

CHARACTERISTICS OF WAVEFORM FOR DETECTION OF PARKINSON'S DISEASE USING EEG WAVES

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Electronics and
Communication Engineering

by

K. SAI CHARAN (37130185)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING**

SCHOOL OF ELECTRICAL AND ELECTRONICS

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade 'A' by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI

SALAI,

CHENNAI - 600 119

MARCH 2021



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

**DEPARTMENT OF ELECTRONICS AND
COMMUNICATION ENGINEERING**

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of K. SAI CHARAN (37130185) who carried out the project entitled '**CHARACTERISTICS OF WAVEFORM FOR DETECTION OF PARKINSON'S DISEASE USING EEG WAVES**' under our supervision from September 2020 to March 2021.

Internal Guide

Dr. P. GRACE KANMANI PRINCE M.E.,Ph.D.

Associate Professor, Department of ECE

HEAD OF THE DEPARTMENT

Dr. T. RAVI

Submitted for Viva Voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **K.SAI CHARAN (37130185)** hereby declare that the Project Report entitled '**CHARACTERISTICS OF WAVEFORM FOR DETECTION OF PARKINSON'S DISEASE USING EEG WAVES**' done by me under the guidance of **Dr. P. GRACE KANMANI PRINCE**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

I are pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.N.M.NANDHITHA, M.E., Ph.D. Dean,School of Electrical and Electronics Engineering** and **Dr. T. RAVI, M.E., Ph.D. Head of the Department, Department of Electronics and Communication Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to our Project Guide **Dr. P. GRACE KANMANI PRINCE , M.E., Ph.D. Associate Professor, Department of Electronics and Communication Engineering** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of the project work.

I wish to express my thanks to all teaching and Non-teaching staff members of the **Department of Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

I express my gratitude to my parents for their constant encouragement and support for the completion of the project.

ABSTRACT

The motive of this work was to explore the automated diagnosis of Parkinson's disease (PD) using electroencephalography (EEG). Diagnosis of Parkinson's disease (PD) in the early stages is very critical for effective treatments. In this project, we propose a simple and low cost biomarker to diagnose PD, using the electroencephalography (EEG) signals. In the proposed method, EEG is used to detect the brain electrical activities in internal regions of brain, e.g., basal ganglia (BG). Neural activity in the β frequency range (13–30 Hz) is excessively synchronized in Parkinson's disease (PD). We show that these features of β waveform shape also distinguish PD patients on and off medication using non-invasive recordings in a dataset of a few PD patients with resting scalp EEG. Based on the high correlation between PD and brain activities in the brain ganglia, the proposed project, it provides a highly accurate PD diagnostic measure. Moreover, a quantitative measure of disease severity, using the spectral analysis of extracted brain sources is obtained.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF ABBREVIATIONS	x
	LIST OF FIGURES	viii
	LIST OF TABLES	ix
1	INTRODUCTION	1
	1.1 OUTLINE OF THE PROJECT	1
	1.2 PARKINSON'S DISEASE	1
	1.3 ELECTROENCEPHALOGRAPHY	3
	1.4 ARTIFICIAL INTELLIGENCE	4
2	LITERATURE REVIEW	4
3	AIM AND SCOPE	11
	3.1 AIM OF THE PROJECT	12
	3.2 SCOPE OF THE PROJECT	12
4	MATERIALS AND METHODS	12
	4.1 SIGNAL PROCESSING	13
	4.2 PRE-PROCESSING	15
	4.2.1 PRINCIPLE COMPONENT COEFFICIENT	15
	4.2.2 XLSREAD	16
	4.2.3 MACHINE LEARNING	16

	4.2.4 CLASSIFICATION LEARNER	17
	4.2.5 CONFUSION MATRIX	20
5	RESULTS AND DISCUSSION	23
	5.1 RESULTS	23
	5.1.1 CLASSIFICATION LEARNER OUTPUT	23
	5.1.2 CONFUSION MATRIX	26
	5.2 DISCUSSION	32
6	CONCLUSION	33
	6.1 CONCLUSION	33
	REFERENCES	34

LIST OF FIGURES

FIG NO	TITLE	PAGE NO
4.1	BINARY CLASSIFICATION METRIC	21
5.1	SCATTER PLOT OF ORIGINAL SET	24
5.2	SCATTER PLOT OF PREDICTED	25
5.3	CONFUSION MATRIX	26
5.4	TRUE POSITIVE RATE AND FALSE NEGATIVE RATE	27
5.5	POSITIVE PREDICTIVE AND FALSE DISCOVERY	28

LIST OF TABLES

TABLE	TITLE	PAGE NO
5.1	CLASSIFICATION LEARNER OUTPUT	23
5.2	BINARY CLASSIFICATION METRIC - YOUNG	29
5.3	BINARY CLASSIFICATION METRIC - OLD	30
5.4	BINARY CLASSIFICATION METRIC - PD	31

LIST OF ABBREVIATIONS

PD	PARKINSON'S DISEASE
EEG	ELECTROENCEPHALOGRAPHY
AI	ARTIFICIAL INTELLIGENCE
SVM	SUPPORT VECTOR MACHINE
KNN	K-NEAREST NEIGHBOURS
DFA	DISCRIMINANT FEATURE ANALYSIS
DSP	DIGITAL SIGNAL PROCESSING
ELM	EXTREME LEARNING MACHINE
MuPAD	MULTIPROCESSING ALGEBRA DATA
PCA	PRINCIPAL COMPONENT ANALYSIS
PPQ	PHOTO PLETHYSMOGRAPHY
PVM	PARALLEL VIRTUAL MACHINE

CHAPTER-1

INTRODUCTION

OUTLINE OF THE PROJECT

In the brain, the maximum number of the neurons is reached at birth, and, unlike other cells in the human body, these neurons cannot be repaired. Thus, over the time, neurons die and fail to be replaced. Usually, Parkinson's disease (PD) arises from the death of these nerve cells. The nerve cells generate dopamine, which is a chemical substance that mainly controls the body's motion. Therefore, the quantity of the generated dopamine decreases when nerve cells die, which begins to affect the multiple communication modes of the brain. This disease appears mostly in people 50 years or older. Unstable posture, muscle stiffness, slow movement, tremor, loss of balance, and reduced fine motor skills are some initial signs of PD. Hence, intelligent detection methods can be valuable for the early diagnosis of abnormal signs. As automated diagnosis systems, these methods are able to objectively detect PD by electroencephalogram (EEG) signals. Using the EEG signals, functions of cortical (or sub-cortical) segments in the brain can be easily detected. Other disease brain-related diseases, such as Alzheimer's and epilepsy, can be identified by these signals. We thus aimed to use these EEG signals to configure a computer-aided system capable of diagnosing PD. Collection of statistical dataset is taken and manipulations are done on them using Classification Learner tool in Matlab and infer whether the person is suffering from Parkinson's Disease with the help of Machine Learning Algorithms accuracy which will be explained later in this report.

PARKINSON'S DISEASE

Parkinson's disease is a condition whose main features are

- Slowed movement
- tremor
- gait or balance problems

More than 1 million people in the United States have Parkinson's disease. Although it more commonly develops in people in their 60s or older, it can occur as early as age 20. Parkinson's disease is caused by a loss of nerve cells in a specific part of the brain called the substantia nigra. Cells in the substantia nigra communicate with other movement control centers in the brain by secreting dopamine and other neurotransmitters. When substantia nigra cells die, they stop secreting dopamine, and the other movement control centers become unregulated. This disturbance in the movement control centers of the brain cause the main symptoms of Parkinson's disease (slowness, tremor and gait or balance problems). In most patients, what causes the loss of substantia nigra cells remains unknown. Environmental toxins that may cause Parkinson's include Manganese, carbon monoxide, organic solvents, certain pesticides. However, most people with Parkinson's disease have not been exposed to these toxins. A number of genes have been shown to cause Parkinson's disease in about 10% of patients, but these gene abnormalities do not occur in most individuals with Parkinson's disease. Currently, researchers suspect that the cause of Parkinson's disease in most individuals reflects a combination of genetic factors and environmental exposure(s). The first symptom and disease course varies considerably in people with Parkinson's disease. However, several general statements can be made. The most common initial symptom is tremor. Symptoms of tremor or slowness usually begin on one side of the body, often involving either an arm or leg. Over time (in some people, 5 years or more) the other limb on the same side, and then limbs on the opposite side are affected. Mobility (arising from a chair or walking) may be limited as the disease progresses. For most individuals on treatment, lifespan is not shortened by Parkinson's

disease. In addition to the well known symptoms of tremor, slowness, and imbalance, other problems such as excess saliva, a soft voice, as well as hand or foot cramps may occur. Loss of the sense of smell is common. Depression, anxiety and a number of sleep disorders are quite common. Constipation, more frequent urination, incontinence, increased sweating, and impotence may develop. About 40% of people with Parkinson's disease will develop changes in memory and intellectual function.

ELECTROENCEPHALOGRAPHY(EEG)

The **electroencephalogram (EEG)** is a recording of the electrical activity of the brain from the scalp. The first recordings were made by Hans Berger in 1929 although similar studies had been carried out in animals as early as 1870. The waveforms recorded are thought to reflect the activity of the surface of the brain, the cortex. This activity is influenced by the electrical activity from the brain structures underneath the cortex. The nerve cells in the brain produce signals that are called action potentials. These action potentials move from one cell to another across a gap called the synapse. Special chemicals called neurotransmitters help the signals to move across the gap. There are two types of neurotransmitters, one will help the action potential to move to the next cell, the other will stop it moving to another nerve cell. The brain normally works hard to keep an equal amount of each of these neurotransmitters in the brain. EEG activity is quite small, measured in microvolts (μV) with the main frequencies of interest up to approximately 30 Hertz (Hz). Small metal discs called electrodes are placed on the scalp in special positions. These positions are identified by the recordist who measures the head using the International 10/20 System. This relies on taking measurements between certain fixed points on the head. The electrodes are then placed at points that are 10% and 20% of these distances. There are a great variety of electrodes that can be used. The majority are small discs of stainless steel, tin, gold or

silver covered with a silver chloride coating. These normally have a lead attached. Alternative methods consist of a cap in which the electrodes are already imbedded.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence is the ability to design smart machines or to develop self-learning software applications that imitate the traits of the human mind like reasoning, problem-solving, planning, optimal decision making, sensory perceptions etc. According to John McCarthy, the father of AI, 'The science and engineering of making intelligent machines, especially intelligent computer programs' is the definition of Artificial Intelligence. AI, as the name suggests, is imparting intelligence to the machines so that the machines operate like human beings.

AI is that sector in computer science that emphasizes the creation of intelligent machines that work, operate and react like human beings. AI is used in decision making by the machines considering the real-time scenario. An Artificially Intelligent machine reads the real-time data, understands the business scenario and reacts accordingly. AI has now become a very important part of Information Technology. This branch aims to create machines that are intelligent. The process of transforming a computer into a computer-controlled robot or designing software that thinks and reacts exactly the way a human being thinks is what AI is all about.

CHAPTER-2

LITERATURE REVIEW

Survey carried out for the diagnosis of PD with different algorithms and approaches:

Several strategies are recorded for early detection of PD based on the different ML techniques. But accuracy in detection and classifying within the time is very important or else, it causes development of more symptoms. There are different kinds of data, brain MRI images, Voice data, posture images, sensor captured data, handwritten data, using which we can predict whether person is having PD or not.

Max A. Little et al presented a new dysphonia measure, pitch period entropy (PPE) and used a kernel support vector machine and has achieved classification accuracy of 91%[1].

Rainer Scho Nweiller et al identified a different approach which used voice analysis with ANN and got good results but observed that cost-effectiveness remains to be a challenge[2].

Marius Ene et al suggested NN based approach with three types of internal methods and discriminated persons having PD with healthy persons[3].

David Gil A, Magnus Johnson B found that with a smaller number of neurons at hidden layer both training set and test sets performed poorly. With higher number of neurons, the training set performed well with high risk of over fitting. The ideal solution for this layer was found to be 13 neurons[4].

Ipsita Bhattacharya et al identified the ROC curve variation and identified that values of TP and FP rates show changes while increase in the CV folds[5].

Freddie Åström et al proposed unique approach of parallel neural networks and then outcome of each neural network is assessed by using a rule-based system for the decision. During the training process, data that is not yet learned of each neural network is collected and applied in the training set of the later neural network. This helped to increase prediction accuracy[6].

AthanasiosTsanas et al developed a novel algorithm based on speech signals but is questionable as most of the features are not considered here, only 10 features are used[7].

Hui-Ling Chen et al proposed FKNN centered system using a 10-fold cross validation method[8].

Mohammad S Islam et al has compared various ML techniques based on their performance accuracies in determining whether person is having PD or not and mentioned that new classifier may be built to get better accuracies[9].

Bo Penga et al suggested Computer Aided Analysis with image data and used BrainLab software for processing the images and calculate thickness of the cortex, volume of gray matter, and surface area of the cortex on each region of interest (ROI). Use of Multilevel ROI-based features improved the classification performance[10].

DeriyaAvci and AkifDogantekin proposed another approach using Genetic Algorithm-Wavelet Kernel-Extreme Learning and achieved good accuracy results[11].

SatyabrataAich proposed a unique approach by using Genetic algorithm and PCA as feature selection methods and applied seven ML algorithms for classification, that saved time and productivity while doing pattern classification with two categories such as PD and not PD[12].

Leandro A. Passos compared ResNet- 50 , Optimum-Path Forest (OPF) classifier with Support Vector Machines (SVM) and Bayes and achieved 96% of identification rate[13].

Deepak Gupta followed a different approach cuttlefish algorithm and used for feature selection, different fitness functions approximations are used to improve cuttlefish algorithm and is termed as Optimized cuttlefish algorithm (OCFA).Decision tree and K-Nearest Neighbor classifiers are applied and achieved 94% of accuracy in detecting PD effected patients[14].

Salama A. Mostafa proposed Multiple Feature Evaluation Approach (MFEA) of a multi-agent system (ii) Implementation of five classification schemas which are Decision Tree, Random Forests, Neural Network, Naïve Bayes and Support Vector Machine on the Parkinson"s diagnosis before and after applying their approach, and (iii)Author approach witnessed the following average rate of accuracies : Decision Tree achieved accuracy of 10.51%, Naïve Bayes shown 15.22%, Neural Network is found with 9.19%, Random Forests and SVM performed with 12.75% and 9.13%respectively[15].

It is important to note that , out of all ML techniques, ANN and SVM classifier are used most of the proposed algorithms to aid faster and accurate the prediction.

DIAGNOSIS OF PARKINSON'S DISEASE USING EEG

Using EEG Signal the most common disabling disturbances of PD with respect to the physical activity are the Freezing of Gait (FOG), Turning Freeze (TF), and Gait Initiation Failure (GIF). The PD affected patient's experiences difficulty in walking as they feel that their feet have been glued to the floor. During this stage the EEG signals can be recorded from the patients given them a TUG tasks. The cognitive process (brain mechanism) can be analyzed for the timing and scaling of movement using EEG signal analysis and could be a useful tool in the detection of above- mentioned disturbances. The below section discusses the related work carried in the analysis of PD using EEG signals. AM ArdiHandojoseno, et.al discusses the method of detection of Freezing of Gait in PD patients using EEG signal based on wavelet decomposition. In this paper the dynamics of FOG was analyzed by extracting the EEG Entropy and Sub-band wavelet energy. 26 Patients around 60-70 years with PD and significant FOG were considered. The author highlights that by using BPNN (Back Propagation Neural Network) classifier the onset of Freezing of PD can be identified with 75% accuracy, specificity and sensitivity using sub bands of EEG.

Also, the statistical analysis indicates that there will be a difference in the Energy sub bands of delta, theta and alpha with Normal and PD affected group. Finally gives a future scope for the usage of different classifiers for still better results . Jonathan Weihenmayer[16] proposes the method of diagnosing of PD patients with a Healthy control by extracting the Raw and Clean EEG data of patients during the quasi resting state for

1 sec of Nine PD patients and ten healthy individuals. Nonlinear delay differential equations (DDE) are applied to classify the PD patients from and Healthy individual.

The classification is done upon the plotting of area under the receiver operating characteristic curve (AROC). It is shown that the raw EEG data (AROC = 0.59-0.86) can classify the PD patients with a Healthy individual more robustly than clean EEG data (0.57-0.72). The author states that increased classification of raw EEG time series originates from muscle artifacts and concludes by stating that a non-linear extraction and classification of raw EEG data can be used as biomarker for PD. This paper can be extended for a further scope for research by considering the patients in true resting states and for a larger data sets and sample size.

Quynh Tran Ly et al.[17] discuss the brain mechanism during the Gait Initiation Failure (GIF), which is a subtype of Freezing of Gait (FOG). EEG signals of 61 samples were recorded from Four PD affected patients (3 Male and 1 Female) to study the brain dynamic changes during the occurrence of GIF. EEG samples were analyzed through the Centroid Frequency (CF) and Power Spectral Density (PSD). The results showed that GIF were associated with significant increase in high beta band (21-38 Hz) in the frontal, central, parietal, occipital regions of brain. Using BPNN classifier the analysis of PSD and CF in EEG provided 84.27 % of performance and 84.80% of sensitivity and accuracy respectively. Concludes by stating that High beta activity was found in EEG with PD affected patients during the analysis of GIF and suggested to use a larger sample size and other multiple electrodes. The identification of combination of sensors and optimal location are also discussed. EEG measures/parameters, in the form of Power Spectral entropy, Power Spectral density and Centroid Frequency for 343seconds from 7 patients(6Mand1F)with, Effective walking and PD affected with FOG during a series of TUG were extracted to analyze the important features that were significantly changed during FOG. Feed- Forward Neural networks were applied to classify the EEG signals amongst the Effective walking and Freezing. The sensitivity and accuracy from the classification using single channel, two channels, three, Four, Six, Seven and Thirty-two channels were discussed. The performance of C4O2 in a two channel and four locations at F3C4P4O2 provided the best indicator of FOG detection with 72.54% and 72.29% of

sensitivity and 69.71% and 70.29% of accuracy respectively. The outcome of the paper is that the best montage is using EEG system with two channels C4-O2. The author concludes by stating that the increase in the number of channels will not increase the accuracy and suggests that the further scope of research can be done by considering larger sample of data and to use of any specific Neural Network which can lead to a better result. A method for the suppression of Artifacts that occur in EEG signals due to Neuro stimulation.

The Neurostimulation is assessed by Advanced Statistical Signal Processing and Machine Learning Methodologies. The unsupervised clustering method is applied to neural recordings in order to suppress Neurostimulation. The author also discusses about the Deep Brain Simulation (D B S) method for the treatment of PD, where electrical simulator is placed into the body and it delivers the electrical current to thalamus in order to reduce the severity of PD. This Simulator can send the EEG signals which indicates the treatment and possible DBS side effects. Time Frequency features in EEG are extracted in patients undergoing DBS treatment of PD using matching pursuit decomposition (MPD). This paper concludes by stating that the behavioral task clustering for PD patients with DBS using Time frequency features gives the differences in the band activity (delta, beta and gamma) of EEG. The authors Quynh Tran Lya et al. [13] briefs about technique of S-transform decomposition of EEG Signals in detection of Turning freeze in PD patients. Turning Freeze (TF) is the most common subtype of FOG and is causing the majority rate of falls in PD patients.

The brain signals were captured to using EEG signal to measure the feet getting stuck while making a turn (TF). Data was collected from six PD affected patients during TUG tasks. The main discussion in this paper is the extraction of EEG features using time frequency technique S-Transform decomposition (Stock well transform). Independent Component analysis by entropy bound minimization technique was used to separate EEG sources. Bayesian neural networks (BNN) classification algorithm was implemented to improve the classification results in detecting TF in PD patients. The

main findings in this paper are the inclination in the high beta activity in the region of cortex near occipital O2 and frontal F4 locations. The author concludes the paper by projecting 84.24% sensitivity, 88.0% of specificity and 86.2% accuracy in detecting the TF. The scope for the future work is to increase the accuracy. AM ArdiHandojoseno et al. proposes a methodology for the prediction of FOG in PD using EEG spatial correlation, wavelet coefficient and cross frequency energy techniques. Integration of spatial, spectral and temporal features of EEG signal with the use of Neural Network and K-NN classifiers is discussed.

CHAPTER-3

AIM AND SCOPE

AIM OF THE PROJECT

We propose detection of Parkinson's Disease using the EEG. To study the characteristics of the waveform obtained by EEG. Machine learning is used to verify the performance of the labeled input readings. Early diagnosis of Parkinson's Disease is necessary as by getting late it becomes more and more hazardous which may ultimately lead to death.

SCOPE OF THE PROJECT

As diagnosis is currently rather subjective, researchers have been trying to find an easier and more scientific method. An EEG records electrical activity produced by the brain via small sensors attached to the scalp.

Attempts to use EEG readings for Parkinson's disease diagnosis have not always produced the results that researchers have been looking for. Different architectures of Artificial Neural Network (ANN) such as Convolutional Neural Network (CNN), Long-short term Memory, and Extreme Learning Machine are well-known deep learning algorithms for future growth (ELM). Finding that a noninvasive method such as an EEG could be a promising diagnostic tool may have important consequences for the future of the disease. One can use the test to track changes related to Parkinson's in the brain over the years.

The Classification Algorithms from Machine Learning and Deep Learning are used to predict and investigate the Parkinson's Disease. The ideal features from the dataset are passed as input to the models and the prediction results are obtained.

CHAPTER- 4

METHODS AND MATERIALS USED

Parkinson's disease (PD) is one of the most common neurodegeneration disorders and it is believed that it affects 0.3% of the global population. PD normally occurs in people aged from 50 to 60 and is not common before 50's. When it appears before age 21, genetic factors are presumed to be the cause. The origin of PD remains unknown until today, although it is believed that the interaction between environmental and genetic factors may explain its appearance. The lack of a cure for PD makes it important to research a method for early stages detection in order to delay the disease progression and its symptoms, thereby improving the patient's quality of life. Patients diagnosed with PD are known by the rest tremor but also present other symptoms, such as: muscular rigidity, bradykinesia or postural instability. Moreover, they present an accumulation of α -synuclein, which leads to the formation of Lewy bodies and a progressive loss of dopaminergic neurons in the substantianigra . In order to describe how the symptoms of PD progress, the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Hoehn and Yahr scale are commonly used.

SIGNAL PROCESSING

Signal Processing Toolbox includes functions and applications for analyzing, pre-processing, and extracting features from signals that are sampled uniformly or non-uniformly. Filter configuration and analysis, resampling, smoothing,

detrending, and power spectrum estimation are all included in the toolbox. Extracting features such as change points and envelopes, detecting peaks and signal patterns, quantifying signal similarities, and performing measurements such as SNR and distortion are all possible with the toolbox. Vibration signals may also be subjected to modal and order review.

Without writing code, Signal Analyzer allows you to pre-process and analyze several signals simultaneously in the time, frequency, and time-frequency domains; explore long signals, and extract regions of interest. Filter Designer allows you to build and analyze digital filters by selecting from a number of different algorithms and responses.

PREPROCESSING

PRINCIPAL COMPONENT COEFFICIENT

The principal component coefficients, also known as loadings, for the n -by- p data matrix X are returned by PCA (x). Observations are represented by rows of X , while variables are represented by columns. It's a p -by- p coefficient matrix. The coefficients for each principal component are mentioned in descending order of component variance in each column of coefficients. By default, PCA uses the Singular Value Decomposition (SVD) algorithm to center the data.

A quantitatively rigorous method for achieving this simplification is principal component analysis. The principal components method creates a new set of variables. The original variables are linearly combined components are orthogonal to one another. The principal components as a whole provide an orthogonal foundation for the data space. in each principal factor. There is no redundant knowledge since all of the principal

An orthogonal basis for multiple columns of data can be constructed in an infinite number of ways.

A single axis in space is the first major part. The values that result from projecting each observation on that axis form a new variable. And this variable's variance is the highest of all the options on the first axis.

The second main part is a perpendicular axis in space to the first. When you project the observations onto this axis, you get a new variable. This attribute has the highest variance out of all the options on the second axis.

The total number of principal components is the same as the number of variables in the original package. The number of the variances of the first few principal components, on the other hand, is often greater than 80% of the total variance of the original results. Researchers also gain a better understanding of the driving forces that produced the original data by analyzing plots of these few new variables. The PCA function aids in the discovery of the primary components. The real calculated data had to be analyzed in order to use PCA.

XLSREAD

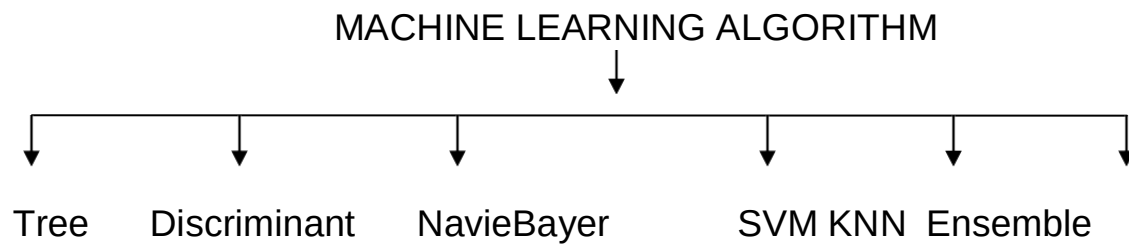
`A = xlsread('filename')` returns numeric data from the first sheet in the file name Microsoft Excel spreadsheet format in array A. Leading rows or columns of text are ignored by `xlsread`. If a cell in a leading row or column is empty or contains text, however, `xlsread` replaces it with a NaN.

MACHINE LEARNING

Machine learning teaches machines to learn from experience in the same way as humans do. Machine learning algorithms employ analytical methods to 'learn' knowledge directly from data rather than relying on a model based on a predetermined equation. As the number of samples available for learning grows, the algorithms adapt their output. Machine Learning uses two ways of technique

- Supervised learning is a method of training a model using known input and output data in order to predict potential outcomes.
- Unsupervised learning is the discovery of hidden patterns or intrinsic structures in data without the use of supervision.

In the face of uncertainty, the aim of supervised machine learning is to create a model that can make predictions based on data. A supervised learning algorithm uses a known collection of input data and known responses to the data (output) to train a model to make accurate predictions about how new data will react. To build predictive models, supervised learning employs classification and regression methods.



CLASSIFICATION LEARNER

Importing data, defining validation schemes, interactively exploring the data, selecting features, training models, and evaluating model efficiency are all possible with Classification Learner.

The Classification Learner app teaches data classification to models. It uses automatic training to find the best classification model sort, which can include decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive Bayes, and ensemble classification. Techniques for classifying data predict categorical responses, such as whether an email is real or spam, or whether a tumor is cancerous or benign. Input data is classified using classification models. Medical imaging, image and speech recognition, and credit scoring are examples of popular applications.

Train Classification Models with the Classification Learner App

The Classification Learner app allows you to use supervised machine learning to train models to classify data. Using different classifiers, the software allows you to interactively explore supervised machine learning.

- Automatically train a number of models and assist you in selecting the right one. Decision trees, discriminant analysis, support vector machines, naive Bayes, and ensemble classification are some of the different types of models.
- Investigate your data, define validation schemes, choose features, and visualize the results. Cross-validation is used by default to defend against over fitting. The validation results will assist you in selecting the best model for your data which are reflected in the plots and performance indicators.
- To make predictions with new data, export models to the workspace. The app often trains a full model as well as a model with the required validation scheme, and the full model is the model you export.
- Build scripts, train with new data, work with large data sets, or change the code for further analysis using MATLAB code generated from the app.

Decision Tree: Decision trees, also known as classification trees and regression trees, are used to forecast data responses. Follow the decisions in the tree from the root (beginning) node down to a leaf node to predict an answer. The answer is stored in the leaf node. Nominal answers, such as 'true' or 'false,' are given by classification trees.

Discriminant Analysis

DFA (Discriminant Feature Analysis) is a mathematical method for classifying unknown persons and determining their likelihood of being classified into a specific category (such as sex or ancestry group). The presumption in discriminant analysis is that the sample is usually distributed for the trait.

Where the parameter or dependent variable is categorical and the indicator or independent variable is the interval in nature, discriminant analysis is a method used by the researcher to interpret the test results.

Support Vector Machine: Support-Vector Machines (also known as support-vector networks) are supervised learning models with associated learning algorithms that analyze data for classification in machine learning. An SVM model is a representation of the examples as points in space, mapped such that the examples of the different categories are separated by a large distance. New examples are then mapped into the same space and assigned to one of the categories depending on which side of the distance they fall on.

Naive Bayer: The Naive Bayes classifier is a basic, but powerful, machine learning classifier. In a Bayesian context, it is a probabilistic classifier that makes classifications using the Maximum A Posteriori decision law. A very simple Bayesian network can also be used to describe it. A similar approach is used by Naive Bayes to estimate the likelihood of various classes based on various attributes. This algorithm is usually used for text classification and problems with multiple groups.

Ensemble Classification: Since ensemble can be trained and then used to make predictions, an ensemble is a supervised learning algorithm. It has been shown that qualified ensembles have greater versatility in the roles they can serve. In theory, this versatility allows them to over-fit the training data rather than a single model, although some ensemble techniques have been shown to minimize problems associated with over-fitting the training data.

Where there is a lot of variety among the models, ensembles seem to produce better results. Many ensemble methods contribute to model diversity by combining several models. More random algorithms (such as random decision trees) may generate a stronger ensemble than very deliberate algorithms (such as entropy-reducing decision trees), which may seem counterintuitive.

KNN Classification: Categorization KNN is a nearest-neighbor classification model in which the distance metric and the number of nearest neighbors can both be modified. You can use a Classification KNN classifier to compute substitution predictions because it stores training data. Alternatively, you can use the predicted approach to identify new findings using the model. The KNN binary (as a two-class system) provides a more reliable data classification, making it advantageous to use k as an odd number to avoid irregular data. In machine learning, the KNN procedure is used. It's an object that's been categorized using a mainstream set of its neighbors, with the determination to assign occurrence for the most common class of its k closest neighbors (k is a positive integer, classically small).

The distance metric used in the past was Euclidean distance; however, this is only suitable for infinite variables. KNN is a modern method for delivering all available cases based on a quantity of assessment (e.g., distance functions). The KNN protocol is similarly straightforward. It regulates the K-nearest neighbors by requiring a minimum distance between the interrogation instance and the training samples.

CONFUSION MATRIX

The rows correspond to the expected class (Output Class) and the columns to the true class on the confusion matrix map (Target Class). The diagonal cells refer to correctly identified observations. The off-diagonal cells refer to findings that were wrongly labeled. Each cell displays the number of observations as well as the percentage of the total number of observations.

The percentages of all the examples predicted to belong to each class that is correctly and incorrectly categorized are shown in the column on the far right of the map. The precision (or positive predictive value) and false discovery rate are two commonly used metrics. The percentages of all the examples belonging to each class that is correctly and incorrectly labeled are shown in the row at the bottom of the plot. The recall (or true positive rate) and false negative rate are two commonly used metrics.

Confusion Matrix for Binary Classification Metric

Confusion Matrix is a square matrix where C_{ij} represents the number of data instances that are known to be in group i (true label) and predicted to be in group j (predicted label).The predicted label for a machine learning algorithm is the predicted label, while the true label for a machine learning algorithm is the true label.

		PREDICTED LABEL	
		NEGATIVE	POSITIVE
TRUE LABEL	NEGATIVE	TRUE NEGATIVE	FALSE POSITIVE
	POSITIVE	FALSE NEGATIVE	TRUE POSITIVE

Fig: 4.1. Binary Classification Metric

When a statement is actually positive and classified as positive, the statement is known as TRUE POSITIVE (TP). When a statement is actually negative and classified as negative, the statement is known as TRUE NEGATIVE (TN). When a statement is actually negative and classified as positive, the statement is known as FALSE POSITIVE (FP). When a statement is actually positive and classified as negative, the statement is known as FALSE NEGATIVE (FN). The Confusion Matrix, which is required to calculate the machine learning algorithm's accuracy in classifying data into its labels.

Accuracy

The number of correctly identified data instances divided by the total number of data instances is known as accuracy. If the set of data's is not balanced, accuracy may

not be in good measures.

$$\text{Accuracy} = \text{TN} + \text{TP} / (\text{TN} + \text{FP} + \text{TP} + \text{FN}).$$

Precision

A successful classifier should have a precision of 1 (high). Only when the numerator and denominator are equal, i.e. $\text{TP} = \text{TP} + \text{FP}$, does precision become 1. This also implies that FP is zero. The value of the denominator becomes greater than the numerator as FP increases, and the precision value decreases (which we don't want).

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}).$$

Recall

A successful classifier should have a recall of 1 (high). Only when the numerator and denominator are equal, i.e. $\text{TP} = \text{TP} + \text{FN}$, does recall become 1. This also implies that FN is zero. The value of the denominator becomes greater than the numerator as FN increases, and the recall value decreases (which we don't want).

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}).$$

F1 Score

Only when precision and recall are both 1 does the F1 Score become 1. Only when both precision and recall are high does the F1 score rise. The F1 score is a better metric than accuracy since it is the harmonic mean of precision and recall.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}).$$

CHAPTER-5

RESULT AND DISCUSSION

RESULT

A dataset of different categories of people(young, old and people suffering from Parkinson's Disease) are collected. We used supervised machine learning to verify our reading efficiency percentage. In machine learning, the filtered signals are then passed on to a classification learner. The parameters are evaluated and the efficiency is verified in the Confusion matrix which produces a precise result.

CLASSIFICATION LEARNER OUTPUT

The Classification Learner is a program that teaches models how to classify data. It uses automatic training to find the best classification model sort, which can include decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive Bayes, and ensemble classification.

Table 5.1. Classification Learner Output

Fine Tree	77.2%
Medium Tree	78.3%
Linear Discriminant	75.5%
Quadratic Discriminant	72.5%
Linear SVM	72.8%
Cubic SVM	71.6%
Fine Guassian SVM	74.5%
Medium Guassian SVM	73.0%

The Classification Learner is a program that teaches models how to classify data. It aids in the exploration of supervised machine learning using various classifiers and data, as

well as the selection of features, specification of validation schemes, training of models, and evaluation of performance.

Scatter Plot of Original Datasets

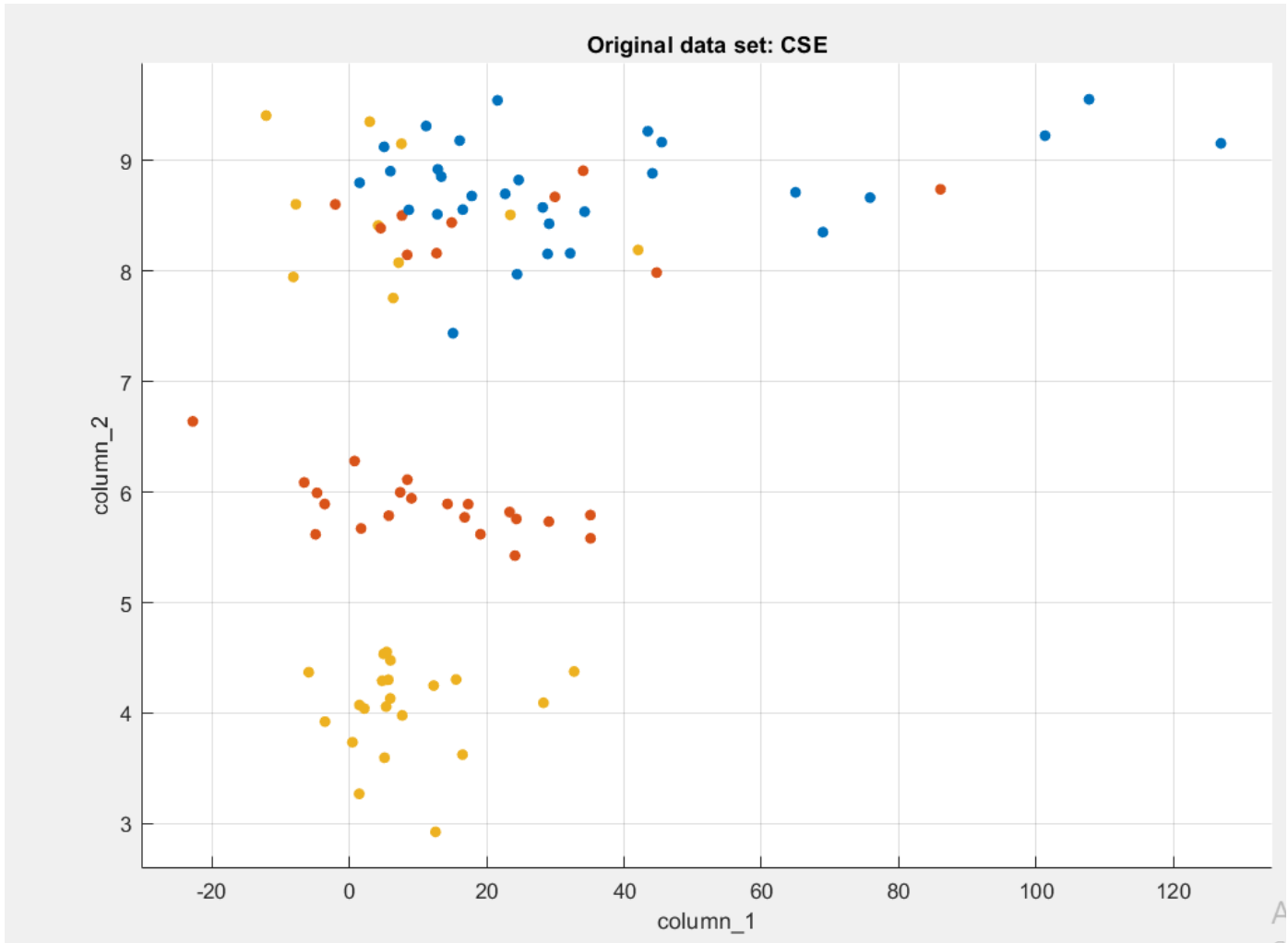


Fig 5.1 Scatter Plot of Original Datasets

In scatter plot the original set of data is obtained. The colors blue, red, yellow, and blue indicates categories like young, old, and people suffering from Parkinson's Disease.

Scatter Plot of Predicted Datasets

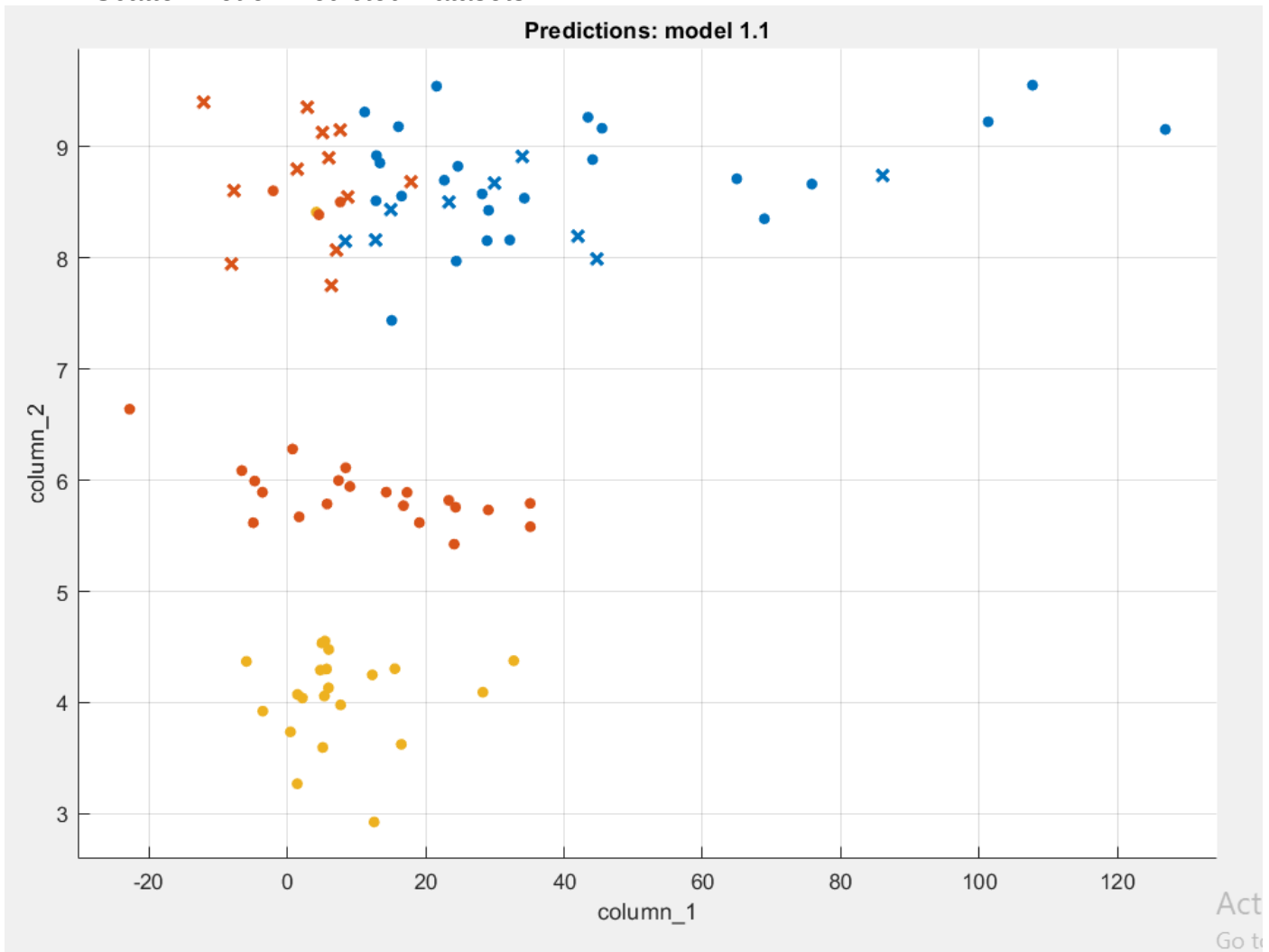


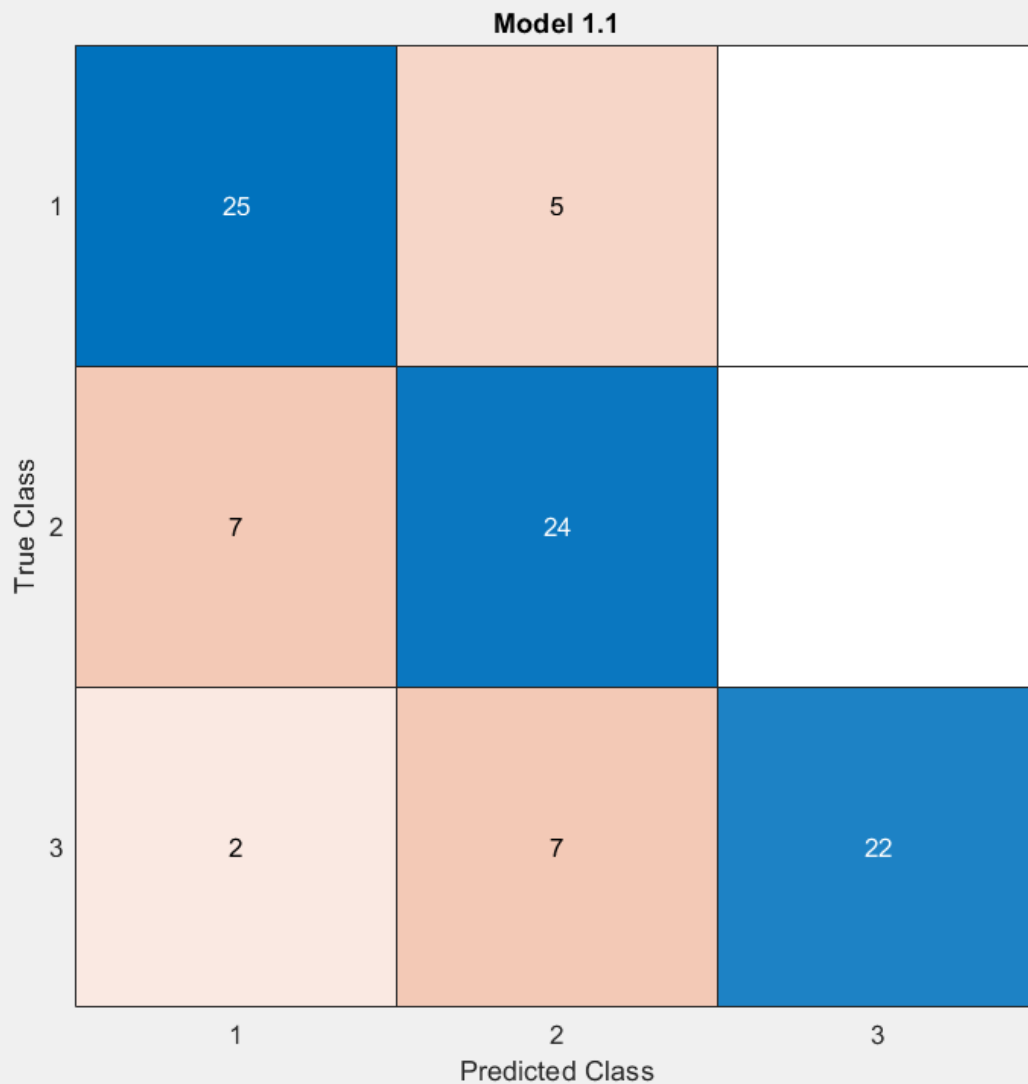
Fig 5.2 Scatter Plot of Predictions

In scatter plot the predicted set of data is obtained. All machine learning algorithms are performed and trained and the results are as shown in the figure. The colors blue, red, yellow, and blue indicates categories like young, old, and people suffering from Parkinson's Disease.

CONFUSION MATRIX:

The feature extraction process is seen in the final production. This process takes place in a matrix of uncertainty. K nearest classification is used in this method. KNN is a lazy learning algorithm that is non-parametric. Its aim is to predict the classification of a signal using a database with data points divided into several groups.

Fig 5.3 Confusion Matrix of Original Dataset



In Confusion Matrix for original set of data's under Number of Observation category the set of data is obtained. The diagonal elements in the Confusion Matrix must be high always as it suggests the true value and predicted value are same which in turn says that the accuracy is higher.

Confusion Matrix for True Positive and False Negative Rates

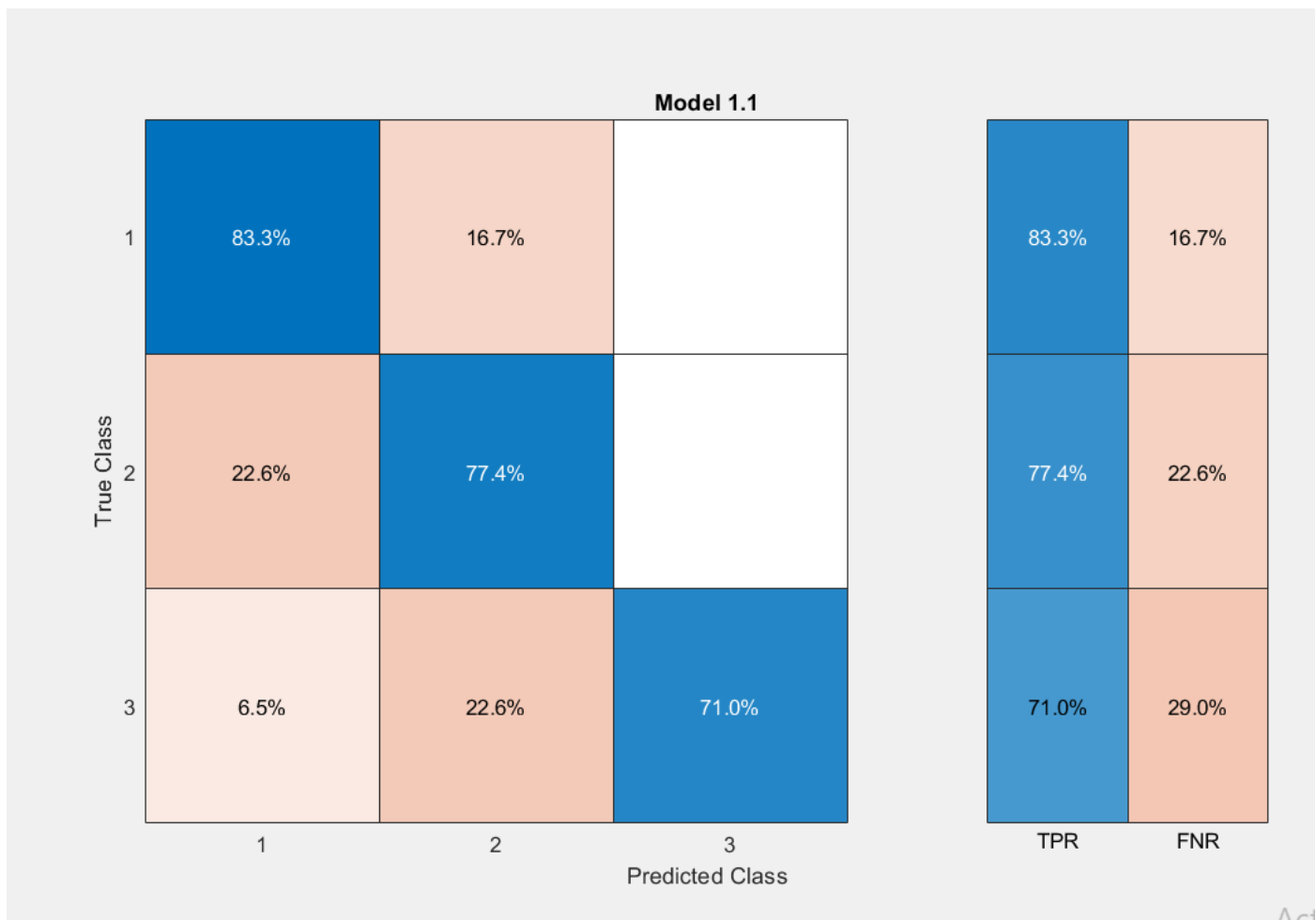


Fig: 5.4. True Positive Rate and False Negative Rate.

In Confusion Matrix for original set of data's under Positive Rate and False Negative Rate category the set of data is obtained. True Positive rates for emotion young is 83.3 %, old is 77.4 %, PD is 71 % and False Negative rates for emotion young is 16.7%, old is 22.6%, PD is 29.0 %.

Confusion Matrix for Positive Predictive and False Discovery

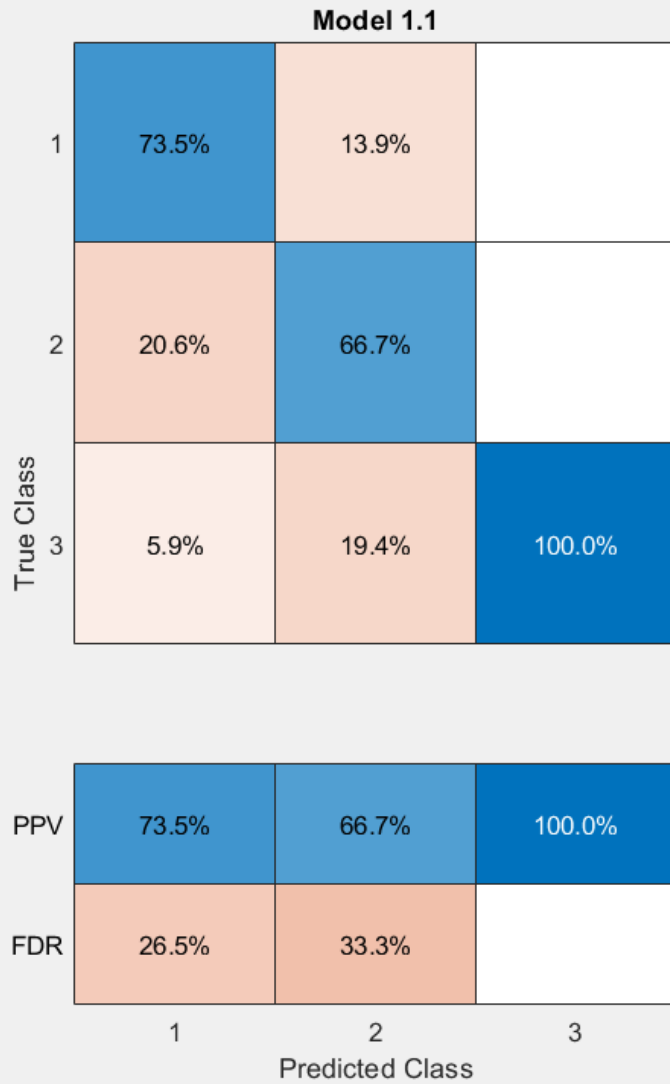


Fig: 5.5. Positive Predictive and False Discovery

In Confusion Matrix for original set of data's under Positive Predictive rate and False Discovery Rate category the set of data is obtained.

Table 5.2. Binary Classification Metric for Young

TRUE NEGATIVE 0.74	FALSE POSITIVE 0.265
FALSE NEGATIVE 0.17	TRUE POSITIVE 0.84

$$\begin{aligned}
 \text{ACCURACY: } &= \text{TN} + \text{TP} / (\text{TN} + \text{FP} + \text{TP} + \text{FN}) \\
 &= 0.74 + 0.83 / (0.74 + 0.83 + 0.17 + 0.26) \\
 &= 0.785 \\
 &= 78.5\%
 \end{aligned}$$

$$\begin{aligned}
 \text{PRECISION: } &= \text{TP} / (\text{TP} + \text{FP}) \\
 &= 0.83 / 0.83 + 0.26 \\
 &= 0.761
 \end{aligned}$$

$$\begin{aligned}
 \text{RECALL: } &= \text{TP} / (\text{TP} + \text{FN}) \\
 &= 0.83 / 1 \\
 &= 0.83
 \end{aligned}$$

$$\begin{aligned}
 \text{F1 SCORE: } &= 2 * ((\text{Precision} * \text{recall}) / (\text{Precision} + \text{Recall})) \\
 &= 2 * (1.26 / 1.59) \\
 &= 0.793
 \end{aligned}$$

Table 5.3. Binary Classification Metric for Old

TRUE NEGATIVE 0.66	FALSE POSITIVE 0.34
FALSE NEGATIVE 0.23	TRUE POSITIVE 0.77

ACCURACY: $= \text{TN} + \text{TP} / (\text{TN} + \text{FP} + \text{TP} + \text{FN})$
 $= 0.66 + 0.77 / (0.66 + 0.34 + 0.23 + 0.77)$
 $= 0.715$
 $= 71.5\%$

PRECISION: $= \text{TP} / (\text{TP} + \text{FP})$
 $= 0.77 / 1.1$
 $= 0.7$

RECALL: $= \text{TP} / (\text{TP} + \text{FN})$
 $= 0.77 / 1$
 $= 0.77$

F1 SCORE: $= 2 * ((\text{Precision} * \text{recall}) / (\text{Precision} + \text{Recall}))$
 $= 2 * (0.54 / 1.47)$
 $= 0.73$

Table 5.4. Binary Classification Metric for PD

TRUE NEGATIVE 1	FALSE POSITIVE 0
FALSE NEGATIVE 0.29	TRUE POSITIVE 0.71

ACCURACY: = $TN+TP / (TN+FP+TP+FN)$
= $1+0.71 / 2$
=0.85
=85%

PRECISION: = $TP / (TP+FP)$
= $0.71 / 0.71$
=1

RECALL: = $TP / (TP+FN)$
= $0.71 / 1$
=0.71

F1 SCORE: = $2 * ((Precision*recall) / (Precision + Recall))$
= $2 * (0.71 / 1.71)$
=0.83

DISCUSSION

Machine Learning techniques has got prominent role as they are applied in variety of domains especially in the healthcare. Unlike traditional methods, the models generated by applying ML techniques show dynamic outputs as data is fed into it. One shall make note that significant and narrow research is needed to obtain knowledge in diagnosing the disease. Various machine learning algorithms and techniques are being proposed rapidly, out of which some are observed to be promising with the results and few demonstrated their usage in different fields. Advantage with the ML generated models is that when more data is used, the precision values gets increased and the much accuracy in predictions can be gained. In the past several years, several non-invasive techniques have been proposed to identify PD using voice and gait signals. Different automated techniques were explored to develop the best automated model to differentiate normal and PD subjects.

It can be noted that various machine learning techniques were adopted to discriminate the EEG signals in normal and PD subjects. The obtained spectral analysis dataset is trained with Machine Learning Algorithms and the data is manipulated. From the results obtained, the scatter plot of original dataset and the scatter plot of predictions along with the confusion matrix is shown and from the results, the parameters like accuracy, precision, recall and f1 score have been calculated.

CHAPTER-6

CONCLUSION

6.1 CONCLUSION

This project is an effort to present broad review about Parkinson disease diagnosis system that have applied various machine learning techniques. The summary of results obtained by different researchers is made available in literature survey table , almost all the authors/researchers made great efforts to predict the Parkinson disease with novel approaches. It can be identified that maximum of all ML techniques used by various authors worked better. We trained the original datasets using Machine Learning algorithms that has a fair recognition rate across all three categories using multiple classification techniques and signal processing of sample readings using denoising filters, supervised machine learning, classification learner, and confusion matrix. The results have produced a good accuracy for PD patients of around 85%. Hence the proposed method can be efficient enough to diagnose PD. The proposed approach was tested using machine learning and deep learning functions. Preprocessing is a technique for reducing noise in sample reading input by using statistics coefficient parameters. Computer learning has been sent the filtered info.

REFERENCES

- [1] Max A. Little et al, Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING,VOL. 56, NO. 4, APRIL 2009.
- [2] Rainer Scho¨ Nweiler et al, Novel Approach to Acoustical Voice Analysis Using Artificial Neural Networks ,JARO 01: 270–282 (2000) DOI:10.1007/s101620010020.
- [3] Marius Ene et al, Neural network-based approach to discriminate healthy people from those with Parkinson's disease, Annals of the University of Craiova, Math.Comp. Sci. Ser. Volume 35, 2008, Pages 112–116 ISSN: 1223-6934.
- [4] David Gil A, Magnus Johnson B, Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines, Global Journal of Computer Science and Technology,2009.
- [5] Ipsita Bhattacharya et al, SVM Classification to Distinguish Parkinson Disease Patients, Conference: Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 16-17, 2010, Tamilnadu, India.
- [6] Freddie Åström et al, A parallel neural network approach to prediction of Parkinson'sDisease,2011Elsevier,doi:10.1016/j.eswa.2011.04.028.
- [7] AthanasiosTsanas et al, Novel Speech Signal Processing Algorithms forHigh-Accuracy Classification of Parkinson's Disease, IEEE Transactions On Biomedical Engineering, Vol. 59, No. 5, May2012.
- [8] Hui-Ling Chen et al, An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach, Expert Systems with Applications 40 (2013)263–271.
- [9] Mohammad S Islam et al, Performance Comparison of Heterogeneous Classifiers for Detection of Parkinson's Disease Using Voice Disorder (Dysphonia), 3rd International Conference On Informatics, Electronics & Vision2014.
- [10] Bo Penga et al, Computer Aided Analysis of Cognitive Disorder in Patients with Parkinsonism using Machine Learning Method with Multilevel ROI-based Features, J Med Syst (2015) 39:179 DOI10.1007/s10916-015-0353-9.

- [11] DeryaAvci and AkifDogantekin et al, An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel Extreme Learning Machine, Hindawi Publishing Corporation Parkinson's Disease Volume2016, Article ID5264743, 9 pages<http://dx.doi.org/10.1155/2016/526474>.
- [12] SatyabrataAich, A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease, ICACT Transactions on Advanced Communications Technology (TACT) Vol. 7, Issue 3, May2018.
- [13] Leandro A. Passos, Parkinson Disease Identification using Residual Networks and Optimum-Path Forest, SACI 2018, IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, May 17-19, Timiúoara, Romania.
- [14] Deepak Gupta, Optimized cuttlefish algorithm for diagnosis of Parkinson's disease, Cognitive Systems Research, Volume 52, December 2018, Pages 36-48 <https://doi.org/10.1016/j.cogsys.2018.06.006>.
- [15] S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research , 2018.
- [16] Jonathan Weyhenmeyer (2014) Muscle artifacts in single trail EEG data distinguish patients with Parkinson's disease from healthy individuals. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- [17]. Quynh Tran Lya (2016) Detection of gait initiation failure in Parkinson's disease patients using EEG signals. 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- [18] ClaasAhlrichsae al, Parkinson's Disease Motor Symptoms In Machine Learning: A Review, Health Informatics- An International Journal (HIJ) Vol.2,No.4,November2013.
- [19] Erikarovini et al, Comparative Motor Pre-clinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches, Annals of Biomedical Engineering (2018)<https://doi.org/10.1007/s10439-018-2104-9>.
- [20] PratibhaSurathietal Research in Parkinson's disease in India:A review,AnnIndian Acad Neurol. 2016 Jan-Mar; 19(1): 9– 20.doi:10.4103/0972-2327.167713.
- [21] EduardoTolosaetal,Decision Support System for the Diagnosis of Parkinson's Disease, SCIA 2005, LNCS 3540, pp. 740–749,2005.

- [22] Gert Cauwenberghs et al, Incremental and Decremental Support Vector Machine Learning, Proceedings of NIPS 2000, Pages 384-394.
- [23] Chien-Wen Cho et al, A vision-based analysis system for gait recognition in patients with Parkinson's disease, Expert Systems with Applications 36(2009)7033–7039
- [24] Resul Das et al, A comparison of multiple classification methods for diagnosis of Parkinson disease, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO.2, MARCH 2002.

37130185,37130194_report_1_- converted.docx

by

Submission date: 16-Apr-2021 06:18AM (UTC-0700)

Submission ID: 1560933708

File name: 37130185,37130194_report_1_-converted.docx (241.49K)

Word count: 8003

Character count: 44730

CHARACTERISTICS OF WAVEFORM FOR DETECTION OF PARKINSON'S DISEASE USING EEG WAVES

²
Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Electronics and Communication
Engineering

by

K. SAI CHARAN (37130185)

K. YASWANTH (37130194)



²
**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING
SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade 'A' by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI
SALAI,**

CHENNAI - 600 119

MARCH 2021



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)



Accredited with Grade 'A' by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

www.sathyabama.ac.in

**DEPARTMENT OF ELECTRONICS AND
COMMUNICATION ENGINEERING**

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of K. SAI CHARAN(37130185) and K. Yaswanth(37130194) who carried out the project entitled 'CHARACTERISTICS OF WAVEFORM FOR DETECTION OF PARKINSON'S DISEASE USING EEG WAVES' under our supervision from September 2020 to March 2021.

Internal Guide

Dr. P. GRACE KANMANI PRINCE (M.E.,Ph.D.)

Associate Professor, Department ofECE

**HEAD OF THE DEPARTMENT
Dr. T. RAVI**

Submitted for Viva Voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

We, **K.SAI CHARAN (37130185)** and **K.YASWANTH(37130194)** hereby declare that the Project Report entitled '**CHARACTERISTICS OF WAVEFORM FOR DETECTION OF PARKINSON'S DISEASE USING EEG WAVES**' done by us under the guidance of **Dr. P. GRACE KANMANI PRINCE,** ² is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

1)

2)

DATE:

PLACE:

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr.N.M.NANDHITHA, M.E., Ph.D. Dean, School of Electrical and Electronics Engineering** and **Dr. T. RAVI, M.E., Ph.D. Head of the Department, Department of Electronics and Communication Engineering** for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **Dr. P. GRACE KANMANI PRINCE , M.E., Ph.D. Associate Professor, Department of Electronics and Communication Engineering** for her² valuable guidance, suggestions and constant encouragement paved way for the successful completion of the project work.

We wish to express our thanks to all teaching and Non-teaching staff members of the **Department of Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

We express our gratitude to our parents for their constant encouragement and support² for the completion of the project.

ABSTRACT

The motive of this work was to explore the automated diagnosis of Parkinson's disease (PD) using electroencephalography (EEG). Diagnosis of Parkinson's disease (PD) in the early stages is very critical for effective treatments. In this project, we propose a simple and low cost biomarker to diagnose PD, using the electroencephalography (EEG) signals. In the proposed method, EEG is used to detect the brain electrical activities in internal regions of brain, e.g., basal ganglia (BG). Neural activity in the β frequency range (13–30 Hz) is excessively synchronized in Parkinson's disease (PD). We show that these features of β waveform shape also distinguish PD patients on and off medication using non-invasive recordings in a dataset of a few PD patients with resting scalp EEG. Based on the high correlation between PD and brain activities in the brain ganglia, the proposed project, it provides a highly accurate PD diagnostic measure. Moreover, a quantitative measure of disease severity, using the spectral analysis of extracted brain sources is obtained.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF ABBREVIATIONS	x
	LIST OF FIGURES	viii
	LIST OF TABLES	ix
1	INTRODUCTION	1
	1.1 OUTLINE OF THE PROJECT	1
	1.2 PARKINSON'S DISEASE	1
	1.3 ELECTROENCEPHALOGRAPHY	3
	1.4 ARTIFICIAL INTELLIGENCE	4
2	LITERATURE REVIEW	4
3	AIM AND SCOPE	11
	3.1 AIM OF THE PROJECT	12
	3.2 SCOPE OF THE PROJECT	12
4	MATERIALS AND METHODS	12
	4.1 SIGNAL PROCESSING	13
	4.2 PRE-PROCESSING	15
	4.2.1 PRINCIPLE COMPONENT COEFFICIENT	15
	4.2.2 XLSREAD	16
	4.2.3 MACHINE LEARNING	16

	4.2.4	CLASSIFICATION LEARNER	17
	4.2.5	CONFUSION MATRIX	20
5		RESULTS AND DISCUSSION	23
	5.1	RESULTS	23
	5.1.1	CLASSIFICATION LEARNER OUTPUT	23
	5.1.2	CONFUSION MATRIX	26
	5.2	DISCUSSION	32
6		CONCLUSION	33
	6.1	CONCLUSION	33
		REFERENCES	34

LIST OF FIGURES

FIG NO	TITLE	PAGE NO
4.1	BINARY CLASSIFICATION METRIC	21
5.1	SCATTER PLOT OF ORIGINAL SET	24
5.2	SCATTER PLOT OF PREDICTED	25
5.3	CONFUSION MATRIX	26
5.4	TRUE POSITIVE RATE AND FALSE NEGATIVE RATE	27
5.5	POSITIVE PREDICTIVE AND FALSE DISCOVERY	28

LIST OF TABLES

TABLE	TITLE	PAGE NO
5.1	CLASSIFICATION LEARNER OUTPUT	23
5.2	BINARY CLASSIFICATION METRIC - YOUNG	29
5.3	BINARY CLASSIFICATION METRIC - OLD	30
5.4	BINARY CLASSIFICATION METRIC - PD	31

LIST OF ABBREVIATIONS

PD	PARKINSON'S DISEASE
EEG	ELECTROENCEPHALOGRAPHY
AI	ARTIFICIAL INTELLIGENCE
SVM	SUPPORT VECTOR MACHINE
KNN	K-NEAREST NEIGHBOURS
DFA	DISCRIMINANT FEATURE ANALYSIS
DSP	DIGITAL SIGNAL PROCESSING
ELM	EXTREME LEARNING MACHINE
MuPAD	MULTIPROCESSING ALGEBRA DATA
PCA	PRINCIPAL COMPONENT ANALYSIS
PPQ	PHOTO PLETHYSMOGRAPHY
PVM	PARALLEL VIRTUAL MACHINE

CHAPTER-1 INTRODUCTION

OUTLINE OF THEPROJECT

In the brain, the maximum number of the neurons is reached at birth, and, unlike other cells in the human body, these neurons cannot be repaired. Thus, over the time, neurons die and fail to be replaced. Usually, Parkinson's disease (PD) arises from the death of these nerve cells. The nerve cells generate dopamine, which is a chemical substance that mainly controls the body's motion. Therefore, the quantity of the generated dopamine decreases when nerve cells die, which begins to affect the multiple communication modes of the brain. This disease appears mostly in people 50 years or older. Unstable posture, muscle stiffness, slow movement, tremor, loss of balance, and reduced fine motor skills are some initial signs of PD. Hence, intelligent detection methods can be valuable for the early diagnosis of abnormal signs. As automated diagnosis systems, these methods are able to objectively detect PD by electroencephalogram (EEG) signals. Using the EEG signals, functions of cortical (or sub-cortical) segments in the brain can be easily detected. Other disease brain-related diseases, such as Alzheimer's and epilepsy, can be identified by these signals. We thus aimed to use these EEG signals to configure a computer-aided system capable of diagnosing PD. Collection of statistical dataset is taken and manipulations are done on them using Classification Learner tool in Matlab and infer whether the person is suffering from Parkinson's Disease with the help of Machine Learning Algorithms accuracy which will be explained later in this report.

PARKINSON's DISEASE

Parkinson's sickness is a condition whose fundamental highlights are

- Slowed development
- tremor
- gait or balance issues

More than 1 million individuals in the United States have Parkinson's illness. In spite of the fact that it all the more regularly creates in individuals in their 60s or more seasoned, it can happen as ahead of schedule as age 20. Parkinson's infection is brought about by a deficiency of nerve cells in a particular piece of the cerebrum called the substantianigra. Cells in the substantianigra speak with other development control focuses in the mind by discharging dopamine and different synapses. When substantianigra cells bite the dust, they quit emitting dopamine, and the other development control focuses become unregulated. This unsettling influence in the development control focuses of the mind cause the fundamental indications of Parkinson's sickness (gradualness, quake and stride or equilibrium issues). In many patients, what causes the deficiency of substantianigra cells stays obscure. Natural poisons that may cause Parkinson is incorporate Manganese, carbon monoxide, natural solvents, certain pesticides. However, the vast majority with Parkinson's sickness have not been presented to these poisons. Various qualities have been appeared to cause Parkinson's sickness in about 10% of patients, however these quality anomalies don't happen in many people with Parkinson's illness. Presently, analysts presume that the reason for Parkinson's infection in many people mirrors a blend of hereditary components and ecological exposure(s). The principal indication and illness course fluctuates extensively in individuals with Parkinson's sickness. Be that as it may, a few general assertions can be made. The most well-known introductory manifestation is quake. Manifestations of quake or gradualness typically start on one side of the body, frequently including either an arm or leg. Over the long run (in certain individuals, 5 years or more) the other appendage on a similar side, and afterward appendages on the contrary side are influenced. Versatility (emerging from a seat or strolling) might be restricted as the infection advances. For most people on treatment, life expectancy isn't shorted by Parkinson's infection. Notwithstanding the notable manifestations of quake, gradualness, and lopsidedness, different issues like abundance spit, a delicate voice, just as hand or foot spasms may happen. Loss of the feeling of smell is normal. Melancholy, tension and various rest problems are very normal. Stoppage, more successive pee, incontinence, expanded perspiring, and feebleness may create. About 40% of individuals with Parkinson's infection will create changes in memory and scholarly capacity.

ELECTROENCEPHALOGRAPHY(EEG)

The electroencephalogram (EEG) is a chronicle of the electrical movement of the cerebrum from the scalp. The primary accounts were made by Hans Berger in 1929 albeit comparative investigations had been completed in creatures as right on time as 1870. The waveforms recorded are thought to mirror the action of the outside of the mind, the cortex. This action is affected by the electrical movement from the mind structures under the cortex. The nerve cells in the cerebrum produce flags that are called activity possibilities. These activity possibilities move starting with one cell then onto the next across a hole called the neurotransmitter. Uncommon synthetic substances called synapses assist the signs with getting across the hole. There are two sorts of synapses, one will help the activity potential to move to the following cell, the other will stop it moving to another nerve cell. The cerebrum typically strives to keep an equivalent measure of every one of these synapses in the mind. EEG movement is minuscule, estimated in microvolts (μV) with the fundamental frequencies of interest up to around 30 Hertz (Hz). Little metal plates called cathodes are put on the scalp in uncommon positions. These positions are distinguished by the recordist who estimates the head utilizing the International 10/20 System. This depends on taking estimations between certain fixed focuses on the head. The anodes are then positioned at focuses that are 10% and 20% of these distances. There are an incredible assortment of cathodes that can be utilized. The larger part are little plates of tempered steel, tin, gold or silver covered with a silver chloride covering. These ordinarily have a lead appended. Elective strategies comprise of a cap wherein the cathodes are as of now imbedded.

ARTIFICIAL INTELLIGENCE

Artificial Intelligence is the ability to design smart machines or to develop self-learning software applications that imitate the traits of the human mind like reasoning, problem-solving, planning, optimal decision making, sensory perceptions etc. According to John McCarthy, the father of AI, 'The science and engineering of making intelligent machines, especially intelligent computer programs' is the definition of Artificial Intelligence. AI, as the name suggests, is imparting intelligence to the machines so that the machines operate like human beings.

AI is that sector in computer science that emphasizes the creation of intelligent machines that work, operate and react like human beings. AI is used in decision making by the machines considering the real-time scenario. An Artificially Intelligent machine reads the real-time data, understands the business scenario and reacts accordingly. AI has now become a very important part of Information Technology. This branch aims to create machines that are intelligent. The process of transforming a computer into a computer-controlled robot or designing software that thinks and reacts exactly the way a human being thinks is what AI is all about.

CHAPTER-2

LITERATURE REVIEW

Overview did for the determination of PD with various calculations and approaches:

A few procedures¹ are recorded for early discovery of PD dependent on the diverse ML strategies. Be that as it may, exactness in recognition and arranging inside the time is vital or something bad might happen, it causes advancement of more indications. There are various types of information, cerebrum MRI pictures, Voicedata, pose pictures, sensor caught information, manually written information, utilizing which we can anticipate whether individual is having PD or not.

Max A. Little et al introduced another dysphonia measure, pitch period entropy (PPE) and utilized a part support vector machine and has accomplished grouping precision of 91%[1].

Rainer Scho Nweiller et al recognized an alternate methodology which utilized voice investigation with ANN and got great outcomes however seen¹ that cost-adequacy remains

to be a challenge[2].

Marius Ene et al recommended NN based methodology with three kinds of inward techniques and separated people having PD with sound persons[3].

5

David Gil A, Magnus Johnson B found that with¹ fewer neurons at covered up layer both preparing set and test sets performed ineffectively. With higher number of neurons, the preparation set performed well with high danger of over fitting. The ideal answer for this layer was discovered to be 13 neurons[4].

Ipsita Bhattacharya et al recognized the ROC bend variety and distinguished that estimations of TP and FP rates show changes while expansion in the CV folds[5].

Freddie Åström et al proposed novel methodology of equal neural organizations and afterward result of each neural organization is surveyed by utilizing a standard based framework for the choice. During the preparation interaction, information that isn't yet scholarly of each neural organization is gathered and applied in the preparation set of the later neural organization. This assisted with expanding forecast accuracy[6].

AthanasiosTsanas et al built up a novel calculation dependent on discourse flags however it's problematic as the greater part of the highlights are not considered here, just 10 highlights are used[7].

Hui-Ling Chen et al proposed FKNN focused framework utilizing a 10-crease cross approval method[8].

Mohammad S Islam et al has looked at different ML methods dependent on their presentation correctnesses in deciding if individual is having PD or not and referenced that new classifier might be worked to improve accuracies[9].

6

Bo Penga et al proposed Computer Aided Analysis with picture information and utilized BrainLab programming for handling the pictures and figure thickness of the cortex, volume of dark matter, and surface space of the cortex on every district of interest (ROI). Utilization of Multilevel ROI-based highlights improved the order performance[10].

DeriyaAvci and AkifDogantekin proposed another methodology utilizing Genetic Algorithm-Wavelet Kernel-Extreme Learning and accomplished great precision results[11].

SatyabrataAich proposed an exceptional methodology by utilizing Genetic calculation

and PCA as highlight choice strategies and applied seven ML calculations for characterization, that saved time and profitability while doing design arrangement with two classifications, for example, PD and not PD[12].

Leandro A. Passos looked at ResNet-50 , Optimum-Path Forest (OPF) classifier with Support Vector Machines (SVM) and Bayes and accomplished 96% of identification rate[13].

Deepak Gupta followed an alternate methodology cuttlefish calculation and utilized for highlight choice, distinctive wellness capacities approximations are utilized to improve cuttlefish calculation and is named as Optimized cuttlefish calculation (OCFA).Decision tree and K-Nearest Neighbor classifiers are applied and accomplished 94% of precision in recognizing PD affected patients[14].

Salama A. Mostafa proposed Multiple Feature Evaluation Approach (MFEA) of a multi-specialist framework (ii) Implementation of five characterization constructions which are Decision Tree, Random Forests, Neural Network, Naïve Bayes and Support Vector Machine on the Parkinson"s diagnosis before and after applying their approach, and (iii)Author approach witnessed the following average rate of accuracies : Decision Tree achieved accuracy of 10.51%, Naïve Bayes shown 15.22%, Neural Network is found with 9.19%, Random Forests and SVM performed with 12.75% and 9.13%respectively[15].

It is important to note that , out of all ML techniques, ANN and SVM classifier are used most of the proposed algorithms to aid faster and accurate the prediction.

DIAGNOSIS OF PARKINSON'S DISEASE USING EEG

Using EEG Signal the most common disabling disturbances of PD with respect to the physical activity are the Freezing of Gait (FOG), Turning Freeze (TF), and Gait Initiation Failure (GIF). The PD affected patient's experiences difficulty in walking as they feel that their feet have been glued to the floor. During this stage the EEG signals can be recorded from the patients given them a TUG tasks. The cognitive process (brain mechanism) can be analyzed for the timing and scaling of movement using EEG signal analysis and could be a useful tool in the detection of above-mentioned disturbances. The below section discusses the related work carried in the analysis of PD using EEG signals. AMardiHandojoseno, et.al discusses the method of detection of Freezing of Gait in PD patients using EEG signal based on wavelet decomposition. In this paper the dynamics of FOG was analyzed by extracting the EEG Entropy and Sub-band wavelet energy. 26 Patients around 60-70 years with PD and significant FOG were considered. The author highlights that by using BPNN (Back Propagation Neural Network) classifier the onset of Freezing of PD can be identified with 75% accuracy, specificity and sensitivity using sub bands of EEG.

Also, the statistical analysis indicates that there will be a difference in the Energy sub bands of delta, theta and alpha with Normal and PD affected group. Finally gives a future scope for the usage of different classifiers for still better results . Jonathan Weihenmayer[16] proposes the method of diagnosing of PD patients with a Healthy control by extracting the Raw and Clean EEG data of patients during the quasi resting state for

1 sec of Nine PD patients and ten healthy individuals. Nonlinear delay differential equations (DDE) are applied to classify the PD patients from and Healthy individual.

The classification is done upon the plotting of area under the receiver operating characteristic curve (AROC). It is shown that the raw EEG data (AROC = 0.59-0.86) can classify the PD patients with a Healthy individual more robustly than clean EEG data (0.57-0.72). The author states that increased classification of raw EEG time series originates from muscle artifacts and concludes by stating that a non-linear extraction and classification of raw EEG data can be used as biomarker for PD. This paper can be extended for a further scope for research by considering the patients in true resting states and for a larger data sets and sample size.

Quynh Tran Ly et al.[17] discuss the brain mechanism during the Gait Initiation Failure (GIF), which is a subtype of Freezing of Gait (FOG). EEG signals of 61 samples were recorded from Four PD affected patients (3 Male and 1 Female) to study the brain dynamic changes during the occurrence of GIF. EEG samples were analyzed through the Centroid Frequency (CF) and Power Spectral Density (PSD). The results showed that GIF were associated with significant increase in high beta band (21-38 Hz) in the frontal, central, parietal, occipital regions of brain. Using BPNN classifier the analysis of PSD and CF in EEG provided 84.27 % of performance and 84.80% of sensitivity and accuracy respectively. Concludes by stating that High beta activity was found in EEG with PD affected patients during the analysis of GIF and suggested to use a larger sample size and other multiple electrodes. The identification of combination of sensors and optimal location are also discussed. EEG measures/parameters, in the form of Power Spectral entropy, Power Spectral density and Centroid Frequency for 343seconds from 7 patients(6Mand1F)with, Effective walking and PD affected with FOG during a series of TUG were extracted to analyze the important features that were significantly changed during FOG. Feed- Forward Neural networks were applied to classify the EEG signals amongst the Effective walking and Freezing. The sensitivity and accuracy from the classification using single channel, two channels, three, Four, Six, Seven and Thirty-two channels were discussed. The performance of C4O2 in a two channel and four locations at F3C4P4O2 provided the best indicator of FOG detection with 72.54% and 72.29% of

sensitivity and 69.71% and 70.29% of accuracy respectively. The outcome of the paper is that the best montage is using EEG system with two channels C4-O2. The author concludes by stating that the increase in the number of channels will not increase the accuracy and suggests that the further scope of research can be done by considering larger sample of data and to use of any specific Neural Network which can lead to a better result. A method for the suppression of Artifacts that occur in EEG signals due to Neuro stimulation.

The Neurostimulation is assessed by Advanced Statistical Signal Processing and Machine Learning Methodologies. The unsupervised clustering method is applied to neural recordings in order to suppress Neurostimulation. The author also discusses about the Deep Brain Stimulation (DBS) method for the treatment of PD, where electrical stimulator is placed into the body and it delivers the electrical current to thalamus in order to reduce the severity of PD. This Simulator can send the EEG signals which indicates the treatment and possible DBS side effects. Time Frequency features in EEG are extracted in patients undergoing DBS treatment of PD using matching pursuit decomposition (MPD). This paper concludes by stating that the behavioral task clustering for PD patients with DBS using Time frequency features gives the differences in the band activity (delta, beta and gamma) of EEG. The authors Quynh Tran Lya et al. [13] briefs about technique of S-transform decomposition of EEG Signals in detection of Turning freeze in PD patients. Turning Freeze (TF) is the most common subtype of FOG and is causing the majority rate of falls in PD patients.

The brain signals were captured to using EEG signal to measure the feet getting stuck while making a turn (TF). Data was collected from six PD affected patients during TUG tasks. The main discussion in this paper is the extraction of EEG features using time frequency technique S-Transform decomposition (Stockwell transform). Independent Component analysis by entropy bound minimization technique was used to separate EEG sources. Bayesian neural networks (BNN) classification algorithm was implemented to improve the classification results in detecting TF in PD patients. The

main findings in this paper are the inclination in the high beta activity in the region of cortex near occipital O2 and frontal F4 locations. The author concludes the paper by projecting 84.24% sensitivity, 88.0% of specificity and 86.2% accuracy in detecting the TF. The scope for the future work is to increase the accuracy. AM ArdiHandojoseno et al. proposes a methodology for the prediction of FOG in PD using EEG spatial correlation, wavelet coefficient and cross frequency energy techniques. Integration of spatial, spectral and temporal features of EEG signal with the use of Neural Network and K-NN classifiers is discussed.

CHAPTER-3

AIM AND SCOPE

AIM OF THE PROJECT

We propose detection of Parkinson's Disease using the EEG. To study the characteristics of the waveform obtained by EEG. Machine learning is used to verify the performance of the labeled input readings. Early diagnosis of Parkinson's Disease is necessary as by getting late it becomes more and more hazardous which may ultimately lead to death.

SCOPE OF THE PROJECT

As diagnosis is currently rather subjective, researchers have been trying to find an easier and more scientific method. An EEG records electrical activity produced by the brain via small sensors attached to the scalp.

Attempts to use EEG readings for Parkinson's disease diagnosis have not always produced the results that researchers have been looking for. Different architectures of Artificial Neural Network (ANN) such as Convolutional Neural Network (CNN), Long-short term Memory, and Extreme Learning Machine are well-known deep learning algorithms for future growth (ELM). Finding that a noninvasive method such as an EEG could be a promising diagnostic tool may have important consequences for the future of the disease. One can use the test to track changes related to Parkinson's in the brain over the years.

The Classification Algorithms from Machine Learning and Deep Learning are used to predict and investigate the Parkinson's Disease. The ideal features from the dataset are passed as input to the models and the prediction results are obtained.

CHAPTER- 4

METHODS AND MATERIALS USED

Parkinson's disease (PD) is one of the most common neurodegeneration disorders and it is believed that it affects 0.3% of the global population. PD normally occurs in people aged from 50 to 60 and is not common before 50's. When it appears before age 21, genetic factors are presumed to be the cause. The origin of PD remains unknown until today, although it is believed that the interaction between environmental and genetic factors may explain its appearance. The lack of a cure for PD makes it important to research a method for early stages detection in order to delay the disease progression and its symptoms, thereby improving the patient's quality of life. Patients diagnosed with PD are known by the rest tremor but also present other symptoms, such as: muscular rigidity, bradykinesia or postural instability. Moreover, they present an accumulation of α -synuclein, which leads to the formation of Lewy bodies and a progressive loss of dopaminergic neurons in the substantianigra . In order to describe how the symptoms of PD progress, the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Hoehn and Yahr scale are commonly used.

SIGNAL PROCESSING

Signal Processing Toolbox includes functions and applications for analyzing, pre-processing, and extracting features from signals that are sampled uniformly or non-uniformly. Filter configuration and analysis, resampling, smoothing,

detrending, and power spectrum estimation are all included in the toolbox. Extracting features such as change points and envelopes, detecting peaks and signal patterns, quantifying signal similarities, and performing measurements such as SNR and distortion are all possible with the toolbox. Vibration signals may also be subjected to modal and order review.

Without writing code, Signal Analyzer allows you to pre-process and analyze several signals simultaneously in the time, frequency, and time-frequency domains; explore long signals, and extract regions of interest. Filter Designer allows you to build and analyze digital filters by selecting from a number of different algorithms and responses.

PREPROCESSING

PRINCIPALCOMPONENTCOEFFICIENT

The principal component coefficients, also known as loadings, for the n -by- p data matrix X are returned by `PCA(x)`. Observations are represented by rows of X , while variables are represented by columns. It's a p -by- p coefficient matrix. The coefficients for each principal component are mentioned in descending order of component variance in each column of coefficients. By default, PCA uses the Singular Value Decomposition (SVD) algorithm to center the data.

A quantitatively rigorous method for achieving this simplification is principal component analysis. The principal components method creates a new set of variables. The original variables are linearly combined components are orthogonal to one another. The principal components as a whole provide an orthogonal foundation for the data space. in each principal factor. There is no redundant knowledge since all of the principal

An orthogonal basis for multiple columns of data can be constructed in an infinite number of ways.

A single axis in space is the first major part. The values that result from projecting each observation on that axis form a new variable. And this variable's variance is the highest of all the options on the first axis.

The second main part is a perpendicular axis in space to the first. When you project the observations onto this axis, you get a new variable. This attribute has the highest variance out of all the options on the second axis.

The total number of principal components is the same as the number of variables in the original package. The number of the variances of the first few principal components, on the other hand, is often greater than 80% of the total variance of the original results. Researchers also gain a better understanding of the driving forces that produced the original data by analyzing plots of these few new variables. The PCA function aids in the discovery of the primary components. The real calculated data had to be analyzed in order to use PCA.

XLSREAD

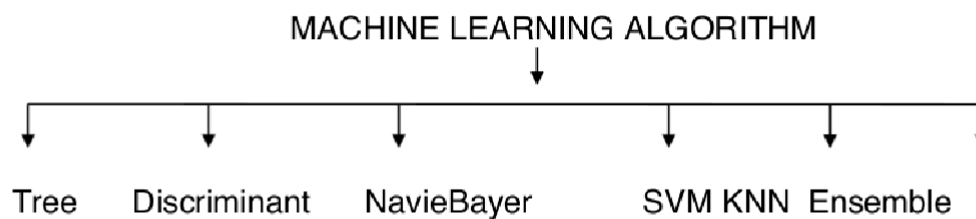
`A = xlsread('filename')` returns numeric data from the first sheet in the file name Microsoft Excel spreadsheet format in array A. Leading rows or columns of text are ignored by `xlsread`. If a cell in a leading row or column is empty or contains text, however, `xlsread` replaces it with a NaN.

MACHINE LEARNING

Machine learning teaches machines to learn from experience in the same way as humans do. Machine learning algorithms employ analytical methods to 'learn' knowledge directly from data rather than relying on a model based on a predetermined equation. As the number of samples available for learning grows, the algorithms adapt their output. Machine Learning uses two ways of technique

- Supervised learning is a method of training a model using known input and output data in order to predict potential outcomes.
- Unsupervised learning is the discovery of hidden patterns or intrinsic structures in data without the use of supervision.

In the face of uncertainty, the aim of supervised machine learning is to create a model that can make predictions based on data. A supervised learning algorithm uses a known collection of input data and known responses to the data (output) to train a model to make accurate predictions about how new data will react. To build predictive models, supervised learning employs classification and regression methods.



CLASSIFICATION LEARNER

Importing data, defining validation schemes, interactively exploring the data, selecting features, training models, and evaluating model efficiency are all possible with Classification Learner.

The Classification Learner app teaches data classification to models. It uses automatic training to find the best classification model sort, which can include decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive Bayes, and ensemble classification. Techniques for classifying data predict categorical responses, such as whether an email is real or spam, or whether a tumor is cancerous or benign. Input data is classified using classification models. Medical imaging, image and speech recognition, and credit scoring are examples of popular applications.

³ Train Classification Models with the Classification Learner App

The Classification Learner app allows you to use supervised machine learning to train models to classify data. Using different classifiers, the software allows you to interactively explore supervised machine learning.

- ³ Automatically train a number of models and assist you in selecting the right one. ³ Decision trees, discriminant analysis, support vector machines, naive Bayes, and ensemble classification are some of the different types of models.
- ³ Investigate your data, define validation schemes, choose features, and visualize the results. Cross-validation is used by default to defend against over fitting. The validation results will assist you in selecting the best model for your data which are reflected in the plots and performance indicators.
- ³ To make predictions with new data, export models to the workspace. The app often trains a full model as well as a model with the required validation scheme, and the full model is the model you export.
- ³ Build scripts, train with new data, work with large data sets, or change the code for further analysis using MATLAB code generated from the app.

Decision Tree: Decision trees, also known as classification trees and regression trees, are used to forecast data responses. Follow the decisions in the tree from the root (beginning) node down to a leaf node to predict an answer. The answer is stored in the leaf node. Nominal answers, such as 'true' or 'false,' are given by classification trees.

Discriminant Analysis

DFA (Discriminant Feature Analysis) is a mathematical method for classifying unknown persons and determining their likelihood of being classified into a specific category (such as sex or ancestry group). The presumption in discriminant analysis is that the sample is usually distributed for the trait.

Where the parameter or dependent variable is categorical and the indicator or independent variable is the interval in nature, discriminant analysis is a method used by the researcher to interpret the test results.

Support Vector Machine: Support-Vector Machines (also known as support-vector networks) are supervised learning models with associated learning algorithms that analyze data for classification in machine learning. An SVM model is a representation of the examples as points in space, mapped such that the examples of the different categories are separated by a large distance. New examples are then mapped into the same space and assigned to one of the categories depending on which side of the distance they fall on.

Naive Bayer: The Naive Bayes classifier is a basic, but powerful, machine learning classifier. In a Bayesian context, it is a probabilistic classifier that makes classifications using the Maximum A Posteriori decision law. A very simple Bayesian network can also be used to describe it. A similar approach is used by Naive Bayes to estimate the likelihood of various classes based on various attributes. This algorithm is usually used for text classification and problems with multiple groups.

Ensemble Classification: Since ensemble can be trained and then used to make predictions, an ensemble is a supervised learning algorithm. It has been shown that qualified ensembles have greater versatility in the roles they can serve. In theory, this versatility allows them to over-fit the training data rather than a single model, although some ensemble techniques have been shown to minimize problems associated with over-fitting the training data.

Where there is a lot of variety among the models, ensembles seem to produce better results. Many ensemble methods contribute to model diversity by combining several models. More random algorithms (such as random decision trees) may generate a stronger ensemble than very deliberate algorithms (such as entropy-reducing decision trees), which may seem counterintuitive.

KNN Classification: Categorization KNN is a nearest-neighbor classification model in which the distance metric and the number of nearest neighbors can both be modified. You can use a Classification KNN classifier to compute substitution predictions because it stores training data. Alternatively, you can use the predicted approach to identify new findings using the model. The KNN binary (as a two-class system) provides a more reliable data classification, making it advantageous to use k as an odd number to avoid irregular data. In machine learning, the KNN procedure is used. It's an object that's been categorized using a mainstream set of its neighbors, with the determination to assign occurrence for the most common class of its k closest neighbors (k is a positive integer, classically small).

The distance metric used in the past was Euclidean distance; however, this is only suitable for infinite variables. KNN is a modern method for delivering all available cases based on a quantity of assessment (e.g., distance functions). The KNN protocol is similarly straightforward. It regulates the K-nearest neighbors by requiring a minimum distance between the interrogation instance and the training samples.

CONFUSION MATRIX

The rows correspond to the expected class (Output Class) and the columns to the true class on the confusion matrix map (Target Class). The diagonal cells refer to correctly identified observations. The off-diagonal cells refer to findings that were wrongly labeled. Each cell displays the number of observations as well as the percentage of the total number of observations.

The percentages of all the examples predicted to belong to each class that is correctly and incorrectly categorized are shown in the column on the far right of the map. The precision (or positive predictive value) and false discovery rate are two commonly used metrics. The percentages of all the examples belonging to each class that is correctly and incorrectly labeled are shown in the row at the bottom of the plot. The recall (or true positive rate) and false negative rate are two commonly used metrics.

Confusion Matrix for Binary Classification Metric

Confusion Matrix is a square matrix where C_{ij} represents the number of data instances that are known to be in group i (true label) and predicted to be in group j (predicted label).The predicted label for a machine learning algorithm is the predicted label, while the true label for a machine learning algorithm is the true label.

		PREDICTED LABEL	
		NEGATIVE	POSITIVE
TRUE LABEL	NEGATIVE	TRUE NEGATIVE	FALSE POSITIVE
	POSITIVE	FALSE NEGATIVE	TRUE POSITIVE

Fig: 4.1. Binary Classification Metric

When a statement is actually positive and classified as positive, the statement is known as TRUE POSITIVE (TP). When a statement is actually negative and classified as negative, the statement is known as TRUE NEGATIVE (TN). When a statement is actually negative and classified as positive, the statement is known as FALSE POSITIVE (FP). When a statement is actually positive and classified as negative, the statement is known as FALSE NEGATIVE (FN). The Confusion Matrix, which is required to calculate the machine learning algorithm's accuracy in classifying data into its labels.

Accuracy

The number of correctly identified data instances divided by the total number of data instances is known as accuracy. If the set of data's is not balanced, accuracy may

not be in good measures.

$$\text{Accuracy} = \text{TN} + \text{TP} / (\text{TN} + \text{FP} + \text{TP} + \text{FN}).$$

Precision

A successful classifier should have a precision of 1 (high). Only when the numerator and denominator are equal, i.e. $\text{TP} = \text{TP} + \text{FP}$, does precision become 1. This also implies that FP is zero. The value of the denominator becomes greater than the numerator as FP increases, and the precision value decreases (which we don't want).

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}).$$

Recall

A successful classifier should have a recall of 1 (high). Only when the numerator and denominator are equal, i.e. $\text{TP} = \text{TP} + \text{FN}$, does recall become 1. This also implies that FN is zero. The value of the denominator becomes greater than the numerator as FN increases, and the recall value decreases (which we don't want).

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}).$$

F1 Score

Only when precision and recall are both 1 does the F1 Score become 1. Only when both precision and recall are high does the F1 score rise. The F1 score is a better metric than accuracy since it is the harmonic mean of precision and recall.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}).$$

CHAPTER-5

RESULT AND DISCUSSION

RESULT

A dataset of different categories of people(young, old and people suffering from Parkinson's Disease) are collected. We used supervised machine learning to verify our reading efficiency percentage. In machine learning, the filtered signals are then passed on to a classification learner. The parameters are evaluated and the efficiency is verified in the Confusion matrix which produces a precise result.

CLASSIFICATION LEARNER OUTPUT

The Classification Learner is a program that teaches models how to classify data. It uses automatic training to find the best classification model sort, which can include decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, naive Bayes, and ensemble classification.

Table 5.1. Classification Learner Output

Fine Tree	77.2%
Medium Tree	78.3%
Linear Discriminant	75.5%
Quadratic Discriminant	72.5%
Linear SVM	72.8%
Cubic SVM	71.6%
Fine Guassian SVM	74.5%
Medium Guassian SVM	73.0%

The Classification Learner is a program that teaches models how to classify data. It aids in the exploration of supervised machine learning using various classifiers and data, as

well as the selection of features, specification of validation schemes, training of models, and evaluation of performance.

Scatter Plot of Original Datasets

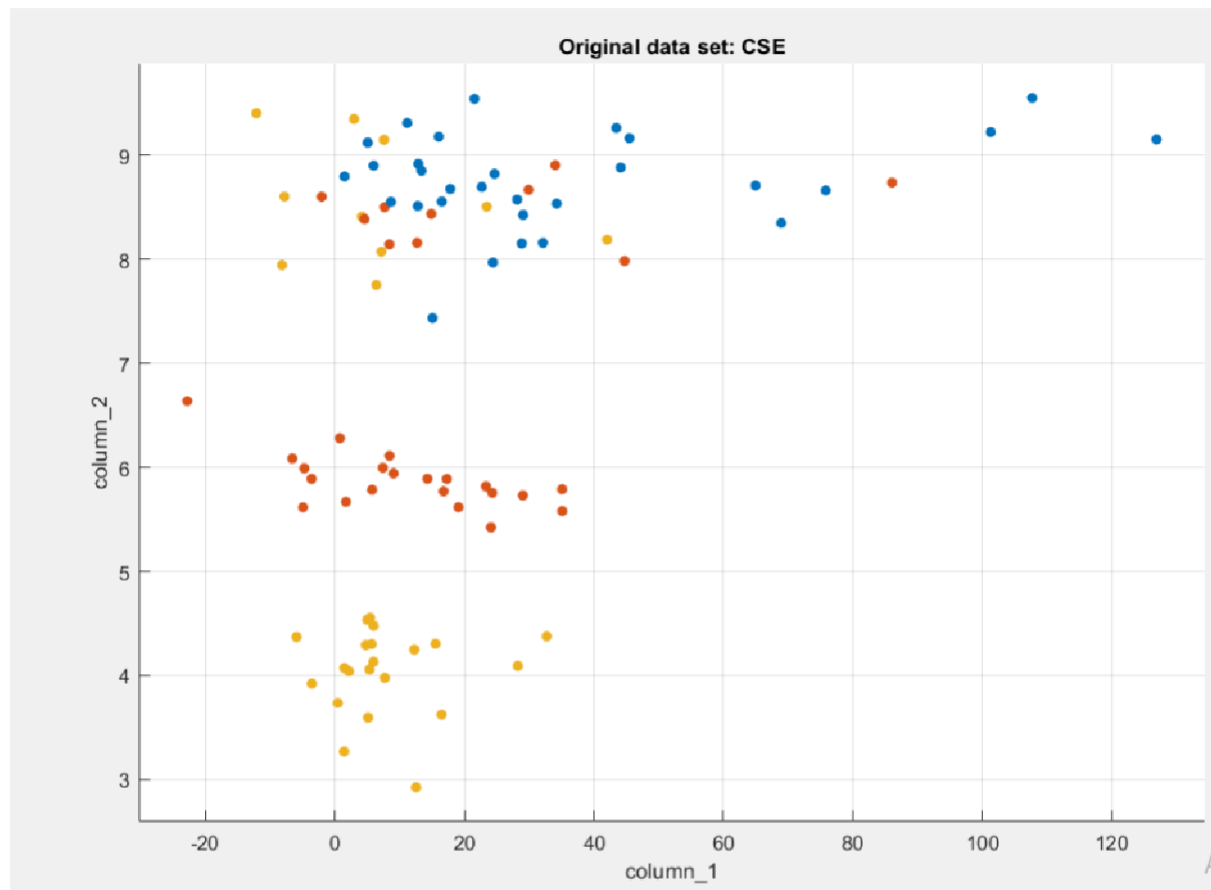


Fig 5.1 Scatter Plot of Original Datasets

In scatter plot the original set of data is obtained. The colors blue, red, yellow, and blue indicates categories like young, old, and people suffering from Parkinson's Disease.

Scatter Plot of Predicted Datasets

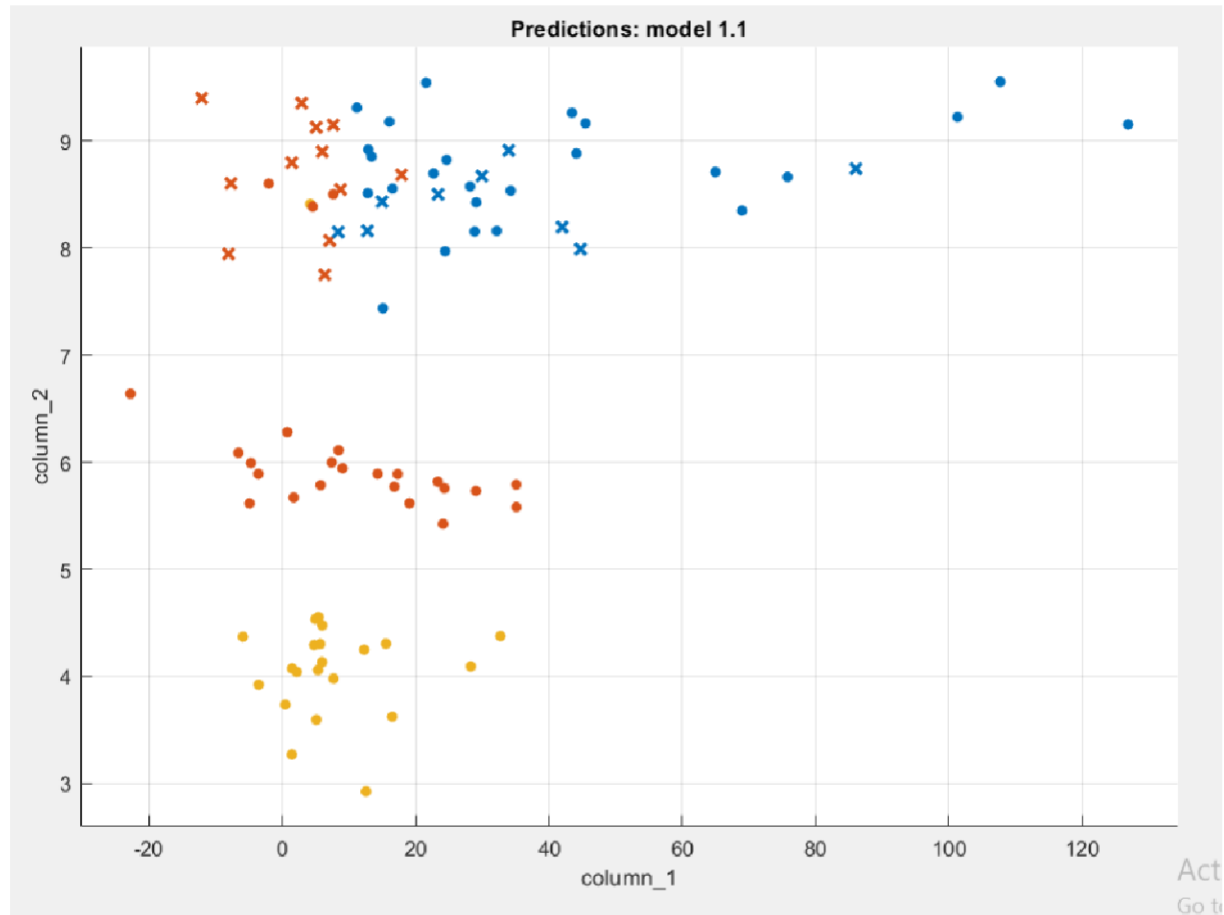


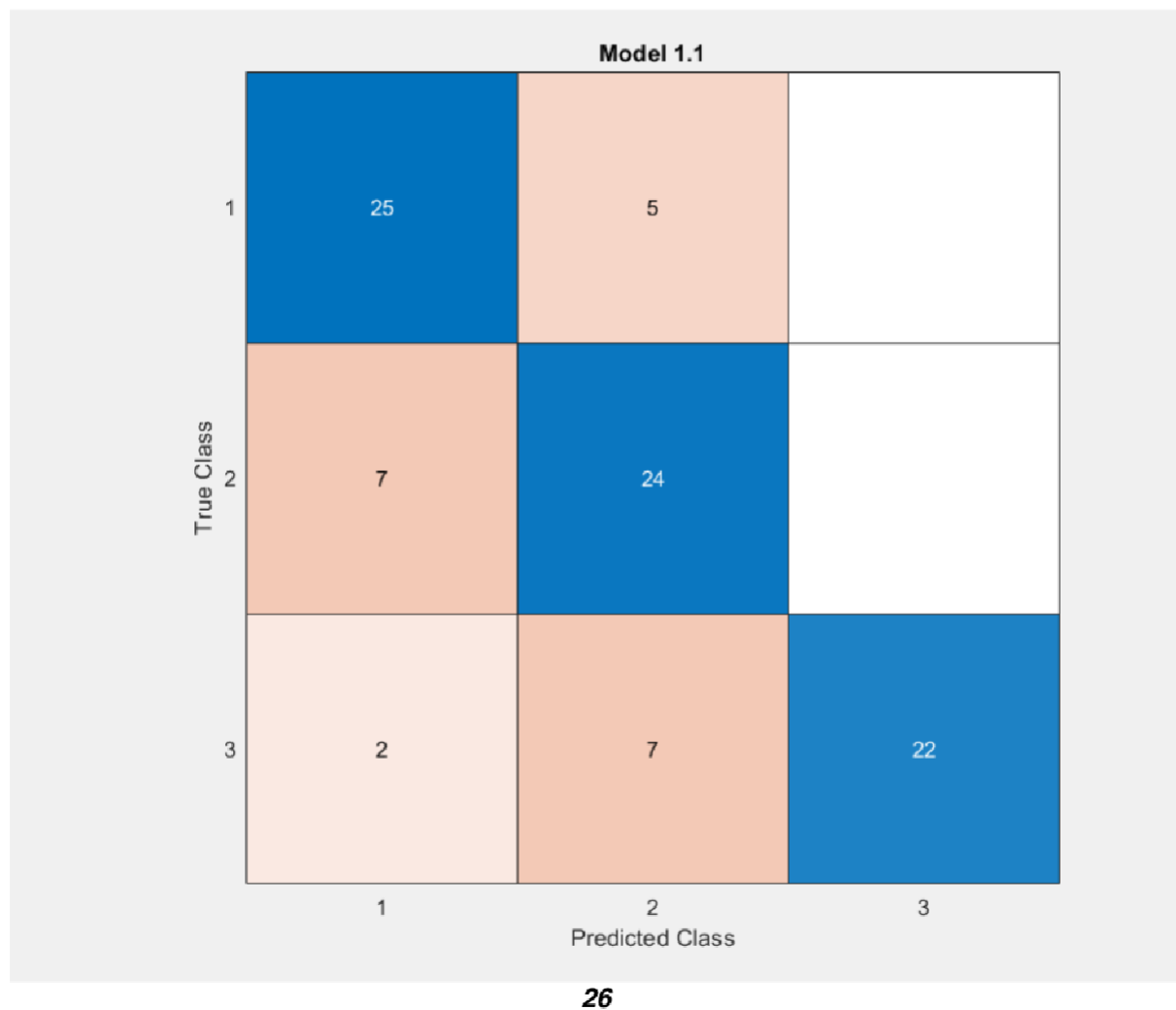
Fig 5.2 Scatter Plot of Predictions

In scatter plot the predicted set of data is obtained. All machine learning algorithms are performed and trained and the results are as shown in the figure. The colors blue, red, yellow, and blue indicates categories like young, old, and people suffering from Parkinson's Disease.

CONFUSION MATRIX:

The feature extraction process is seen in the final production. This process takes place in a matrix of uncertainty. K nearest classification is used in this method. KNN is a lazy learning algorithm that is non-parametric. Its aim is to predict the classification of a signal using a database with data points divided into several groups.

Fig 5.3 Confusion Matrix of Original Dataset



In Confusion Matrix for original set of data's under Number of Observation category the set of data is obtained. The diagonal elements in the Confusion Matrix must be high always as it suggests the true value and predicted value are same which in turn says that the accuracy is higher.

Confusion Matrix for True Positive and False Negative Rates

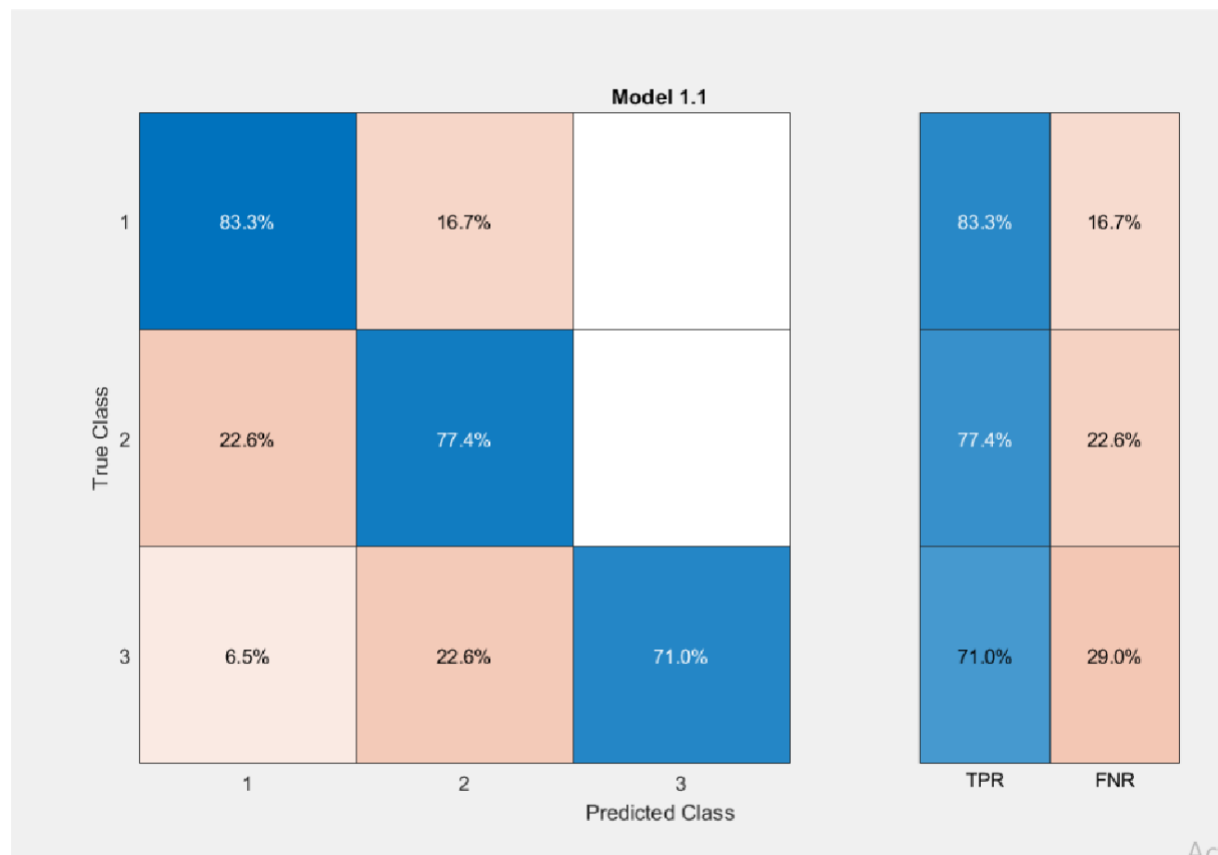


Fig: 5.4. True Positive Rate and False Negative Rate.

In Confusion Matrix for original set of data's under Positive Rate and False Negative Rate category the set of data is obtained. True Positive rates for emotion young is 83.3 %, old is 77.4 %, PD is 71 % and False Negative rates for emotion young is 16.7%, old is 22.6%, PD is 29.0 %.

Confusion Matrix for Positive Predictive and False Discovery

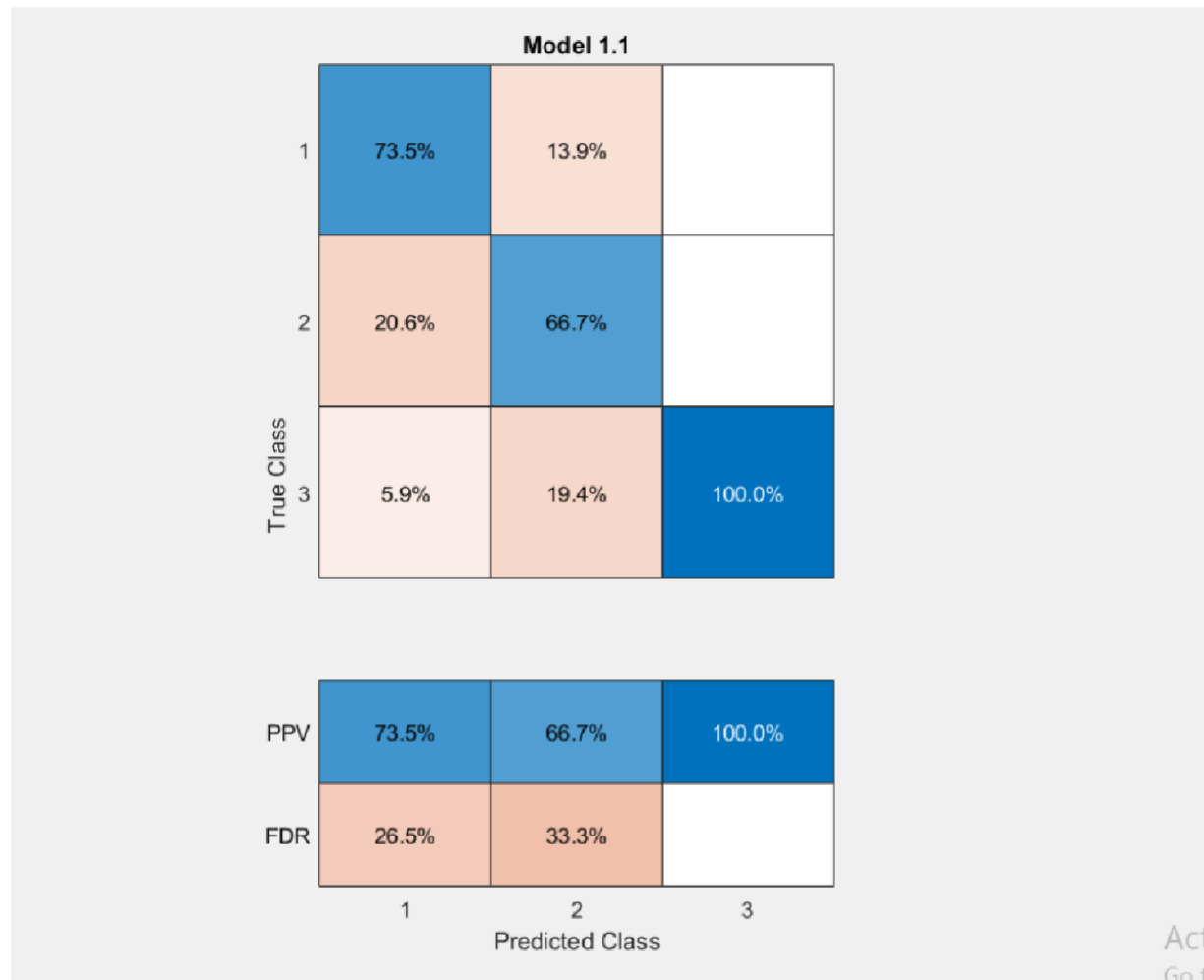


Fig: 5.5. Positive Predictive and False Discovery

In Confusion Matrix for original set of data's under Positive Predictive rate and False Discovery Rate category the set of data is obtained.

Table 5.2. Binary Classification Metric for Young

TRUE NEGATIVE 0.74	FALSE POSITIVE 0.265
FALSE NEGATIVE 0.17	TRUE POSITIVE 0.84

ACCURACY: $= \text{TN} + \text{TP} / (\text{TN} + \text{FP} + \text{TP} + \text{FN})$
 $= 0.74 + 0.83 / (0.74 + 0.83 + 0.17 + 0.26)$
 $= 0.785$
 $= 78.5\%$

PRECISION: $= \text{TP} / (\text{TP} + \text{FP})$
 $= 0.83 / 0.83 + 0.26$
 $= 0.761$

RECALL: $= \text{TP} / (\text{TP} + \text{FN})$
 $= 0.83 / 1$
 $= 0.83$

F1 SCORE: $= 2 * ((\text{Precision} * \text{recall}) / (\text{Precision} + \text{Recall}))$
 $= 2 * (1.26 / 1.59)$
 $= 0.793$

Table 5.3. Binary Classification Metric for Old

TRUE NEGATIVE 0.66	FALSE POSITIVE 0.34
FALSE NEGATIVE 0.23	TRUE POSITIVE 0.77

ACCURACY: $= \text{TN} + \text{TP} / (\text{TN} + \text{FP} + \text{TP} + \text{FN})$
 $= 0.66 + 0.77 / (0.66 + 0.34 + 0.23 + 0.77)$
 $= 0.715$
 $= 71.5\%$

PRECISION: $= \text{TP} / (\text{TP} + \text{FP})$
 $= 0.77 / 1.1$
 $= 0.7$

RECALL: $= \text{TP} / (\text{TP} + \text{FN})$
 $= 0.77 / 1$
 $= 0.77$

F1 SCORE: $= 2 * ((\text{Precision} * \text{recall}) / (\text{Precision} + \text{Recall}))$
 $= 2 * (0.54 / 1.47)$
 $= 0.73$

Table 5.4. Binary Classification Metric for PD

TRUE NEGATIVE 1	FALSE POSITIVE 0
FALSE NEGATIVE 0.29	TRUE POSITIVE 0.71

ACCURACY: = $TN+TP / (TN+FP+TP+FN)$
= $1+0.71 / 2$
=0.85
=85%

PRECISION: = $TP / (TP+FP)$
= $0.71 / 0.71$
=1

RECALL: = $TP / (TP+FN)$
= $0.71 / 1$
=0.71

F1 SCORE: = $2 * ((Precision*recall) / (Precision + Recall))$
= $2 * (0.71 / 1.71)$
=0.83

DISCUSSION

¹ Machine Learning techniques has got prominent role as they are applied in variety of domains especially in the healthcare. Unlike traditional methods, the models generated by applying ML techniques show dynamic outputs as data is fed into it. One shall make note that significant and narrow research is needed to obtain knowledge in diagnosing the disease. Various machine learning algorithms and techniques are being proposed rapidly, out of which some are observed to be promising with the results and few demonstrated their usage in different fields. Advantage with the ML generated models is that when more data is used, the precision values gets increased and the much accuracy in predictions can be gained. In the past several years, several non-invasive techniques have been proposed to identify PD using voice and gait signals. Different automated techniques were explored to develop the best automated model to differentiate normal and PD subjects.

It can be noted that various machine learning techniques were adopted to discriminate the EEG signals in normal and PD subjects. The obtained spectral analysis dataset is trained with Machine Learning Algorithms and the data is manipulated. From the results obtained, the scatter plot of original dataset and the scatter plot of predictions along with the confusion matrix is shown and from the results, the parameters like accuracy, precision, recall and f1 score have been calculated.

CHAPTER-6

CONCLUSION

6.1 ¹ CONCLUSION

This project is an effort to present broad review about Parkinson disease diagnosis system that have applied various machine learning techniques. The summary of results obtained by different researchers is made available in literature survey table , almost all the authors/researchers made great efforts to predict the Parkinson disease with novel approaches. It can be identified that maximum of all ML techniques used by various authors worked better. We trained the original datasets using Machine Learning algorithms that has a fair recognition rate across all three categories using multiple classification techniques and signal processing of sample readings using denoising filters, supervised machine learning, classification learner, and confusion matrix. The results have produced a good accuracy for PD patients of around 85%. Hence the proposed method can be efficient enough to diagnose PD. The proposed approach was tested using machine learning and deep learning functions. Preprocessing is a technique for reducing noise in sample reading input by using statistics coefficient parameters. Computer learning has been sent the filtered info.

REFERENCES

- [1] Max A. Little et al, Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING,VOL. 56, NO. 4, APRIL 2009.
- [2] Rainer Scho¨ Nweiler et al, Novel Approach to Acoustical Voice Analysis Using Artificial Neural Networks ,JARO 01: 270–282 (2000) DOI:10.1007/s101620010020.
- [3] Marius Ene et al, Neural network-based approach to discriminate healthy people from those with Parkinson's disease, Annals of the University of Craiova, Math.Comp. Sci. Ser. Volume 35, 2008, Pages 112–116 ISSN: 1223-6934.
- [4] David Gil A, Magnus Johnson B, Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines, Global Journal of Computer Science and Technology,2009.
- [5] Ipsita Bhattacharya et al, SVM Classification to Distinguish Parkinson Disease Patients, Conference: Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 16-17, 2010, Tamilnadu, India.
- [6] Freddie Åström et al, A parallel neural network approach to prediction of Parkinson's Disease,2011Elsevier,doi:10.1016/j.eswa.2011.04.028.
- [7] Athanasios Tsanas et al, Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease, IEEE Transactions On Biomedical Engineering, Vol. 59, No. 5, May 2012.
- [8] Hui-Ling Chen et al, An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach, Expert Systems with Applications 40 (2013)263–271.
- [9] Mohammad S Islam et al, Performance Comparison of Heterogeneous Classifiers for Detection of Parkinson's Disease Using Voice Disorder (Dysphonia), 3rd International Conference On Informatics, Electronics & Vision 2014.
- [10] Bo Peng et al, Computer Aided Analysis of Cognitive Disorder in Patients with Parkinsonism using Machine Learning Method with Multilevel ROI-based Features, J Med Syst (2015) 39:179 DOI10.1007/s10916-015-0353-9.

- [11] DeryaAvci and AkifDogantekin et al, An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel Extreme Learning Machine, Hindawi Publishing Corporation Parkinson's Disease Volume2016, Article ID5264743, 9 pages<http://dx.doi.org/10.1155/2016/526474>.
- [12] SatyabrataAich, A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease, ICACT Transactions on Advanced Communications Technology (TACT) Vol. 7, Issue 3, May2018.
- [13] Leandro A. Passos, Parkinson Disease Identification using Residual Networks and Optimum-Path Forest, SACI 2018, IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, May 17-19, Timiúoara, Romania.
- [14] Deepak Gupta, Optimized cuttlefish algorithm for diagnosis of Parkinson's disease, Cognitive Systems Research, Volume 52, December 2018, Pages 36-48 <https://doi.org/10.1016/j.cogsys.2018.06.006>.
- [15] S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease, Cognitive Systems Research , 2018.
- [16] Jonathan Weyhenmeyer (2014) Muscle artifacts in single trial EEG data distinguish patients with Parkinson's disease from healthy individuals. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- [17]. Quynh Tran Lya (2016) Detection of gait initiation failure in Parkinson's disease patients using EEG signals. 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- [18] ClaasAhlrichsae al, Parkinson's Disease Motor Symptoms In Machine Learning: A Review, Health Informatics- An International Journal (HIJ) Vol.2,No.4,November2013.
- [19] Erikarovini et al, Comparative Motor Pre-clinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches, Annals of Biomedical Engineering (2018)<https://doi.org/10.1007/s10439-018-2104-9>.
- [20] PratibhaSurathietal Research in Parkinson's disease in India:A review,AnnIndian Acad Neurol. 2016 Jan-Mar; 19(1): 9– 20.doi:10.4103/0972-2327.167713.
- [21] EduardoTolosaetal,Decision Support System for the Diagnosis of Parkinson's Disease, SCIA 2005, LNCS 3540, pp. 740–749,2005.

[22] Gert Cauwenberghs et al, Incremental and Decremental Support Vector Machine Learning, Proceedings of NIPS 2000, Pages 384-394.

[23] Chien-Wen Cho et al, A vision-based analysis system for gait recognition in patients with Parkinson's disease, Expert Systems with Applications 36(2009)7033–7039

[24] Resul Das et al, A comparison of multiple classification methods for diagnosis of Parkinson disease, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO.2, MARCH 2002.

ORIGINALITY REPORT

12%
SIMILARITY INDEX

12%
INTERNET SOURCES

0%
PUBLICATIONS

0%
STUDENT PAPERS

PRIMARY SOURCES

1 www.ijert.org 5%
Internet Source

2 www.coursehero.com 4%
Internet Source

3 uk.mathworks.com 3%
Internet Source

Exclude quotes On

Exclude matches < 3%

Exclude bibliography On