EXPLORING PARKINSON'S DIAGNOSIS USING MACHINE LEARNING TECHNIQUES

RESEARCH REPORT – F21RP

Author

Supervisor

Co-Supervisor

MACS, HERIOT WATT
RICCARTON, EDINBURGH

ABSTRACT

Series of research have been done on Parkinson's disease for more than 200 years since it was first medically described as a disease by James Parkinson in 1817. With around 145 thousand active cases in the UK and 9 million all over the world, Parkinson's disease is one of the two most common incurable neurodegenerative diseases having motor and non-motor implications. Other than being incurable, lack of any biomarkers and an alarming misdiagnosis rate of 17% have necessitated to harness the potential of technology in detecting the Parkinson's disease at an early stage to mitigate its growth. For two decades now, countless attempts have been made to use the machine learning methods for an early discovery of the disease and to distinguish it from other diseases expressing similar symptoms. Although many of these research have been reasonably successful but at present there is no single identified machine learning model to serve this purpose, only a string of experiments with varying outcomes. Thus, there is a substantial prospect for further study in this area. Consequently, in this research, the author first explores an array of such initiatives, ranging from simpler machine learning techniques to intricate deep learning models. Further, by applying machine learning methods such as support vector machines, neural networks and extreme gradient boost to the dataset obtained from Leeds Teaching Hospitals NHS Trust, the author seeks to draw an analytical summary to develop a comprehensive understanding of the subject area. While reflecting on the conclusive advantages and limitations, this study intends to be a contribution to the global battle against Parkinsons.

"Success is not final, failure is not fatal; It is the courage to continue that counts."

Winston Churchill

Acknowledgements

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

TABLE OF CONTENTS

List of Figures –

Abbreviations

PD – Parkinson's Disease
NHS – National Health Services
ML – Machine Learning
RNN – Recurrent Neural Network
CNN – Convolutional Neural Network
SVM – Support Vector Machines
XGBoost – Extreme Gradient Boost
FS – Feature selection
SWEDD – Scans Without Evidence of Dopaminergic Deficit

Chapter 1

INTRODUCTION

National Health Services United Kingdom (NHS-UK) defines PD as “a condition in which parts of the brain become progressively damaged over many year. Parkinson's disease is the second most common neurodegenerative disease that results in motor and non-motor indications (“Parkinson’s disease,” 2017). Mainly, PD is evidenced by tremors (unconscious shaking of limbs), bradykinesia (slow limb movements), muscle inflexibility, motor disorder, reduced agility, hunched posture, and distorted facial expression. While, losing motor abilities over a period of time is more generic to PD patients, non-motor symptoms like depression, loss of motivation, and anxiety have also been witnessed in PD patients (Jellinger, 2003). PD is believed to be corresponding to a significant loss of dopaminergic neurons which are the primary source of dopamine (aids nerve cell communication) in the human nervous system accompanied by abnormal protein deposits in the brain called the Lewy bodies, followed by inflammation of brain neurons (Goetz, 2011). Collectively, these abnormal activities affect brain chemicals leading to impediments in basic functioning of the human brain like thinking, movement, behaviour, and mood. Although, the underlying cause of the abnormalities triggering the PD is still unknown still PD can be categorised into simply PD received through parents (hereditary) and Idiopathic PD where the causes are thoroughly unknown or sometimes environmental factors (Mei et al., 2021).

Worldwide, it accounts for a significant number of disability-induced life years and deaths, resulting in an extremely high demand for relevant health resources. The condition affects approximately 10 million people worldwide, accounting for around 1% of the total population(“What is Parkinson’s? | Parkinson’s Europe,” n.d.). As the incidence of Parkinson's disease increases with age and people are living longer lives, the prevalence of Parkinson's disease is expected to rise dramatically in the future. According to the Global Burden of Disease Study 2015, there could be nearly 13 million people with Parkinson's disease by 2040 (“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,” n.d.).

Compounding these statistics with the fact that there is currently no cure for Parkinson's disease, monitoring the occurrence of Parkinson's disease, diagnosing it at an early stage, distinguishing it from other parkinsonian syndromes and other diseases, and tracking its response to treatment and progression is of paramount importance (Emamzadeh and Surguchov, 2018a)(Emamzadeh and Surguchov, 2018).

Pertinently, everything boils down to cautious diagnosis of the disease. Often the non-motor signs in PD patients are present way before than the motor symptoms but for the clinical diagnosis to be accurate, it is based on the motor symptoms because the non-motor symptoms are frequently difficult to detect and are usually attributed to other diseases and age (Pan et al., 2012). Additionally, no specific test has been developed for PD so far due to the lack of exact causes. So, it is a common medical practice to take all the aspects of human conditions into account like medical history, examinations, tests, mental tasks, etc to conclude on PD (Goetz, 2011). Advanced medical assessments such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) Scans, Positron Emission Tomography (PET) Scans and other radiographic techniques are also utilised to observe the brain activity, especially to rule out other medical conditions (Emamzadeh and Surguchov, 2018).

1.1 Current Knowledge Gaps

There is substantial gap in the diagnosis of PD as it all depends on how well and early the above-mentioned medical assessments are carried out. So, it can be said that other than being an incurable disease, patients are also suffering due to late or misdiagnosis of their condition. According to a study, approximate rate of misdiagnosis of PD is 10-17% and it usually takes around 3 years to reach a 90% diagnostic accuracy (A J Hughes et al., 1992; Hughes et al., 2002).

1.2 Motivation

To bridge this gap, several ML techniques have been implemented and a number of other ML aided models are being introduced to increase the accuracy of diagnosis of PD. Artificial Intelligence follows the rules and expertise derived from human knowledge. Characteristically, ML is a more adaptable and data-driven subfield of AI where a computational machine gains experience and improves its performance through iterative processes for a certain task (LeCun et al., 2015). Deep learning is another field of ML which works on a connected network of neurons inspired by the working and structure of human brain. The modern ML techniques have contributed to early diagnosis of PD by training ML algorithms on data collected from PD patients. Typically, symptoms are recorded by using wearable and other motion detecting devices which transform the data into usable forms. The data when fed to advanced machine learning techniques can potentially identify the patterns in the symptoms which might go undetected by the experts. Over the last few decades, ML techniques such as k nearest-neighbours (k-NN) algorithm and Bayes classifier, regression trees (RT), support vector machines (SVM), decision trees (DT), naïve Bayes (NB), artificial neural networks (ANN), and k-Means clustering have often been used to design the computer-based diagnosis systems for PD diagnosis.

A few studies have been put together which involve developing models for drawing data collected by asking patients to perform straightforward drawing tasks on electronic tablets. The data produced is analysed through the models to find the patterns and features to establish PD. (A V et al., 2021) used the drawing data as a time series, and after working on it using RNN and CNN techniques of deep learning, observed that the CNN approach was more balanced. (Lahmiri et al., 2018) investigated the data obtained from drawing tasks performed by random PD and non-PD subjects. The dataset was used to train a CNN deep learning architecture, which achieved an accuracy of 93.53% on previously unseen data when used to distinguish PD subjects from healthy controls.

1.2 Aim

The aim of this paper is to –

- explore the PD diagnosis using ML techniques.
- assess the current state of knowledge regarding the use of ML algorithms applied to data gathered from different sources for the measurement of PD symptoms.
- locates the knowledge gaps in the literature.
- examine the application of the most common independent ML techniques.
- analyse the individual models against a few ensemble models which are being developed for PD and are gaining popularity in the clinical world.
- apply SVM, ANN, CNN and others to the data set acquired from Leeds Teaching Hospitals NHS Trust.

By performing this study, the author wants to comprehend the potential and limitations of ML techniques being utilised as a PD diagnostic tool.

Chapter 2

LITERATURE REVIEW

2.1 Literature Search Strategy

As the validity of a review is strongly correlated to the robustness of the search (Dignen, 2008), a thorough search method was used to identify the relevant existing literature on the topic.

2.1.1 Databases

The author made a list of the appropriate databases and chose three of the most suitable academic citation databases to ensure the papers' high quality and effect. Largely, the papers were fetched from the most popular databases of biomedical literature such as PubMed & MEDLINE ProQuest. These databases were selected because PubMed is a valuable source for biomedical publications and MEDLINE ProQuest is the most common database for highlighting health care interventions (Lu, 2011). Also, another recognised database of technical literature, IEEE Xplore was used as well. IEEE XPlore is the go-to reference for engineering research of all kinds. It includes millions of full-text journal articles from all disciplines of engineering ("A-Z Databases: Computer Science; IEEE,").

2.1.2 Keywords

The author chose the most pertinent keywords for the search in order to guarantee the incorporation of studies that provide solutions to the concerns raised in the research. Primary keywords used to search for the research papers were "parkinson's disease", "diagnosis", "technologies", "techniques", "methods", "machine learning", "algorithms", "SVM", "CNN", "k-means" "ensemble", "neural nets". Most of the queries submitted to the databases were of the form – "parkinson's disease" AND "diagnosis" AND "technologies" OR "methods" OR "machine learning". These keywords were used across all the databases.

2.1.3 Screening

Only peer reviewed academic studies published in English language were opted to retrieve the possible literature. Grey literature was also needed to augment the search using "google scholar" which provided a straightforward approach to browse widely for academic articles (Lu, 2011). The author used *Zotero* referencing manager to retain an electronic copy of all bibliographic references, searches and search results.

2.2 Parkinsons – Primary Symptoms

PD is characterised by more than 40 different symptoms (Pan et al., 2012). From tremors or stiffness to issues with sleep and mental health, these symptoms can manifest in a variety of ways but remarkably, experiences of each individual are unique and can vary even by day and hour. A list of a few of the most commonly seen symptoms include tremor, rigidity (stiffness), slowness of movement, mild memory, thinking problems, sleep problems, pain, and mental health problems including anxiety and depression (Hausdorff et al., 2000). Nonetheless, PD symptoms are frequently classified as either motor or non-motor. As this research paper revolves around the drawing data extracted from the hand-tasks performed by the PD patients, so the author wants focus on motor symptoms of PD especially the tremors.

2.2.1 Tremor and Bradykinesia

A tremor is an involuntary shaking of a body part, most commonly the hand. Along with slowness of movement and stiffness, tremor is one of the main signs of PD. Because everyone has a different experience, having a tremor does not always mean that it's PD. It can also be an indication of something else being wrong.("Parkinson's symptoms," n.d.)Essential tremor is a trembling of the hands, head, legs, body, or voice that is most noticeable when the individual is in motion. This common tremor is sometimes misdiagnosed as PD. Dystonic tremor can appear in people who have dystonia (a range of movement disorders that cause muscle spasms and contractions). It may be challenging to differentiate between an essential tremor, a dystonic tremor, and a PD tremor. It may be challenging to differentiate between an essential tremor, a dystonic tremor, and a PD tremor. That's the primary objective of the data scientists currently working on PD to assist in accurately point out the PD symptoms from other similar looking symptoms.

A tremor caused by Parkinson's can appear in two ways:

Resting tremor	Action tremor
Resting tremors can occur even when the patient is relaxed, still, or lying in bed.	The patient experiences action tremor when doing voluntary actions.

(Gallicchio et al., 2018; "Parkinson's symptoms,")

Bradykinesia which means the slowing of any voluntary movement and a gradual decline in scale and momentum as the movement continues.

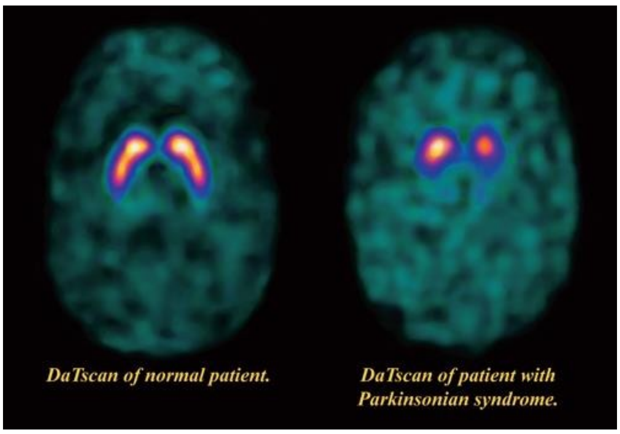
The reason for author's digging into the tremor and Bradykinesia here is that the International Parkinson and Movement Disorder Society has offered a set of criteria that might be thought of as a modernised version of the Queen's Square Brain Bank Criteria, which have been the gold standard for decades. These criteria are based on a professional neurological exam that reveals PD by confirming presence of bradykinesia and at least one other cardinal motor characteristic, such as stiffness or resting tremor (Tolosa et al., 2021).

2.3 Clinical Methods used for PD Diagnosis

PD affects balance, coordination, and mobility of the affected person. PD has no definite test, making diagnosis challenging. Instead, medical history, physical examination, and symptoms are used to diagnose. The most prevalent PD diagnosis process starts with doctors asking the patient about all the symptoms and PD in the family followed by physical examination to closely look and make eye-judgements of tremors, stiffness, slow movement where the posture, balance, and coordination is also checked. Afterwards, designated brain imagery are involved to assess the brain function, such as CT Scan, MRI Scan and SPECT scan (bgdteam, 2022).

A special type of imaging called *DaTSCAN*, is growing popular and has been widely used by medical fraternity in indicating PD. In this imaging, a scanner detects a dye injected into the patient's bloodstream. The scan can then detect the decrease of brain dopamine transporter. DaTSCAN has been considerably effective with a sensitivity of 98% and a specificity of 67% in early PD diagnosis (de la Fuente-Fernández, 2012). It has shown considerable increase in the specificity to 94% after confirmation of clinical diagnosis. The clinical diagnosis is 84 percent accurate in the early stages of PD and 98 percent accurate in the later stages (de la Fuente-Fernández, 2012).

Brain imagery from DatSCAN [Image source – (de la Fuente-Fernández, 2012)



2.4 Challenges in Clinical Diagnostic Of PD

Most frequent dispute in the diagnosis of PD has been the misdiagnosis. According to a recent meta-analysis, only 86.6% of 11 clinico-pathological investigations had a pooled diagnostic accuracy for the clinical diagnosis of PD (Giovanni Rizzo et al., 2016). Even when using strict clinical diagnostic standards, 10% of patients who were given PD diagnoses by neurologists actually had other illnesses. Moreover, misclassification is quite widespread with error rates of 15-24% in different series (Tolosa et al., 2021). Clinical misdiagnoses include essential tremor and secondary parkinsonism. The greatest challenge by far has been differentiating PD from a variety of neurodegenerative disorders (atypical parkinsonian) where Parkinsonian syndrome is one of the most important clinical feature, but the full underlying medical spectrum differs fundamentally from PD.

2.5 ML Techniques Used For PD Diagnosis

Since the introduction of ML algorithms to the context of PD, they have played an increasingly important role in its diagnosis and treatment. ML have progressively been more recognised by the medical fraternity because it can aid in the analysis of large amounts of patient data more quickly and accurately than the conventional diagnostic methods. The symptoms of PD can be ambiguous and similar to those of other conditions, making diagnosis difficult. Consequently, ML algorithms can be trained on large patient datasets to identify PD-specific patterns and markers. Predictive models to help identify patients at risk of developing PD can also be developed using ML algorithms. As a result, doctors will be able to intervene sooner, potentially halting the disease's progression or slowing it down. Depending on the needs of a particular study or clinical application, the most popular ML algorithm used in PD diagnostic may differ. To list a few –

- Support Vector Machines (SVM)
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- K-Nearest Neighbour (KNN)
- Hybrid methods
- Linear regression
- Logistic regression
- Decision trees
- Random forest
- Naïve bayes

2.5.1 SVM and PD Diagnostics

Support Vector Machines (SVM) is an ML technique which classifies new cases using non-probabilistic binary classifiers. It is a supervised kernel-based technique that first analyses known data and then classes unknown test samples. The linear SVM classifier divides examples into two locations in space and maximizes the gap width. SVMs classify or regress using infinite-dimensional hyperplanes. Classifier error is lowest for a hyperplane with largest separation from nearby training data points. These hyperplanes are built by the help of the data points which are closest to the hyperplane, these datapoint are called Support Vectors.

SVM can be executed in a linear or non-linear fashion. When a satisfactory fit cannot be obtained using the linear margin hyperplane, non-linear SVM performs better. Non-Linear SVM maximizes hyperspace by kernel-transforming feature space (S. Shetty and Y. S. Rao, 2016).

SVM has become remarkably useful for PD diagnosis due to its high-dimensional data handling, robustness to overfitting, non-linear classification, robustness to outliers, well-established theory and implementation, and potential interpretability (Karapinar Senturk, 2020).

(S. Shetty and Y. S. Rao, 2016) considers an approved and previously verified data set containing 12 features related to human gait cycle. The data set consists of records from 64 patients obtained from ‘The National Institutes of Health-sponsored Research Resource for complex physiological signals’(Hausdorff et al., 2000). SVM technique was applied to the data to classify PD and healthy patients.

During the pre-processing of data, Feature Selection (FS) which identifies the key features of the problem was employed. It was found there was a complete negative correlation among some features and were hence discarded to avoid having redundancies. Improving classification precision depends heavily on an accurate determination of characteristics and dimensionality reduction can boost the overall efficiency of ML techniques (Jain and Singh, 2018). A good correlation was observed between 10 features, so only those were used for the actual experiment. Before actually testing the data using the SVM, feature vectors were calculated using a host of statistical tools. Further reduction of the preliminary vectors was done by computing the SVM for each vector individually. The author of this paper developed the classifier model using MATLAB.

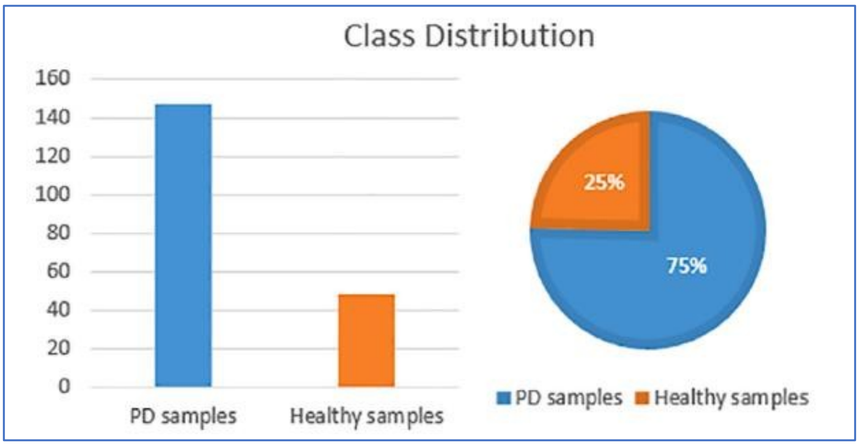
Result – The classifier achieved an accuracy of 83.33%. The model gave 75% true positive and 25% false negative outcomes.

Evaluation – Even though FS was used before applying the classifier to the dataset, accuracy was not up to the mark. The author discusses SVM further.

2.5.2 ANN and PD Diagnostics

Neural networks (also known as neural nets) are a subset of machine learning that are at the heart of deep learning algorithms. They are also known as artificial neural networks (ANNs) or simulated neural networks (SNNs). Their name and form are inspired by the human brain, and they replicate the way biological neurons communicate with one another. ANNs are made up of node layers, which include an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, links to another and has its own weight and threshold. If the output of any particular node exceeds the defined threshold value, that node is activated and sends data to the next tier of the network. Otherwise, no data is transmitted to the next network layer ("What are Neural Networks?," IBM).

(Karapinar Senturk, 2020) used a dataset created by Max Little with the cooperation of the National Voice and Speech Centre of the University of Colorado and the University of Oxford. The dataset was taken from the UCI – Machine Learning Repository ("UCI Machine Learning Repository: Parkinsons Data Set,"). The dataset consists of 195 biomedical sound measurements taken from 31 people consisting of 8 healthy subjects and 23 with PD.



Class distribution in the dataset [Image source – (Karapinar Senturk, 2020)]

Three different classification algorithms were applied to this data – CART, SVM & ANN and the end results were compared based on multiple criteria.

Classification and Regression Tree (CART)

The CART method is a nonparametric approach in identifying the most significant independent and interactive variables. If the outcome is a continuous variable, CART makes regression trees. If the outcome is a categorical variable, CART makes classification trees. If the outcome is a continuous variable, CART makes regression trees. If the outcome is a categorical variable, CART makes classification trees (Yohannes and Webb, 1999).

Further, as good practice, during the pre-processing FS was done on 23 voice features and just effective features were used, reducing the analysis cost. Different FS methods were used such as Feature Importance (FI) method for CART and Recursive Feature Elimination (RFE) method for SVM and ANN. For an efficient classification, the dataset was rearranged so that there were fewer columns (features).

Also, interestingly the whole experiment was designed in such a way that while comparing the 3 different classifying methods, benefit of application of FS was also evaluated (discussed in the results). The experiments were performed using Python programming language and its libraries Keras, Tensorflow, and sci-kit learn using 7 features for CART and 13 features for SVM and ANN methods.

Result –

Classification method	Accuracy before FS	Accuracy after FS
CART	85.23%	90.76%
SVM	79.98%	93.84%
ANN	80.25%	91.54%

Evaluation – It is clear from the results of the experiment that although SVM with FS had the best performance at classifying (93.84%) but ANN also performed (91.54%) reasonable as well. Again, we have established the efficient application of SVM to PD diagnosis and simultaneously established the advantage of careful use of FS in data pre-processing. Further, from these two experiments it can be said that SVM classifier is fairly successful for PD diagnosis.

2.5.3 CNN and PD Diagnostic

Convolutional Neural Network (CNN) is a well-known form of ANN which is primarily used in the field of pattern recognition within images (O'Shea and Nash, 2015). CNNs are different from other neural networks based on the way they work better when images, speech, or audio signals are fed into them. They are made up of three main types of layers – Convolutional layer, Pooling layer and Fully-connected (FC) layer. The first layer of a convolutional network is the convolutional layer. Even though more convolutional layers or pooling layers can come after convolutional layers, the fully connected layer is the last layer. With each layer, the CNN gets more complicated and can recognize more parts of the image. The first layers are all about simple things like colors and edges. As the image data moves through the CNN's layers, it starts to recognize larger parts or shapes of the object until it finds the right one ("What are Convolutional Neural Networks?," IBM).

(Gil-Martín et al., 2019) did a study where CNN was applied to a public dataset (Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet dataset) comprising of 77 unique spiral drawings from 15 healthy and 62 people diagnosed with PD. The CNN model had two parts: The first part was composed of two convolutional layers, considering 16 filters with dimensions 1*5 with an intermediate Maxpooling layer between the convolutional ones, to extract the main features from the inputs. The second part included three fully connected layers for classification. A dropout layer was also introduced to avoid over fitting. The inputs were compiled in a 2 D matrix with N*125 dimensions. And because output layer classifies between two classes only, it provides one output with a sigmoid function and uses the binary cross-entropy as loss metric. The primary experiment employed 25 epochs, 100 batches, and ReLU as the activation function for the deep learning framework. Root-mean-square propagation was fixed as the optimizer.

Result – After employing a five-fold cross-validation to split the data into train & test data and repeating the experiment five times modifying the test and training sets each time, the system reported an accuracy of 96.5% and F1-score of 97.7%.

Evaluation – On comparing the results with similar based previous experiments carried out on the same dataset, (Gallicchio et al., 2018) obtained an accuracy of 89.3% while (Khatamino et al., 2018) reported an accuracy of 72.5% for the subject-wise cross-validation. The authors believe the use of the spectrum points as inputs to the CNN instead of the raw data has wholly improved the output accuracy.

2.5.4 Multi-class SVM and PD Diagnostics

A similar study was done by (Singh et al., 2016) where MRI data was fed into multi-class SVM for classification of PD and non-PD patients. The data was obtained from Parkinson's Progression Markers Initiative (PPMI, a public repository) and contained 150 MRIs, 50 each for healthy control (HC), *de novo* PD and Scans Without Evidence of Dopaminergic Deficit (SWEDD) subjects.

FS method called principal component analysis (PCA) was employed for feature extraction and reduction in data-dimensionality from the high-dimensional imaging data. Fisher discriminant ratio (FDR) was applied for choosing features with higher discriminative ability, considering only two classes at once. The ability of a diagnostic test to differentiate between healthy and diseased patients is known as its *discriminative ability* and the discriminative ability of a diagnostic procedure is called diagnostic accuracy ("Feature Importance," codeacademy). After arranging the features in descending order based on their combined FDR scores, a binary and multi-class SVM was applied to the MRI (image) data.

Utilizing 10-fold cross-validation, 9 parts i.e., 45 patients for each subject class – PD, HC & SWEDD, were used for training and the rest i.e., 5 patients for each subject class for testing. Evaluation of the performance of the methodology was based on the accuracy for every comparison identifying the grey matter (GM) & the white matter (WM) in the brain.

Result –

Overall mean accuracy for *binary classification* was approximately 99%, 96% and 96% for GM brain images and approximately 88%, 94% and 94% for WM.

Overall mean accuracy for *multi-class classification* were 86% and 87% for GM & WM respectively.

Evaluation –

Evaluating the performance of the methodology based on the accuracy, a similar study was done by (Duchesne et al., 2009) where an accuracy of 91% was achieved, which is relatively lower than (Singh et al., 2016).

2.6 Ensemble Models

2.6.1 (RF and XGBoost)

A *random forest* (RF) is made up of numerous separate decision trees that work together to form a combination. Each tree in the random forest produces a class prediction, and the class with the most returns becomes the model's prediction (Yiu, 2021).

Extreme Gradient Boosting (XGBoost) is a supervised tree-based ML algorithm. It includes a scalable and distributed gradient-boosted decision tree that offers concurrent tree boosting and is an important ML library for regression, classification, and ranking issues (Chen et al., 2015).

(Xing et al., 2022) conducted an experiment on 398 patients with confirmed upper limb tremors. The data was obtained from department of Neurology of Rui Jin Hospital (Shanghai, China). ML techniques namely random forest (RF) and eXtreme gradient boosting (XGBoost) among others were applied in combination for classification of either PD or essential tremor (ET). During data preprocessing, 40 tremor variables and two demographics (sex and age) were used as the variables for training. To find the optimal parameter combination for each model, parameters were selected and altered using the grid-search approach, and then randomly split into a training set (80%) and a validation set (20%).
The training set was subjected to ten-fold cross-validation in order to determine the ideal model parameters. The training set was split into 10 pieces, of which nine were utilised to train the model sequentially and the remaining one to test it.

Result –
RF and XGBoost showed a competent performance with an accuracy rate of 84%.

2.6.2 Hybrid Models

(Varalakshmi et al., 2022) built a variety of ML, DL, and hybrid models to determine the best model for classifying PD in the earliest stages. A spiral hand drawing dataset containing 51 HC and 51 PD subjects was used, obtained from Kaggle repository.

A *hybrid model* is the process of merging two or more models to produce new models. Hybrid models can be designed by merging DL networks with ML frameworks or DL with DL models (Drotár et al., 2016).

Results –

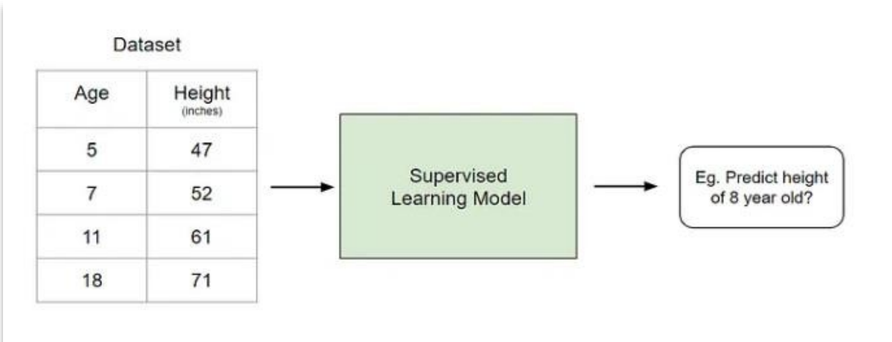
At the end of the experiment, following results were observed –

Model type	Model with highest accuracy	Accuracy (%)
ML	SVM	82.40
DL	RESNET50	96.79
Hybrid (ML + DL)	RESNET50 + SVM	98.45
Hybrid (DL + DL)	RESNET + MLP	98.36

2.7 Prospected ML models

Here, the author wants to elaborate a few prospected ML models that might be utilized in the development of the prototype for PD diagnosis. Based on the requirements of the experiment, all ML models are categorised as supervised and unsupervised. Supervised models are further divided into regression and classification models (Shin, 2022).

In supervised learning, a function is created which learns by analysing pairs of inputs and outputs and accordingly maps a new unseen input to an output.

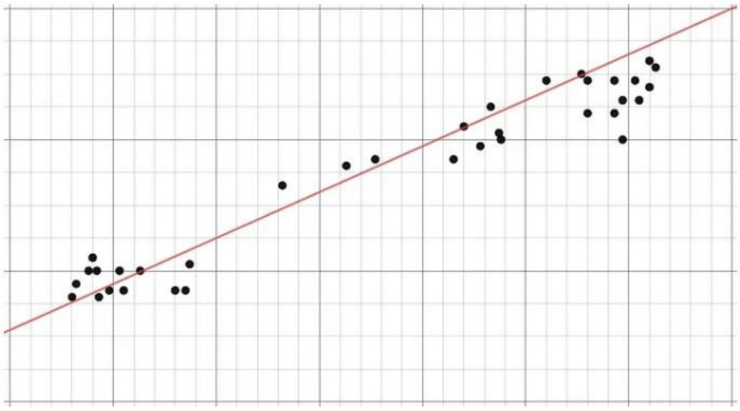


Example of working of supervised learning [Image source – (Shin, 2022)]

From supervised learning, the author will consider –

2.7.1 Linear Regression

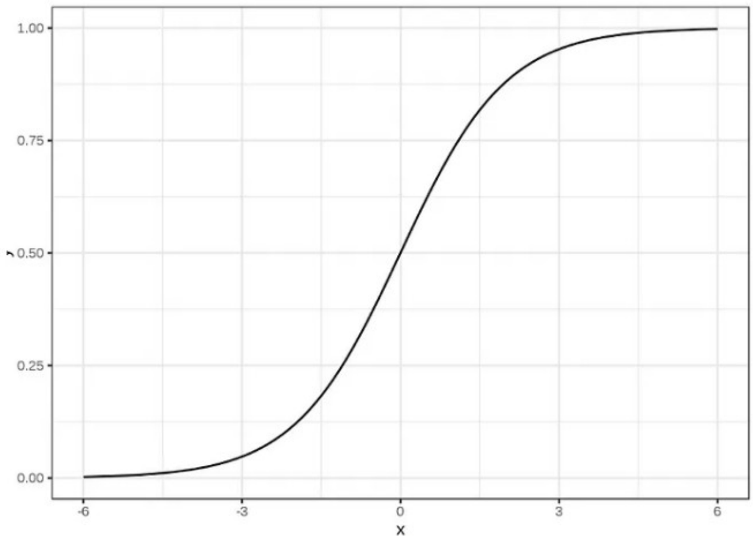
Finding a line that best fits the data is the concept underlying linear regression. Linear regression is a frequent and straightforward technique for predicting numerical values based on input information (Lahmiri et al., 2018). It assumes a linear relationship between input data and the target variable and seeks to choose the best-fitting line that minimizes residual errors. Widespread applications for linear regression include predicting housing prices, clinical data, stock prices, and other numerical quantities (Rana et al., 2022).



Graph showing Linear Regression [Image source – (Shin, 2022)]

2.7.2 Logistic Regression

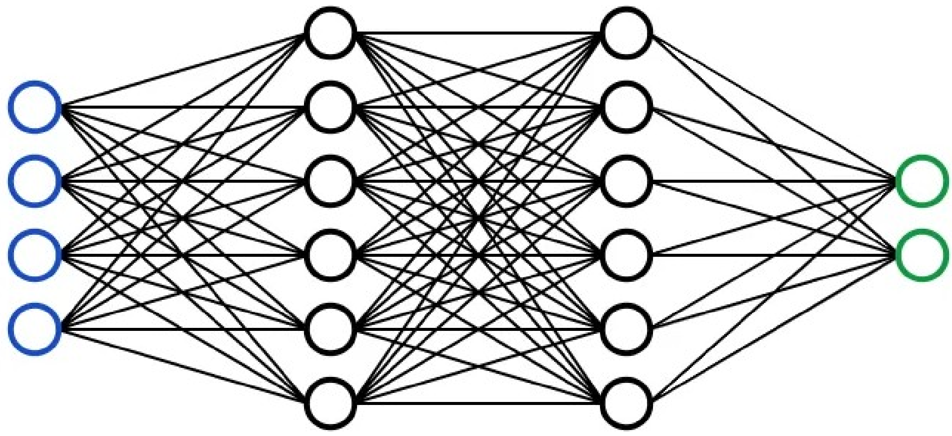
Logistic regression is similar to linear regression in that it models the probability of a finite number of outcomes, often two, hence it is mostly used for binary classification assignments in which the goal is to classify cases into one of two groups. It employs a logistic function to evaluate the likelihood of an instance belonging to a specific class. Logistic regression is frequently used in tasks including spam classification, churn prediction, and medical diagnosis. Basically, a logistic equation is made so that the outputs can only be between 0 and 1 (see below).



Graph showing Logistic Regression [Image source – (Shin, 2022)]

2.7.3 Neural Networks

Neural networks are a sort of deep learning model that can learn complicated patterns from enormous volumes of data. They are made up of numerous interconnected layers of nodes (neurons) and can be utilized for a variety of tasks such as image identification, natural language processing, and speech recognition. Deep learning models, particularly neural networks, have demonstrated cutting-edge performance in a variety of disciplines. A Neural Network is essentially a network of mathematical equations. It takes one or more input variables and produces one or more output variables after passing them through a network of equations.



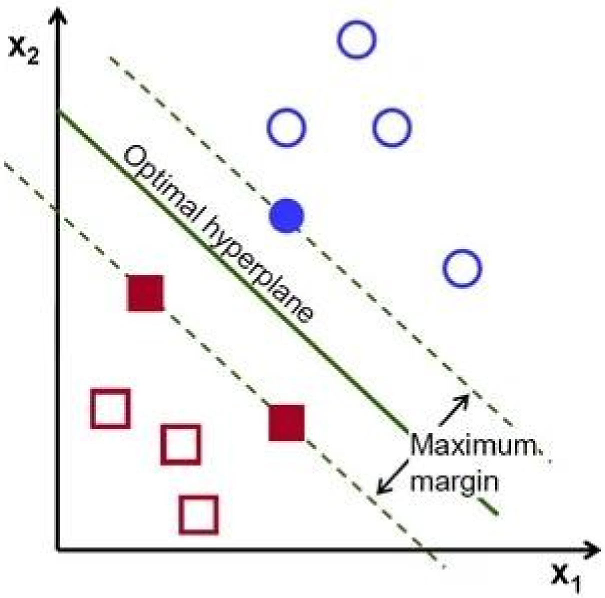
Multi-Layer Neural Network [Image source – (Shin, 2022)]

The blue circles represent the input layer, the black circles represent the hidden layers, and the green circles represent the output layer. Output is shown in green circles, and each hidden layer node represents a linear function and activation function that the preceding layer nodes traverse.

2.7.4 Support Vector Machine (SVM)

SVM is a form of supervised classification that, at its most advanced level, can be rather difficult to grasp, but at its most fundamental level, it is relatively simple to grasp. It locates a hyperplane or a border between the two classes of data in a way that maximizes the difference in margin between the two classes. There are a number of planes that are capable of putting some distance between the two classes; however, there is only one plane that can maximize that margin or distance. It is mostly used for binary classification tasks. Its goal is to find the optimum hyperplane that separates instances of distinct classes by the greatest possible margin.

Additionally, SVM can be utilized for multi-class classification and regression analysis. Image identification, text categorization, and bioinformatics are just a few examples of the applications that make extensive use of SVM, which is famous for its capacity to manage large amounts of data.



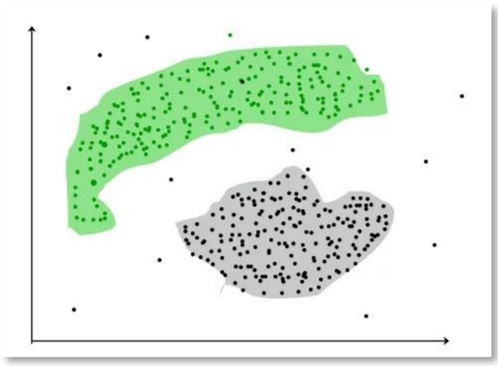
Hyperplanes in SVM [Image source – (Shin, 2022)]

From unsupervised learning, the author will consider –

2.7.5 Clustering

Clustering is an unsupervised approach for grouping or clustering data elements. Customer segmentation, fraud detection, and document classification are all common applications. K-means clustering, hierarchical clustering, mean shift clustering, and density-based clustering are all common clustering techniques. While each methodology uses a different way to locate clusters, they all aim for the same result. Clustering algorithms like K-means, hierarchical clustering, and DBSCAN are used in unsupervised learning tasks to organize instances into similar clusters based on their similarity or distance.

Data bifurcated into separate clusters [Image source – (Shin, 2022)]



2.8 Comparison of ML models

Comparative studies of ML approaches to diagnose PD.

ML algorithm used	Type of Data	No. of test subjects	Accuracy (%)	Reference
SVM	Gait cycle	64	83.33	(S. Shetty and Y. S. Rao, 2016)
SVM with FS	Speech	23 + 8	93.84	(Karapinar Senturk, 2020)
Multi-class SVM with 10-fold cross-validation	MRI	50 + 50 + 50	86-87	(Singh et al., 2016)
ANN with FS	Speech	23 + 8	91.54	(Karapinar Senturk, 2020)
SVM, KNN, Ensemble AdaBoost***	Handwriting	37 + 38	81	(Drotár et al., 2016)
CNN***	Handwriting	438 + 207	97.2	(Wenzel et al., 2019)
CNN with 5-fold cross validation	Hand drawings	62 + 15	96.50	(Gil-Martín et al., 2019)
CART with FS	Speech	23 + 8	90.76	(Karapinar Senturk, 2020)
Ensemble RF and XGBoost	Tremor	398	84	(Xing et al., 2022)
RESNET50	Hand drawings	51 + 51	96.79	(Varalakshmi et al., 2022)
RESNET50 + SVM (Hybrid)	Hand drawings	51 + 51	98.45	(Varalakshmi et al., 2022)
RESNET + MLP	Hand drawings	51 + 51	98.36	(Varalakshmi et al., 2022)

*** are not discussed in this paper but are included in the table for comparison purposes.

2.9 Conclusion

The author concludes this literature review by observing that ML techniques such as SVM, ANNS, CNNs, and other deep learning algorithms are being increasingly used in PD diagnosis. By offering more precise and effective diagnostic tools, the application of ML approaches in PD diagnosis has the potential to revolutionise the clinical practice. These methods can improve early PD identification, individualised PD treatment strategies, and the disease progression tracking, which would eventually improve patient outcomes and healthcare decision-making.

Critically, even though, these techniques have shown promising results in terms of accuracy and performance on specific data sets, they had some shortcomings, including small datasets and lack of standardised diagnostic criteria, among others. Furthermore, there is a need for additional validation in larger and more diverse datasets to assess generalisability and real-world clinical usefulness of ML. Therefore, future study in this field could concentrate on overcoming these constraints.

Chapter 3

METHODOLOGY

Cambridge defines methodology as a system of ways of doing, teaching, or studying something ("methodology," 2023). Methodology is a set of time-dependent, provision-based steps or stages that are meant to make an operation or project more effective but are often hard to measure when applied to a specific problem. Accordingly, in this section, the author provides an overview of the experiment and ML approach to be used for diagnosis of PD. The objective of the experiment is to develop a ML-based method for identifying PD. Utilizing a data set obtained from Leeds Teaching Hospitals NHS Trust which contains clinical information from PD patients and healthy controls will constitute the methodology. The author will follow these steps to develop the predictive model for PD diagnostic –

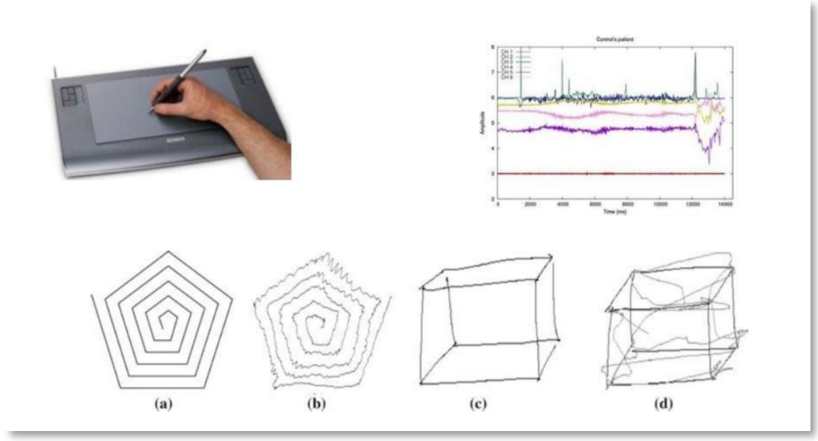
- Data Collection
- Data Preprocessing
- Feature Selection
- Model Selection and Development
- Model Evaluation
- Model Optimization
- Model Validation

3.1 Data Collection

Collecting data is the first stage in developing a methodology for the project. The author will use a dataset which was collected at Leeds Teaching Hospitals NHS Trust for this dissertation. The data has collection from 58 PD patients. The data was created by asking PD patients to copy images and draw spiral pentagons on a digitising pressure-sensitive Wacom tablet (Wacom Technology Corporation) of size (20.3*32.5) cm. Wherein each subject drew once for image copying task, their dominant hands and did four drawings for spiral pentagon tasks, two with each hand. Subjects were instructed that the figures should be drawn as quickly and accurately as feasible (Alissa et al., 2022).

As this is a student-based study and is completely non-funded, so it is not feasible to collect the primary data directly from the patients or hospitals. So, the author opted for the already collected and verified secondary data which is considered suitable for this research.

Data Gathering Method ➡

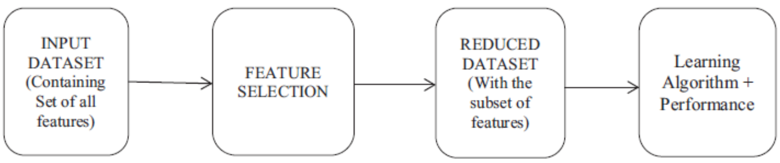


3.2 Data Preprocessing

This is the first stage after obtaining the data. Preprocessing is basically applied to ensure the quality and usefulness of the acquired data for the training of an ML model (Paleyes et al., 2022). At this stage, author will start removing the outliers and noise (unnecessary data points) from the data. Data will be standardised to guarantee that each feature is given the same weight. The data may need to be cleaned, missing values may need to be handled, and features may need to be normalised or standardised. Feature handling is discussed further in the next section.

3.3 Feature Selection

Feature selection is a widely used data preparation approach in data mining that removes unnecessary features from datasets. This technique improves data comprehensibility, visualization, learning algorithm training time, and prediction performance. Consequently, at this stage, all the relevant features that are most indicative of PD diagnosis will be selected from the preprocessed data. There are multiple methods for feature selection such as correlation-based feature selection, recursive feature elimination (RFE) and Principal Component Analysis (PCA) among others (Jain and Singh, 2018).



Steps for Feature Selection

3.4 Model Selection and Development

In many real-world situations, a model's complexity is the most important factor in choosing which ML model should be used. Even though deep learning and reinforcement learning are becoming more popular in the research world, simpler models are often used in practice (Paleyes et al., 2022). Post extensive analysis of data and selecting the appropriate features, the author will select the most effective ML model. The author might use cross-validation or grid-search techniques for the same. Some of the ML algorithms, already discussed in previous sections, will be considered, such as SVM, logistic regression, decision trees, and neural networks.

The data with selected features will be used to train the selected ML model. For that, the author will split the modelling data into training and testing samples. Model training is the process of feeding a collected dataset to the chosen model so it can learn certain patterns or ways to represent the data (Paleyes et al., 2022).

The author will use Python and its libraries to develop the ML model. One such Python code snippet for splitting the data –

> *Using scikit-learn (aka sklearn) train_test_split()*
> *Using numpy 's randn() function*
> *or with built-in pandas method called sample()*

3.5 Model Evaluation

Model evaluation is a crucial stage in ML that assesses the performance of a trained ML model on a given dataset. The purpose of model evaluation is to determine how well a model is expected to function on unseen data and to identify any faults or constraints that may impact its performance in real-world situations (Landolfi et al., 2021).

As explained in the previous section, the data will be split into training and test samples. The author will utilise the test set to assess the developed ML model's performance. As the data used for this study is collected through NHS Trust, the test set will be a representative of the real-world data that the model is expected to encounter after deployment.

Now, to quantify the performance of the model performance metrics will be used. Commonly used metrics for model evaluation include classification accuracy or accuracy, precision, recall, F1 score, area under the curve, and mean squared error (MSE), among others (Mishra, 2020).

In addition to other metrics used to evaluate a classification model, the author will initially use the three main metrics as defined below –

Accuracy is defined as the proportion of correct test data predictions. It is easily determined by dividing the number of accurate forecasts by the total number of predictions (Mishra, 2020).

$$\text{accuracy} = \frac{\text{correct predictions}}{\text{all predictions}}$$

Precision is defined as the proportion of relevant instances (true positives) among all instances projected to belong to a particular class (Mishra, 2020).

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

Recall is defined as the fraction of examples anticipated to belong to a class divided by the total number of examples that actually belong to the class (Mishra, 2020).

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

F1 score – The accuracy of a test can also be measured by its F1 Score. F1 Score is the Harmonic Mean between precision and recall, so it tries to find a balance between precision and recall. The F1 Score range is [0, 1]. It gives the number of correctly classified cases and measures the model's robustness, which means that it doesn't miss a large number of cases (Landolfi et al., 2021).

High accuracy but low recall yields an exceptionally precise result, but it misses a huge number of difficult-to-classify cases. The higher the F1 Score, the better our model's performance (Mishra, 2020).

Mathematically, it can be expressed as –

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}}$$

Mean squared error (MSE) is just the average of the squared differences between the predicted and the actual. Squared error is often used because it indicates the wrong predictions (Mishra, 2020).

$$MeanSquaredError = \frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2$$

Area Under Curve (AUC) is one of the most prevalent evaluation metrics. It is used for problems involving binary classification. A classifier's AUC corresponds to the likelihood that it would rank a randomly selected positive example higher than a randomly selected negative example. AUC is calculated by plotting the false positive rate against the true positive rate at various points within the range [0, 1].

3.5.1 Overfitting and Underfitting

The author is aware of being cautious regarding the overfitting and underfitting of the model.

In statistics, a fit refers to the measure of approximating a target function. Based on that, a terminology is used in ML to analyse the model's learning rate from the training data and its ability to generalise the new data, namely *overfitting* and *underfitting*. They are the two major causes for lower performance of ML algorithms (Brownlee, 2016).
Overfitting occurs when a model performs well on the training set but unsuccessfully on the test set, this indicates that it has memorised the training data rather than learnt to generalise from it. Underfitting, on the other hand, establishes the inadequacy of the developed model to learn and represent the underlying data patterns (Brownlee, 2016).
The author might need to employ some reducing approaches like Constraining Model Complexity. Overfitting can also simply be reduced by training the network on more examples (training data) or by just holding back a validation dataset. And underfitting can be addressed by changing the complexity of the network.

3.6 Model Optimisation

Optimisation is fundamental to ML model development. It is a theme which is recurring at every step of the ML model. In its core, training an ML model is an optimisation problem in itself, as the model is learning to perform a function in the most efficient manner. The tweaking and adjusting of model configurations or hyperparameters are the most critical aspect of ML optimisation. The components of the model that the model developer sets are known as hyperparameters (Castillo, 2021).

Therefore, the author plans to optimise the chosen model first by selecting the optimal hyperparameters and by hyperparameter tuning at the later stages. There are multiple approaches available in the ML world (Castillo, 2021) some of which might have to bring in at this point such as Random and Grid searches, Evolutionary optimisation and Bayesian optimisation for the aforementioned purpose.

3.7 Model Validation

The optimized model will be further tested and validated using additional datasets to ensure its generalisability and reliability in real-world scenarios. The author will also include the cross-validation technique to assess the performance of the ML model by repeatedly training and evaluating it on different subsets of the data by using *k*-fold cross-validation. With this method, The data is separated into *k* folds of equal size. Then, the model is trained and tested *k* times, with stratified cross-validation ensuring that the class distribution is preserved in each fold (Gallicchio et al., 2018).

Chapter 4

REQUIREMENT ANALYSIS

Prioritisation of requirements is a crucial phase of the software development process. It must be done in order to limit the chance of costly software failures. It is acknowledged as being one of the most significant decision-making processes used to examine the requirements of a project. Prioritisation provides numerous advantages, such as minimising the likelihood that precious project resources would be misallocated and facilitating clients' and developers' comprehension of a project's primary aims (Khan et al., 2015).

The author in this project seeks to develop a system based on ML for the diagnosis of Parkinson's disease. The technology will analyse patient data and make precise predictions regarding the presence or absence of PD. The objective is to assist medical practitioners with early diagnosis and to reduce the current PD misdiagnosis rate. The project will involve the development of a ML model using Python. The project will use open-source libraries such as scikit-learn for machine learning and matplotlib for data visualization.

Numerous techniques have been developed to analyse the requirements of any specific project. A careful attention was given by the author to a few of such methods such as analytic network process (ANP), scrum, agile, waterfall, analytic hierarchy process (AHP), hierarchy AHP, MoSCoW, spanning tree matrix, bubble sort, binary search tree and priority groups. Most of these methods include explanation of the methodologies. Because the methodologies have already been discussed in the previous chapter, so the author decided to use MoSCoW which has a different approach towards the requirement analysis.

Comparison of multiple techniques [Source – (Khan et al., 2015)]

Evaluation Criteria	Simple Ranking	MoScoW	100 dollar	AHP
Ratio Scale Information			Yes	Yes
High Confidence from User	Yes	Yes	Yes	
Consistent	Yes	Yes	Yes	Yes
Low difficulty	Yes	Yes	Yes	
Low effort	Yes	Yes	Yes	
Able to handle large number of alternatives		Yes		

4.1 MoSCoW

(Museum of Soviet Calculators on the Web) MoSCoW is an acronym that stands for four hierarchical priority groupings. Each requirement in a group is assigned the same priority. This method categorizes requirements into four groups based on their priority: MUST-have, SHOULD-have, COULD-have, and WON'T-have (Black, 2020). Each criterion will be assigned to one of the groups depending on its relative importance.

MUST-*have* – signifies that this group of requirements must be implemented in the software prior to its release. These requirements are non-negotiable; failing to meet them will result in the failure of the entire project.

SHOULD-*have* – indicates that the product/software will benefit from the implementation of these requirements. Optional features that would be a plus if they could be included.

COULD-*have* – means that if requirement from this group exist then it will be good for the project/software. Features that would be nice to have if at all possible but slightly less advantageous than the SHOULD-have.

WON'T-*have* – says that these needs cannot be implemented in the current iteration since they are low priority. This is sometimes referred to as a "wish list." These criteria are not inconsequential, but they may be implemented at a later time.

(Black, 2020; Overby, 2021)

Therefore, the author has categorised the *specific requirements* of this project into four groups stated below –

MUST-have

Requirement	Priority
Data Collection	*High*
Data Preprocessing	*High*
Feature Selection	*Medium*
Model Selection and Development	*High*
Model Evaluation	*High*
Model Optimization	*High*
Model Validation	*High*

All the stages of the project explained in the methodology section of this report are important for its successful completion. So, the stages of getting access to the dataset, choosing the right ML algorithms, putting data preparation and preprocessing techniques into practice, including data cleaning and feature selection, assessing the performance of the developed model using the right metrics, and validation of the model are non-negotiables (must-haves) for this project.

SHOULD-have

In addition to the above-mentioned non-negotiable elements, the author is planning to utilise various techniques to improve the project's overall performance by comparing various model evaluation strategies, researching multiple ML algorithms to find the one that is most suitable for this project, and exploring various feature extraction techniques.

Requirement –	Priority
Comparison to other models	*High*
Research multiple ML methods	*Medium*
Feature Extraction Techniques	*Medium*

COULD-have

The author wants to test the model on multiple datasets from different sources to improve the accuracy and reliability of the diagnostic model. Further, ensemble models could be developed by combining different ML methods to enhance the model's performance. Incorporation of complex deep learning or neural network models might also be considered.

WON'T-have

Commercial datasets or tools that are unavailable or impractical for the study. In addition, the system will not have a user interface.

4.2 Hardware Requirements

The author will be using a personal computer having an Intel Core i7 processor with 500 GBs of hard disk and 16 GBs of ram which should be sufficient for carrying out coding part of the experiments.

Chapter 5

PROFESSIONAL, LEGAL, ETHICAL AND SOCIAL ISSUES

5.1 Professional Issues

The author will write and test the code while adhering to British Computing Society (BCS) code of conduct. Code will be written to a high standard and commented throughout for clarity. There will be adequate documentation presented. All third-party software, libraries, and other goods will only be utilized if their licenses let it. Any citations or outside data will be properly referenced.

5.2 Legal Issues

PD diagnosis requires the use of sensitive patient data, so the author will ensure proper data privacy and security measures are followed in accordance with GDPR UK.

5.3 Ethical Issues

In general, when working with the data obtained from real subjects, it is crucial to take ethical factors such as informed consent, data privacy, data security, fairness, and transparency into account (Lamba et al., 2022).

In particular, the author will ensure that –

- the research is conducted with integrity and transparency.
- the data is managed carefully, respecting the rights of test subjects, supervisors, and Heriot Watt University as well as their right to privacy.
- the participants' data is kept confidential.
- the information is securely kept and guarded against unauthorized access or disclosure.

Finally, the author will need to ensure that the data gathering process is handled responsibly and in accordance with GDPR UK.

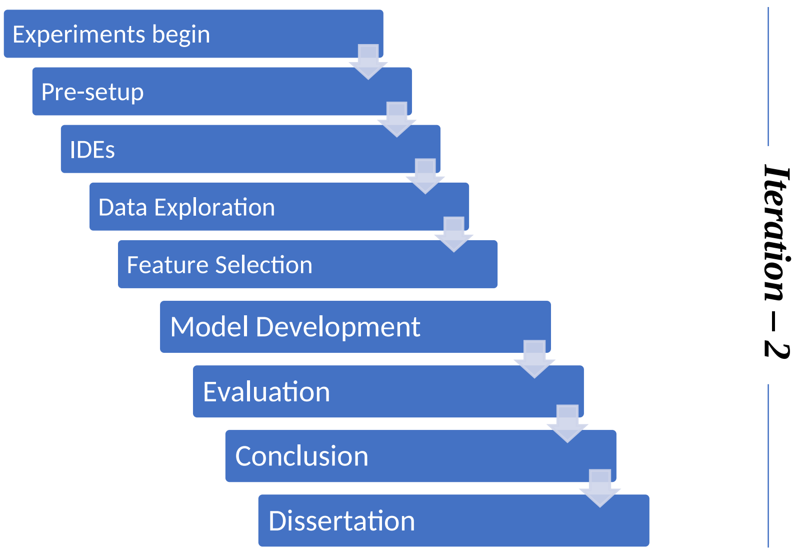
5.4 Social Issues

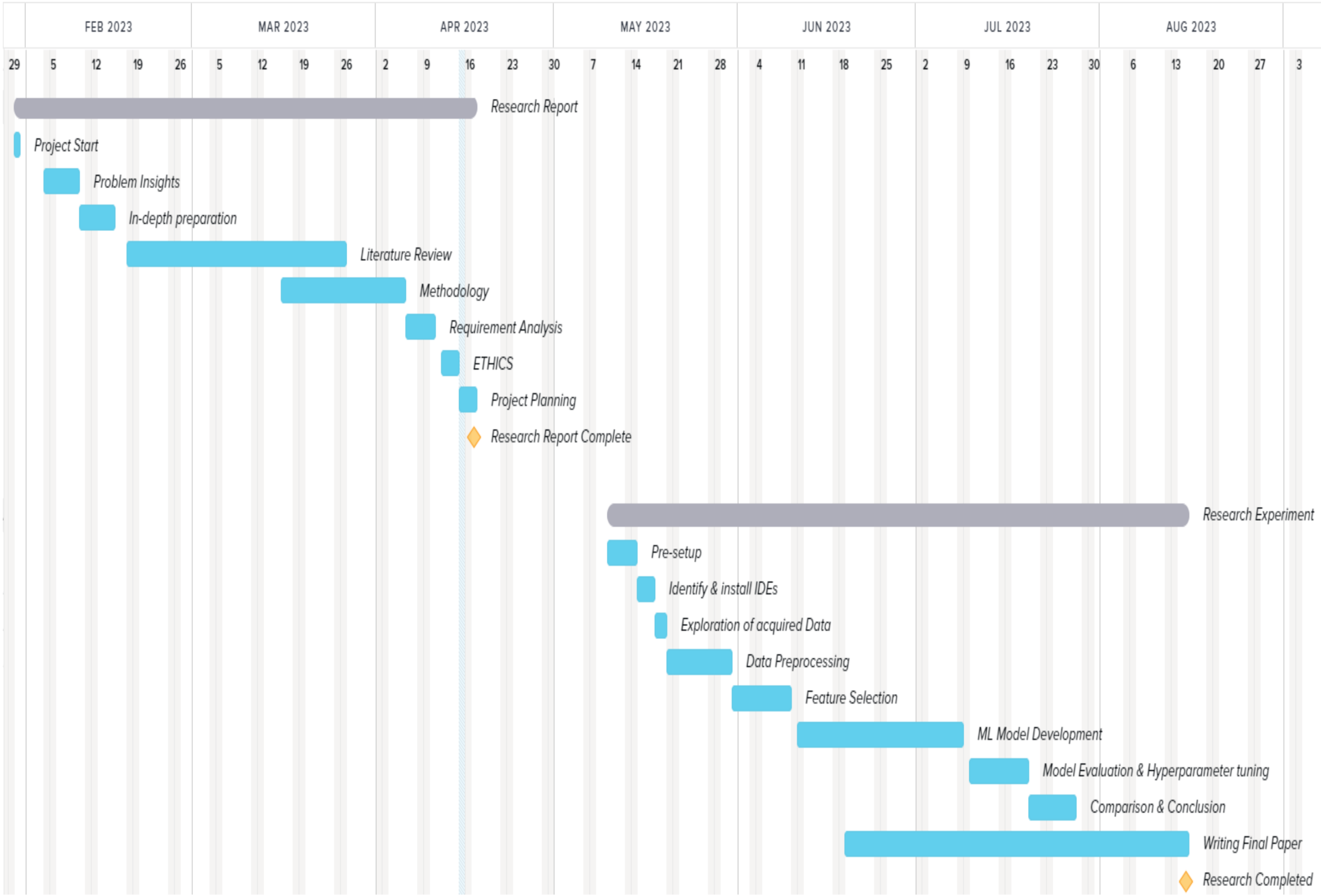
As there is no direct communication with the patients so the project does not come across any social issues.

Chapter 6

PROJECT PLAN

A well-devised plan is essential for finishing a project on schedule and within budget. The steps that must be taken and the goals that must be accomplished are laid out in this section of the project plan. However, the timeline provided is an estimate and may need adjustment based on the circumstances at a later stage.





<u>Project Gantt Chart</u>

Bibliography

A J Hughes, S E Daniel, L Kilford, A J Lees, 1992. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55, 181. https://doi.org/10.1136/jnnp.55.3.181

A V, A.S., Lones, M.A., Smith, S.L., Vallejo, M., 2021. Evaluation of Recurrent Neural Network Models for Parkinson's Disease Classification Using Drawing Data, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1702–1706. https://doi.org/10.1109/EMBC46164.2021.9630106

Alissa, M., Lones, M.A., Cosgrove, J., Alty, J.E., Jamieson, S., Smith, S.L., Vallejo, M., 2022. Parkinson's disease diagnosis using convolutional neural networks and figure-copying tasks. Neural Computing and Applications 34, 1433–1453. https://doi.org/10.1007/s00521-021-06469-7

A-Z Databases: Computer Science; IEEE [WWW Document], n.d. URL https://isguides.hw.ac.uk/az.php?q=IEEE&s=73911&p=1 (accessed 4.11.23).

bgdteam, 2022. Parkinson's Disease | Fact Sheet | Health Information | Brain & Spine Foundation. URL https://www.brainandspine.org.uk/health-information/fact-sheets/parkinsons-disease/ (accessed 4.7.23).

Black, S., 2020. How to Prioritize and Execute your Machine Learning Research [WWW Document]. Medium. URL https://towardsdatascience.com/how-to-prioritize-and-execute-your-machine-learning-research-eb6e6ce09f81 (accessed 4.14.23).

Brownlee, J., 2016. Overfitting and Underfitting With Machine Learning Algorithms. MachineLearningMastery.com. URL https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/ (accessed 4.12.23).

Castillo, D., 2021. Machine Learning Optimisation - Why is it so Important? [WWW Document]. Seldon. URL https://www.seldon.io/machine-learning-optimisation (accessed 4.12.23).

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2 1, 1–4.

de la Fuente-Fernández, R., 2012. Role of DaTSCAN and clinical diagnosis in Parkinson disease. Neurology 78, 696. https://doi.org/10.1212/WNL.0b013e318248e520

Dignen, B., 2008. English for presentations, Business minimax. York Associates International, York.

Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M., 2016. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artificial Intelligence in Medicine 67, 39–46. https://doi.org/10.1016/j.artmed.2016.01.004

Duchesne, S., Rolland, Y., Vérin, M., 2009. Automated computer differential classification in Parkinsonian Syndromes via pattern analysis on MRI. Acad Radiol 16, 61–70. https://doi.org/10.1016/j.acra.2008.05.024

Emamzadeh, F.N., Surguchov, A., 2018a. Parkinson's Disease: Biomarkers, Treatment, and Risk Factors. Frontiers in Neuroscience 12.

Emamzadeh, F.N., Surguchov, A., 2018b. Parkinson's Disease: Biomarkers, Treatment, and Risk Factors. Frontiers in Neuroscience 12.

Feature Importance [WWW Document], n.d. . Codecademy. URL https://www.codecademy.com/article/fe-feature-importance-final (accessed 4.9.23).

Gallicchio, C., Micheli, A., Pedrelli, L., 2018. Deep Echo State Networks for Diagnosis of Parkinson’s Disease. CoRR abs/1802.06708.
Gil-Martín, M., Montero, J.M., San-Segundo, R., 2019a. Parkinson’s Disease Detection from Drawing Movements Using Convolutional Neural Networks. Electronics 8. https://doi.org/10.3390/electronics8080907
Gil-Martín, M., Montero, J.M., San-Segundo, R., 2019b. Parkinson’s Disease Detection from Drawing Movements Using Convolutional Neural Networks. Electronics 8. https://doi.org/10.3390/electronics8080907
Giovanni Rizzo, Massimiliano Copetti, Simona Arcuti, Davide Martino, Andrea Fontana, Giancarlo Logroscino, 2016. Accuracy of clinical diagnosis of Parkinson disease. Neurology 86, 566. https://doi.org/10.1212/WNL.0000000000002350
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 [WWW Document], n.d. URL https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(18)30295-3/fulltext (accessed 4.7.23).
Goetz, C.G., 2011. The history of Parkinson’s disease: early clinical descriptions and neurological therapies. Cold Spring Harb Perspect Med 1, a008862. https://doi.org/10.1101/cshperspect.a008862
Hausdorff, J.M., Lertratanakul, A., Cudkowicz, M.E., Peterson, A.L., Kaliton, D., Goldberger, A.L., 2000. Gait Dynamics in Neuro-Degenerative Disease Data Base. https://doi.org/10.13026/C27G6C
Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y., Lees, A.J., 2002. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain 125, 861–870. https://doi.org/10.1093/brain/awf080
Jain, D., Singh, V., 2018. Feature selection and classification systems for chronic disease prediction: A review. Egyptian Informatics Journal 19, 179–189. https://doi.org/10.1016/j.eij.2018.03.002
Jellinger, K.A., 2003. How valid is the clinical diagnosis of Parkinson’s disease in the community? J Neurol Neurosurg Psychiatry 74, 1005. https://doi.org/10.1136/jnnp.74.7.1005
Karapinar Senturk, Z., 2020. Early diagnosis of Parkinson’s disease using machine learning algorithms. Medical Hypotheses 138, 109603. https://doi.org/10.1016/j.mehy.2020.109603
Khan, J.A., Rehman, I.U., Khan, Y.H., Khan, I.J., Rashid, S., 2015. Comparison of Requirement Prioritization Techniques to Find Best Prioritization Technique. International Journal of Modern Education & Computer Science 7.
Khatamino, P., Canturk, I., Ozyilmaz, L., 2018. A Deep Learning-CNN Based System for Medical Diagnosis: An Application on Parkinson’s Disease Handwriting Drawings. https://doi.org/10.1109/CEIT.2018.8751879
Lahmiri, S., Dawson, D.A., Shmuel, A., 2018. Performance of machine learning methods in diagnosing Parkinson’s disease based on dysphonia measures. Biomedical Engineering Letters 8, 29–39. https://doi.org/10.1007/s13534-017-0051-2
Lamba, R., Gulati, T., Alharbi, H.F., Jain, A., 2022. A hybrid system for Parkinson’s disease diagnosis using machine learning techniques. International Journal of Speech Technology 25, 583–593. https://doi.org/10.1007/s10772-021-09837-9
Landolfi, A., Ricciardi, C., Donisi, L., Cesarelli, G., Troisi, J., Vitale, C., Barone, P., Amboni, M., 2021. Machine Learning Approaches in Parkinson’s Disease. Curr Med Chem 28, 6548–6568. https://doi.org/10.2174/0929867328999210111211420

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539
Lu, Z., 2011. PubMed and beyond: a survey of web tools for searching biomedical literature. Database 2011, baq036. https://doi.org/10.1093/database/baq036
Mei, J., Desrosiers, C., Frasnelli, J., 2021. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. Frontiers in Aging Neuroscience 13.
methodology [WWW Document], 2023. URL https://dictionary.cambridge.org/dictionary/english/methodology (accessed 4.9.23).
Mishra, A., 2020. Metrics to Evaluate your Machine Learning Algorithm [WWW Document]. Medium. URL https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 (accessed 4.12.23).
O'Shea, K., Nash, R., 2015. An Introduction to Convolutional Neural Networks. ArXiv e-prints.
Overby, S., n.d. How to prioritize Artificial Intelligence (AI) projects: 6 tips [WWW Document]. The Enterprisers Project. URL https://enterprisersproject.com/article/2021/8/how-prioritize-artificial-intelligence-ai-projects-6-tips (accessed 4.14.23).
Paleyes, A., Urma, R.-G., Lawrence, N.D., 2022. Challenges in Deploying Machine Learning: A Survey of Case Studies. ACM Comput. Surv. 55, 114:1-114:29. https://doi.org/10.1145/3533378
Pan, S., Iplikci, S., Warwick, K., Aziz, T.Z., 2012. Parkinson's Disease tremor classification – A comparison between Support Vector Machines and neural networks. Expert Systems with Applications 39, 10764–10771. https://doi.org/10.1016/j.eswa.2012.02.189
Parkinson's disease [WWW Document], 2017. . nhs.uk. URL https://www.nhs.uk/conditions/parkinsons-disease/ (accessed 4.3.23).
Parkinson's symptoms [WWW Document], n.d. . Parkinson's UK. URL https://www.parkinsons.org.uk/information-and-support/parkinsons-symptoms (accessed 4.3.23).
Rana, A., Dumka, A., Singh, R., Panda, M.K., Priyadarshi, N., Twala, B., 2022. Imperative Role of Machine Learning Algorithm for Detection of Parkinson's Disease: Review, Challenges and Recommendations. Diagnostics 12. https://doi.org/10.3390/diagnostics12082003
S. Shetty, Y. S. Rao, 2016. SVM based machine learning approach to identify Parkinson's disease using gait analysis, in: 2016 International Conference on Inventive Computation Technologies (ICICT). Presented at the 2016 International Conference on Inventive Computation Technologies (ICICT), pp. 1–5. https://doi.org/10.1109/INVENTIVE.2016.7824836
Shin, T., 2022. All Machine Learning Models Explained in 6 Minutes [WWW Document]. Medium. URL https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a (accessed 4.11.23).
Singh, G., Vadera, M., Samavedham, L., Lim, E.C.-H., 2016. Machine Learning-Based Framework for Multi-Class Diagnosis of Neurodegenerative Diseases: A Study on Parkinson's Disease. IFAC-PapersOnLine 49, 990–995. https://doi.org/10.1016/j.ifacol.2016.07.331
Tolosa, E., Garrido, A., Scholz, S.W., Poewe, W., 2021. Challenges in the diagnosis of Parkinson's disease. The Lancet Neurology 20, 385–397. https://doi.org/10.1016/S1474-4422(21)00030-2
Varalakshmi, P., Tharani Priya, B., Anu Rithiga, B., Bhuvaneaswari, R., Sakthi Jaya Sundar, R., 2022. Diagnosis of Parkinson's disease from hand drawing utilizing hybrid

models. Parkinsonism & Related Disorders 105, 24–31. https://doi.org/10.1016/j.parkreldis.2022.10.020

Wenzel, M., Milletari, F., Krüger, J., Lange, C., Schenk, M., Apostolova, I., Klutmann, S., Ehrenburg, M., Buchert, R., 2019. Automatic classification of dopamine transporter SPECT: deep convolutional neural networks can be trained to be robust with respect to variable image characteristics. European Journal of Nuclear Medicine and Molecular Imaging 46, 2800–2811. https://doi.org/10.1007/s00259-019-04502-5

What are Convolutional Neural Networks? | IBM [WWW Document], n.d. URL https://www.ibm.com/se-en/topics/convolutional-neural-networks (accessed 4.11.23).

What are Neural Networks? | IBM [WWW Document], n.d. URL https://www.ibm.com/se-en/topics/neural-networks (accessed 4.11.23).

What is Parkinson's? | Parkinson's Europe [WWW Document], n.d. URL https://www.parkinsonseurope.org/about-parkinsons/what-is-parkinsons/ (accessed 4.3.23).

Xing, X., Luo, N., Li, S., Zhou, L., Song, C., Liu, J., 2022. Identification and Classification of Parkinsonian and Essential Tremors for Diagnosis Using Machine Learning Algorithms. Frontiers in Neuroscience 16.

Yiu, T., 2021. Understanding Random Forest [WWW Document]. Medium. URL https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (accessed 4.18.23).

Yohannes, Y., Webb, P., 1999. Classification and regression trees, CART: a user manual for identifying indicators of vulnerability to famine and chronic food insecurity. Intl Food Policy Res Inst.