausdorff, and G. Tröster, "Feature learning for detection and prediction of freezing of gait in parkinsons disease," in *International workshop on machine learning and data mining in pattern Recognition*. Springer, 2013, pp. 144–158.

[53] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster, "Wearable assistant for parkinsons disease patients with the freezing of gait symptom," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 436–446, 2009.

[54] Y. Mittra and V. Rustagi, "Classification of subjects with parkinson's disease using gait data analysis," in *2018 International Conference on Automation and Computational Engineering (ICACE)*. IEEE, 2018, pp. 84–89.

[55] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, "Classification of parkinson's disease gait using spatial-temporal gait features," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1794–1802, 2015.

[56] E. Rastegari, S. Azizian, and H. Ali, "Machine learning and similarity network approaches to support automatic classification of parkinsons diseases using accelerometer-based gait analysis," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[57] A. Khorasani and M. R. Daliri, "Hmm for classification of parkinsons disease based on the raw gait data," *Journal of medical systems*, vol. 38, no. 12, p. 147, 2014.

[58] K. Polat, "Freezing of gait (fog) detection using logistic regression in parkinson's disease from acceleration signals," in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*. IEEE, 2019, pp. 1–4.

[59] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, "Automatic classification and prediction models for early parkinson's disease diagnosis from spect imaging," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3333–3342, 2014.

[60] T. J. Hirschauer, H. Adeli, and J. A. Buford, "Computer-aided diagnosis of parkinsons disease using enhanced probabilistic neural network," *Journal of Medical Systems*, vol. 39, no. 11, p. 179, 2015.

[61] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, "High-accuracy detection of early parkinson's disease through multimodal features and machine learning," *International Journal of Medical Informatics*, vol. 90, pp. 13–21, 2016.

[62] F. P. Oliveira and M. Castelo-Branco, "Computer-aided diagnosis of parkinson's disease based on [123i] fp-cit spect binding potential images, using the

voxels-as-features approach and support vector machines," *Journal of Neural Engineering*, vol. 12, no. 2, p. 026008, 2015.

[63] H. D. Tagare, C. DeLorenzo, S. Chelikani, L. Saperstein, and R. K. Fulbright, "Voxel-based logistic analysis of ppmi control and parkinson's disease datscans," *NeuroImage*, vol. 152, pp. 299–311, 2017.

[64] D. J. Towey, P. G. Bain, and K. S. Nijran, "Automatic classification of 123i-fp-cit (datscan) spect images," *Nuclear medicine communications*, vol. 32, no. 8, pp. 699–707, 2011.

[65] A. Rojas, J. Górriz, J. Ramírez, I. Illán, F. J. Martínez-Murcia, A. Ortiz, M. G. Río, and M. Moreno-Caballero, "Application of empirical mode decomposition (emd) on datscan spect images to explore parkinson disease," *Expert Systems with Applications*, vol. 40, no. 7, pp. 2756–2766, 2013.

[66] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for parkinsons disease diagnosis from eeg signals," *Neural Computing and Applications*, pp. 1–7, 2018.

[67] F. J. Martinez-Murcia, A. Ortiz, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, and I. A. Illán, "A 3d convolutional neural network approach for the diagnosis of parkinson?s disease," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer, 2017, pp. 324–333.

[68] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, "Refining diagnosis of parkinson's disease with deep learning-based interpretation of dopamine transporter imaging," *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.

[69] R. E. Burke, "Evaluation of the braak staging scheme for parkinson's disease: Introduction to a panel presentation," *Movement Disorders*, vol. 25, no. S1, pp. 76–77, 2010.

[70] Y. Zhao, L. Tan, P. Lau, W. Au, S. Li, and N. Luo, "Factors affecting health-related quality of life amongst asian patients with parkinsons disease," *European Journal of Neurology*, vol. 15, no. 7, pp. 737–742, 2008.

[71] A. Elkouzi, "Understanding parkinson: Statistics," http://parkinson.org/Understanding-Parkinsons/Causes-and-Statistics/Statistics, 2018.

[72] G. Orru, W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1140–1152, 2012.

[73] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[74] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[75] D. Godoy, "Understanding binary cross-entropy/log loss: a visual explanation," *Towards datascience, recuperado de: https://towardsdatascience.com/understanding-binary-cross-entropy-logloss-a-visual-explanation-a3ac6025181a*, 2018.

[76] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2d and 3d deep learning techniques for cardiac mr image segmen-

tation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 111–119.

[77] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: a system for large-scale machine learning." in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, vol. 16, 2016, pp. 265–283.

[78] K. Documentation, "Keras: The python deep learning library," https://keras.io/, 2018.

[79] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[80] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential data augmentation techniques for medical imaging classification tasks," in *AMIA Annual Symposium Proceedings*, vol. 2017. American Medical Informatics Association, 2017, p. 979.

[81] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.

[82] L. Wang, Q. Zhang, H. Li, and H. Zhang, "Spect molecular imaging in parkinson's disease," *BioMed Research International*, vol. 2012, 2012.

[83] N. P. Foundation, "The stages of parkinson's disease," http://www.parkinson.org/understanding-parkinsons/what-is-parkinsons/The-Stages-of-Parkinsons-Disease, 2017.

[84] W. Caesarendra, F. T. Putri, M. Ariyanto, and J. D. Setiawan, "Pattern recognition methods for multi stage classification of parkinson's disease uti-

lizing voice features," in *2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2015, pp. 802–807.

[85] C. Che, C. Xiao, J. Liang, B. Jin, J. Zho, and F. Wang, "An rnn architecture with dynamic temporal matching for personalized predictions of parkinson's disease," in *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 2017, pp. 198–206.

[86] R. Bhidayasiri and D. Tarsy, *Movement disorders: A video atlas*. Springer Science & Business Media, 2012.

[87] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.

[88] J. H. Medicine, "How parkinson's disease is diagnosed," https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/how-parkinson-disease-is-diagnosed, 2019.

[89] B. Shahbaba and R. Neal, "Nonlinear models using dirichlet process mixtures," *Journal of Machine Learning Research*, vol. 10, no. Aug, pp. 1829–1850, 2009.

[90] R. Das, "A comparison of multiple classification methods for diagnosis of parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568–1572, 2010.

[91] P.-F. Guo, P. Bhattacharya, and N. Kharma, "Advances in detecting parkinsons disease," in *International Conference on Medical Biometrics*. Springer, 2010, pp. 306–314.

[92] D.-C. Li, C.-W. Liu, and S. C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets," *Artificial Intelligence in Medicine*, vol. 52, no. 1, pp. 45–52, 2011.

[93] F. Åström and R. Koker, "A parallel neural network approach to prediction of parkinsons disease," *Expert systems with applications*, vol. 38, no. 10, pp. 12 470–12 474, 2011.

[94] A. A. Spadoto, R. C. Guido, F. L. Carnevali, A. F. Pagnin, A. X. Falcão, and J. P. Papa, "Improving parkinson's disease identification through evolutionary-based feature selection," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 7857–7860.

[95] C. O. Sakar and O. Kursun, "Telediagnosis of parkinsons disease using measurements of dysphonia," *Journal of medical systems*, vol. 34, no. 4, pp. 591–599, 2010.

[96] A. Ozcift and A. Gulten, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. 443–451, 2011.

[97] H. Karimi Rouzbahani and M. R. Daliri, "Diagnosis of parkinson's disease in human using voice signals," *Basic and Clinical Neuroscience*, vol. 2, 2011.

[98] T. Pringsheim, N. Jette, A. Frolkis, and T. D. Steeves, "The prevalence of parkinson's disease: A systematic review and meta-analysis," *Movement Disorders*, vol. 29, no. 13, pp. 1583–1590, 2014.

[99] A. Wagle Shukla, P. Zeilman, H. Fernandez, J. A. Bajwa, and R. Mehanna, "Dbs programming: an evolving approach for patients with parkinson's disease," *Parkinson's Disease*, 2017.

[100] T. C. Sharma and M. Jain, "Weka approach for comparative study of classification algorithm," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 4, pp. 1925–1931, 2013.

[101] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, 2007.

[102] "Feature selection package - algorithms - information gain," http://featureselection.asu.edu/old/documentation/infogain.htm, 2017.

[103] Q. Zou, J. Zeng, L. Cao, and R. Ji, "A novel features ranking metric with application to scalable visual and bioinformatics data classification," *Neurocomputing*, vol. 173, pp. 346–354, 2016.

[104] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4600–4607, 2011.

[105] M. J. F. F. for Parkinson's Research, "Fox insight clinical trials nct0266835 and nct02474329," https://foxinsight.michaeljfox.org/, 2017.