PARKINSONS DISEASE PREDICTION USING ML ALGORITHMS

A FIELD PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE ENGINEERING

BY

B. SRIRAM - Regd. No. 22071A6608
J. HARIHARAN - Regd. No. 22071A6622
M. KHYATHI - Regd. No. 22071A6638
M. SPOORTHY - Regd. No. 22071A6640

UNDER THE GUIDANCE OF

Dr.N.Sandhya

Professor and Head

CS - AIML & IoT Department

DEPARTMENT OF CSE-AIML ENGINEERING

VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING & TECHNOLOGY
PRAGATHI NAGAR, NIZAMPET (S.O),
HYDERABAD - 500 090

SEP – 2023

VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

Estd.1995

An Autonomous Institute, NAAC Accredited with 'A++' Grade
NBA Accreditation for B.Tech. CE, EEE, ME, ECE, CSE, EIE, IT, AME and
M.Tech. STRE, PE, AMS and SE programmes
Approved by AICTE, New Delhi, Affiliated to JNTUH
Recognized as "College with Potential for Excellence" by UGC
Vignana Jyothi Nagar, Pragathi Nagar, Nizampet (S.O), Hyderabad – 500 090, TS, India.
Telephone No: 040-2304 2758/59/60, Fax: 040-23042761
E-mail: postbox@vnrvjiet.ac.in, Website: www.vnrvjiet.ac.in

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the Field Project report entitled "Parkinsons Disease Prediction Using ML Algorithm" being submitted by Mr./Miss. **B.Sriram**, Regd.no. **22071A6608**, **J.Hariharan,** Regd.no**. 22071A6622, M.Khyathi,** Regd.no. **22071A6638, M.Spoorthy,** Regd.no. **22071A6640** in partial fulfillment for the award of BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE AND ENGINEERING(AIML) to the Jawaharlal Nehru Technological University Hyderabad at VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING & TECHNOLOGY, HYDERABAD, is a record of bonafide work carried out by **Dr.N. Sandhya** under our guidance and supervision.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Signature of Supervisor
Dr. N.SANDHYA
Professor & HOD
Computer Science and Engineering
(Department of AIML&IOT)

Signature of HOD
Dr. N.SANDHYA
Professor & HOD
Computer Science and Engineering
(Department of AIML&IOT)

PLAGIARISM CERTIFICATE

PD prediction

ORIGINALITY REPORT

12%	10%	9%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	iarjset.com Internet Source	4%
2	www.ncbi.nlm.nih.gov Internet Source	2%
3	datapunk.net Internet Source	1%
4	ebin.pub Internet Source	1%
5	www.coursehero.com Internet Source	1%
6	www.researchgate.net Internet Source	1%
7	Submitted to West Texas A&M University Student Paper	1%
8	www.ldonline.org Internet Source	<1%
9	Submitted to Kaplan Professional Student Paper	<1%
10	Submitted to University of Sydney Student Paper	<1%
11	Aditi Govindu, Sushila Palwe. "Early detection of Parkinson's disease using machine learning", Procedia Computer Science, 2023 Publication	<1%
12	sedici.unlp.edu.ar Internet Source	<1%
13	"Levodopa and the Progression of Parkinson's Disease", New England Journal of Medicine, 12/09/2004 Publication	<1%
14	doku.pub Internet Source	<1%
15	Submitted to Geneva College Student Paper	<1%
16	www.science.gov Internet Source	<1%

APPROVAL CERTIFICATE

Field project evaluation for the dissertation work entitled **"Parkinsons Disease Prediction Using ML algorithm"** being submitted by **Mr./Miss. 22071A6608 -B.Sriram, 22071A6622 -Hariharan, 22071A6638-M.Khyathi, 22071A6640-M.Spoorty is** conducted on..... and the work is approved for the award of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ENGINEERING(AIML&IOT) ENGINEERING.**

PROJECT REVIEW COMMITTEE

VNR VIGNANA JYOTHI INSTITUTE OF ENGINEERING & TECHNOLOGY
(An Autonomous Institution, Accredited by NAAC with 'A++' grade and NBA)
Pragathi Nagar, Nizampet (S.O.),
Hyderabad - 500090
Telangana

DEPARTMENT OF Computer Science Engineering (AIML)

DECLARATION

I hereby declare that the Field Project report entitled **"Parkinsons Disease Prediction Using ML algorithm",** submitted for B.Tech. degree is my original work and project has not formed the basis for the award of any degree, associateship, fellowship or any similar titles.

Signature of the student :
Mr. B.Sriram
(22071A6608)

Signature of the student
Mr. J.Hariharan
(22071A6622)

Signature of the student
Miss. M.Khyathi Meghana
(22071A6638)

Signature of the student
Miss. M.Spoorthy
(22071A6640)

TABLE OF CONTENTS

CERTIFICATE ii
PLAGIARISM iii
APPROVAL CERTIFICATE iv
DECLARATION v
TABLE OF CONTENTS vi
LIST OF FIGURES viii
LIST OF TABLES ix
NOMENCLATURE x
ACKNOWLEDGEMENTS xi
ABSTRACT xii
CHAPTER 1 - INTRODUCTION
1.1 Introduction to the project 11
1.2 Background 12
1.3 Tools, equipments and terminology used 13
1.4 Outline of the project report 14
CHAPTER 2 - LITERATURE REVIEW
2.1 Overview 15
2.2 Review of literature 18
2.3 Problem statement 19
2.4 Project objectives 19
2.5 Summary 19
CHAPTER 3 – DEVELOPMENT OF PROJECT
3.1 Project methodology 20
3.2 Development of project (Modeling, Analysis, Fabrication, Programming, Simulation etc.) 21

3.3 Results 22

CHAPTER 4 - CONCLUSIONS

4.1 Conclusions 23

4.2 Recommendations 23

REFERENCES **24**

LIST OF FIGURES

Fig. No.	Title	Page No.
1.1	Parkinson's Disease Symptoms	13
2.1	Design of the model	16
3.1	Result	22

LIST OF TABLES

Table No.	Title	Page No.
1.1	Stages in Parkinon's Disease	12
2.1	History of PD	18
3.1	Current drugs for PD	20

NOMENCLATURE

1. PD- Parkinosn’s Disease
2. SVM- Support Vector Machine
3. SVC- Support Vector Classifier
4. EDA- Exploration Data Analysis
5. PET- Positron Emission Tomography
6. SPECT- Single-Photon Emission Computed Tomography
7. MRI- Magnetic Resonance Imaging
8. ML- Machine Learning

ACKNOWLEDGEMENTS

Over a span of one and a half years, VNRVJIET has helped us transform ourselves from mere amateurs in the field of Computer Science into skilled engineers capable of handling any given situation in real time. We are highly indebted to the institute for everything that it has given us. We would like to express our gratitude towards the principal of our institute, **Dr. Challa Dhanunjaya Naidu** and the Head of the Computer Science & Engineering Department, **Dr.N.Sandhya** for their kind co-operation and encouragement which helped us complete the project in the stipulated time. Although we have spent a lot of time and put in a lot of effort into this project, it would not have been possible without the motivating support and help of our project guide **Mrs.N.Sandhya.** We thank her for her guidance, constant supervision and for providing necessary information to complete this project. Our thanks and appreciations also go to all the faculty members, staff members of VNRVJIET, and all our friends who have helped us put this project together.

B. Sriram
22071A6608
B.Tech. (CSE-AIML)

J. Hariharan
22071A6622
B.Tech. (CSE-AIML)

M. Khyathi
22071A6638
B.Tech. (CSE-AIML)

M. Spoorthy
22071A6640
B.Tech. (CSE-AIML)

ABSTRACT

Parkinson's disease is a neurological disease which has symptoms Bradykinesia, eyes blinking, stiffness etc. and has no treatment as of date. Predicting this disease is now a challenging task which can be done using growing technologies. Our project to predict this disease is built using ML algorithm Support Vector Machine (SVM). SVM is mainly used for classification or regression related works. It creates a hyperplane to classify the data to train and test data. After the classification it predicts whether the person is procrastinated to have Parkinson's or not. It is one of the effective tools to predict disease based on the data given. The model we are developing helps to create a potential change in the medical field to predict the disease in advance using SVM's robustness in handling complex data relations. With the help of this prediction the patient if procrastinated, can start precaution steps to avoid further loss, hoping better outcomes. Our research creates awareness of this disease and highlights the importance of early prediction using the symptoms of the disease.

Key Words: Bradykinesia, Neurodegenerative disorder, Dopamine

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION TO THE PROJECT:

Parkinsons disease (PD) is a neurodegenerative disorder and the exact cause for this disease is still elusive. Scientists and researchers say that it may be due to combination of genetic or environmental factors. This disease is mostly affected to people whose age crossed 60 years and few people whose age is nearly 40s. This is characterized by stiffness, slowing of movement, tremors etc. The main reason for this disease is due to lack of dopamine cells or due to lack of dopamine producing cells. Dopamine is a neurotransmitter which coordinates movements and ensures the smoothness of movements. Based on the scan of these cells, Parkinson's can be predicted but by this time the person might reach the third or fourth stage where treatment and cure becomes nearly impossible.

The disease is first described by British physician James Parkinson in the year 1817. In honor of his description the disease is named Parkinsons. This is most common after Alzheimer's disease. PD affects the movement of the person. It is mainly of two types of motor and non-motor based on symptoms. The motor symptoms are tremors, bradykinesia (slowness of movement), rigidity, and postural instability while the non-motor symptoms include cognitive impairment, mood disorders, and autonomic dysfunction. These symptoms impact the movement of the person and cause trouble in the future years. The diagnosis for this disease is clinical and sometimes based on neuroimaging studies.

The Parkinson's disease can be classified into stages similar to that of Cancer. It has 5 main stages. In stage one that is the initial stage where the person has slight symptoms which are covered in daily activities. The symptoms grow slowly from here on. It includes slowness of movement, tremor and rigidity. Slowness is a major step and this in addition to tremor or rigidity should be there for prediction in stage one. Exercise is important so that the disease progress may decrease.

Next is the second stage where symptoms start becoming worse, that is the rigidity of tremors affect the body at neck and trunk. Change in walking and posture is observed but daily tasks handling becomes difficult and time consuming. Due to this there will be reduced arm swing while walking. It also affects sleep quality. The third stage is considered as mid stage where loss if balance is observed which stands as hallmark. In this stage people often fall due to dizziness and motor symptoms become worse. From this stage, a person is restricted to do his day-to-day activities.

Now the person enters into stage four if not taken proper care. In this stage, symptoms are fully developed and is still able to walk but need extra support like walker. Once entered into this stage the person is not suggested to live alone. It becomes difficult even to stand and sit, dress etc. Then comes stage five within no time after entering stage four. This is the most advanced stage and legs stiffness becomes nearly zero making a person unable walk and a person becomes bedridden making life worsen.

Based on the growth of the disease that scientists and researchers found, the importance for the

prediction of this disease and demand for prediction has increased. When predicted early, patients' dopamine cells are being improved by the neurologists and new medications are being discovered and improved for enabling the health of the PD affected person. Using different approaches for diverse aspects researchers are coming up so that they can help treating the person. But the understanding process is ongoing. To develop efficient treatment and efficient cure the disease should be predicted.

STAGE 1	STAGE 2	STAGE 3	STAGE 4	STAGE 5
Mild unsevere symptoms	Moderate symptoms with facial modifications	Progression of disease is occurred.	Drastic change is observed.	Advanced stage with aggressive symptoms
Tremors on one side and postural changes are observed.	Tremors on both sides of the body is observed.	Imbalance of body and improper reflexes are observed.	Personal assistance is required even in simple tasks.	Hallucination and spasm occurs in this stage.

Table1: Stages in Parkinson's Disease

1.2 BACKGROUND:

The background of the project is to develop the ML model to procrastinate Parkinson's disease based on the data we give. The core components of the code are Data Loading and Preprocessing, Data Splitting, Model Training, Model Evaluation, and new instance Prediction. Raw data is converted into structured data under Data Preprocessing which undergoes Exploration Data Analysis (EDA). In data splitting, it trains and tests the data by dividing in ratios 80-20 70-30, or 60-40. We opted for 80-20. Model training is the next important part in which the model is trained to get accuracy. It includes steps of gathering and preparing data, choosing the right algorithm, splitting data for training, and evaluating the model. The next part is instance creation where a new instance of an individual is entered to predict the disease based on the symptoms are stooped posture, masked face, back rigidity, forward tilt of trunk etc.

Fig. 1.1: Parkinson's Disease Symptoms

1.3 TOOLS, EQUIPMENTS AND TERMINOLOGY USED:

Tools, equipment and terminology used in our model is mostly related to python programming and machine learning libraries.

1. Python

Python is versatile and high level interpreted programming language which is known and most used for its easy readability, extensive standard library. It supports multiple programming which includes procedural functional and object-oriented programming. The key features of python programming is readability, versatility, extensive librariesm community support. It has various language components like variables and data types, control flow structures, functions, classes and objects. The ecosystem of python consists of package management, virtual environments. It has popular libraries and frameworks. It has diverse applications like web development, data science, machine learning, automation etc.

2. Pandas

Pandas is powerful library. It is a popular open source data manipulation and analysis library. The key features are dataframe and series, data alignment, handling missing data, data filtering and selection, groupby operations, merge and join operations. The key components are dataframe and series. It supports reading and writing data in various formats which includes CSV, Excel, SQL etc. Pandas provides multiple ways to index and select data. It also offers methods to handle missing data. It is widely used in data science, economics etc.

3. scikit-learn

It is machine learning library which is an open source. It provides tools which are simple and effective. These can be used for data analysis and modeling. The key features of the scikit-

learn are uniform interface, exclusive algorithms, model selcetion and evaluation, data preprocessing, integration with NumPy and pandas. Scikit-learn provides a vide range for model evaluation including accuracy, prediction etc. It also provides techniques for cross-validation. It is widely used in data analysis, predictive modeling, ML research.

4. Support Vector Machine(SVM)

SVM is used for classification and regression tasks. Its primary task is to find a hyperplane that separates data points. The key concepts of SVM are Hyperplane, Support Vectors, Margin. The SVM types are linear SVM, non-linear SVM. SVM is used for classification and regression. It has both pros and cons. The advantages are effective in high-dimensional spaces, versatile due to use of different kernel functions, robust against overfitting. The disadvantages are sensitive to choice and computationally expensive. It is majorly used in text and image classification, face recognition etc. It also has limitations like struggle with large datasets, tuning.

5. Support Vector Classifier

It is a specific way to implement SVM algorithm for classification tasks. It seeks to find a hyperplane that separates the different classes data points. It shares key concepts like hyperplane, support vectors, margin with SVM. It is also classified into 2 types- linear SVC and non-linear SVC. The advantages of SVC are effective for both linear and non-linear classification tasks and robust. It also have few disadvantages. SVC is sensitive to the choice and computationally expensive. It plays major role in bioinformatics.

1.4 OUTLINE OF THE PROJECT REPORT:

The main idea of the project is to develop a machine learning model which predicts whether the person has PD or not. The title of the project is Parkinsons disease prediction using ML algorithms. The model we develop should have minimum cost and maximum accuracy. The previous studies of how researches have detected this in previous stages and the advantages of our model have been discussed. It is the second most common disease. It has 5 stages where in each stage person's health is deteriorated. The model we developed first loads the dataset and the splits the data into train and test data. Then trains the model with train data and then tests it to find the accuracy. It also gives a classification report. Later it takes an instance to detect if person has PD or not. The main advantages of our protect is it creates a real world situation to detect whether person has PD or not.

CHAPTER 2

LITERATURE REVIEW

2.1 OVERVIEW:

Before 2000's:

1) The first medical information was written in 1817 by James Parkinson, he published "An Essay on the Shaking Palsy" in which he described it as a neurological syndrome and identified the following symptoms tremors, abnormal posture and gait, paralysis, and diminished muscle strength, and the way that the disease progresses over time.

2) Charcot was the first to suggest the term- "PARKINSON'S DISEASE", earlier it was recognized just as 'paralysis agitans' or 'shaking palsy'. He also differentiated this disorder from other tremorous disorders especially multiple sclerosis in 1872 which is a major turnover and important part of this disease prediction.

3) before 2000's the three stages of parkinson disease that includes- unilateral disease , bilateral disease and development of postural reflex impairment are detected. Before the diagnosis for the parkinson was only restricted for the non-motor symptoms but in mid 20th century this disease is primarily diagnosed based on motor symptoms

In late 20th century:

1) In late 20th century PET(positron emission tomography) and spect(single-photon emission computed tomography), MRI(magnetic resonance imaging) were used to unerstand the brain changes of a person affected by PD.

2000's:

1) In 2000's machine learning models were introduced to detect PD that analyses voice recordings, patterns, brain imaging individually but they failed to analyze all at once. In 2009 ML model was applied on the data of vocal records and 93.84% was the accuracy.

2010's:

1)In 2010 DAS had made four classification schemes they are decision trees, regression, neural networks and DMneural. Bilal studied the genetic data was the first to introduce SUPPORTVECTOR MACHINE(SVM) to detect PD and the accuracy obtained was 0.889% but it is restricted to only a single symptom of the patient

2016 to recent years:

1) From 2016 to recent years several papers concluded that neural netwrok classifier gave the best outcome with an accuracy 92.9%. Raundale,Thosar,Rane trained random forest calssifier to predict PD but it is completely reliant on MATLAB.

2)Wang implemented 12 ML models and built a Deep learning model which gave 96.45% of accuracy but it is very expensive and it holds a large memory.

3)Ricciardi implemented decision trees, Random forests, KNN but this whole dataset is small and needs artificial data augumentation.
4)Over the past few several years many articles and papers concluded that several ML models trained and tested but most of them are not successfull.
5)SVM model intially in 2010 gave an accuracy of 88.9% but they considered only single symptom.From the available information and articles and researches by several authors the accuracy of SVM model in detecting PD is raised to 89% but they used a small data set i.e they included several symptoms of the patient but not all symptoms.
6)From previous studies we can conclude that initially before 2000's they ML models were not introduced and the PD is detected by some scanning techniques. In 2000's after ML models were introduced several techniques like decision trees , Neural networks ,regression ,DM neural, many more were used and got an higher accuracy in detecting but it was restricted only up to limited symptoms of the patient. In recent years SVM is used and gave an accuracy of 89% but it was trained only with small dataset. We are implementing an SVM model which gives and accuracy of 84% considering all the symptoms of a patient suffering from the disease and we considered a big amount of dataset.

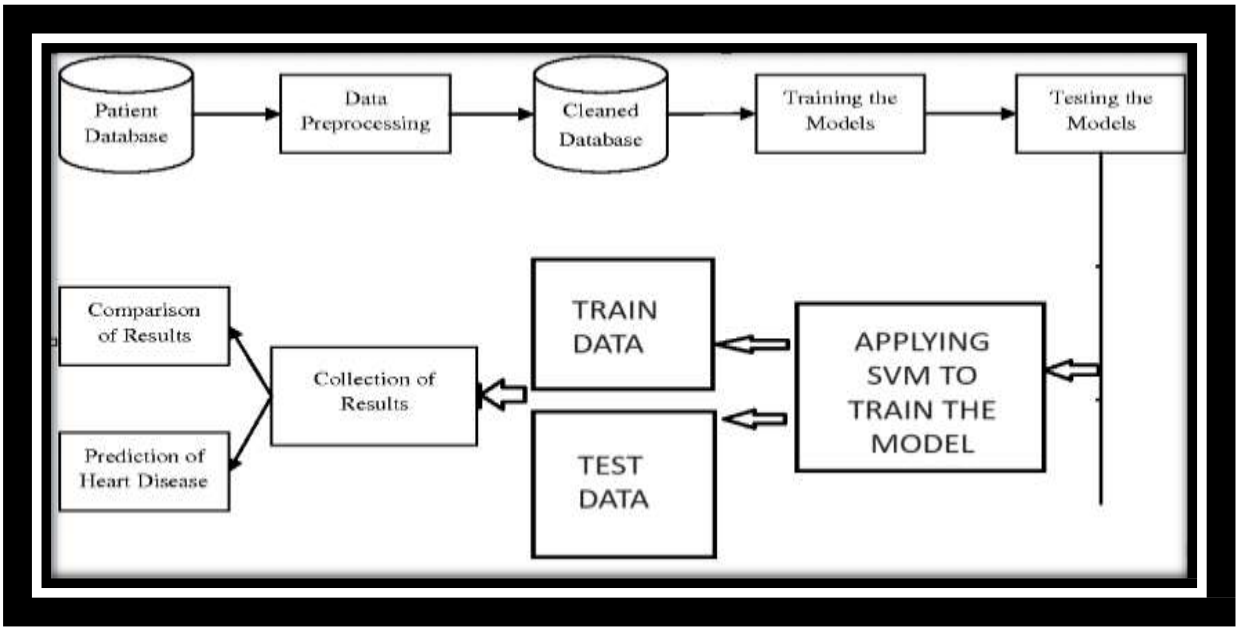

Fig: 2.1 : Design of the Model

2.2 REVIEW OF THE LITERATURE:

While the concept of predicting Parkinson's disease using machine learning algorithms has gained momentum in recent years, it's important to note that the historical data of this disease is looped with the broader understanding and advancements in Parkinson's disease research. Here is a brief historical overview:

Before 2000's-

1) The first medical information was written in 1817 by James Parkinson, he published "An Essay on the Shaking Palsy" in which he described it as a neurological syndrome and identified the following symptoms tremors, abnormal posture and gait, paralysis, and diminished muscle strength, and the way that the disease progresses over time.
2)Charcot was the first to suggest the term- "PARKINGSON'S DISEASE", earlier it was recognized just as 'paralysis agitans' or 'shaking palsy'.He also differentiated this disorder from other tremorous disorders especially multiple sclerosis in 1872 which is a major turnover and important part of this disease prediction.
3)An article was written by Hoehn and Yahr in 1967 from which 3 stages were discovered, stage 1- unilateral disease, stage 2- bilateral disease, and stage 3-the development of postural reflex impairment. stage 3 is the key turning point in the disease's clinical significance.
4)The diagnosis of PD is traditionally based on the non-motor symptoms (e.g: cognitive changes such as sleep disorders, and olfactory dysfunction) but these are the symptoms that are sometimes present before the onset of PD but they lack specificity therefore,non-motor symptoms don't allow the accurate prediction of the PD, so in early to mid-20th century Parkinson's disease is primarily clinically diagnosed based on motor symptoms.

LATE 20th CENTURY-

1)Advances in neuroimaging techniques, such as positron emission tomography(PET) single-photon emission computed tomography(SPECT), and magnetic resonance imaging(MRI) contribute to a better understanding of brain changes in Parkinson's disease.

IN 2000's:

1) Study by Marras et al. explores the usage of machine learning models that can analyze voice recordings to detect vocal characteristics and analyze the various data collected from patients' voice recordings, movement patterns, and brain imaging and detect subtle changes associated with Parkinson's. These technologies aid in early detection of the disease.
2) In 2009 Gil and Manuel gathered some vocal records and with the fewest vocal features necessary to diagnose Parkinson's 93.84% accuracy was obtained.

Advancements in Genetics (2010's):

1) Das compared several classification techniques to make an accurate Parkinson's disease diagnosis and he made four classification schemes they are decision trees, regression, neural networks, and DMneural.
2) Bilal et.al. studied the genetic data of several patients to predict the Parkinson's disease of several senior patients with a support vector machine model. They trained the model of the support vector machine with an accuracy of 0.889.

FROM 2016 to recent years-

1) Al-Fatlawai after studying several papers concluded that The neural network classifier gave the best outcome, its performance is 92.9%. The DBN(deep belief network) is used as a successful method to identify Parkinson's disease, it is employed to categorize Parkinson's illness.

2) Raundale , Thosar, and Rane used the data from the UCI telemonitoring dataset to train a Random forest classifier to predict the Parkinson disease in patients whose age range of 46 to 85 years and also used an audio data set of several patients but they are completely reliant on MATLAB

3) The majority of the articles and research done focus on the usage of deep learning in PD detection. They used the present deep learning models to train the dataset and phonation data to predict the present condition and progress of the disease but their work lacked the use of feature selection that would improve the deep learning model performances which increases the accuracy.

4) Wang et. al. implemented 12 machine learning models on the 401 voice dataset and they built a

deep learning model(DEEP) with an accuracy of 96.45% but the model is very expensive and it holds a large memory.

5) Ricciardi et. al performed a spatial-temporal analysis of brain MRI scans. They also implemented decision trees, random forests, and KNN to detect Mild Cognitive Impairment(MCI) but this whole dataset is small and this also needs artificial data augmentation.

Over the past few years, scientists such as Marras, Espay, Hssayeni, Roy, and others have played a major role in advancing our understanding of the prediction of Parkinson's disease.

Year	Event
1817	James Parkinson publishes "An Essay on the Shaking Palsy," describing Parkinson's symptoms.
1872	Charcot suggests the term "Parkinson's Disease."
1967	Hoehn and Yahr propose a staging system for Parkinson's.
Late 20th century	Advances in neuroimaging techniques contribute to understanding brain changes.
2000	Study by Marras et al. explores machine learning for early detection using voice recordings
2009	Gil and Manuel achieve 93.84% accuracy in diagnosing Parkinson's using vocal features
2010	Das compares classification techniques for accurate diagnosis (genetic

	data). Bilal et al. predict Parkinson's using a support vector machine with 88.9% accuracy.
2016	Al-Fatlawai finds neural network classifiers yield 92.9% accuracy. Raundale, Thosar, and Rane use Random Forest for PD prediction, reliant on MATLAB. Focus on deep learning in PD detection, lacking feature selection (2016 onward). Wang et al. implement 12 ML models on a voice dataset, achieving 96.45% accuracy. Ricciardi et al. analyze brain MRI scans with decision trees, random forests, and KNN for MCI detection, employing artificial data augmentation.

Table 2.1: History of PD

2.3 PROBLEM STATEMENT:

The problem on which we are working is the detection of Parkinson's at the early possible stage so that treatment can be started prior so that the patients procrastinated with the disease will be able to live longer. The problem at hand is marked one of the recent growing diseases. The solution that we come up with should be accurate along with being scalable. The main technology to be used for the prediction should be ML algorithms for early diagnosis as a curing at the further stage will become challenging.

2.4 PROJECT OBJECTIVES:

The primary objective of this project is to develop a ML model which predicts with high accuracy and which is feasible. This model is developed in such a way that it predicts PD at an early stage to prevent deaths. The timely detection is crucial as treatment can be started at an early stage to improve quality of life. The key components of the project are early detection, accuracy, scalability, ML algorithms, timely intervention, patient impact etc.

2.5 SUMMARY:

PD is neurodegenerative disease which destroys the dopamine cells and causing the person to depend on others for doing their daily activities. This project aims to develop a ML model to detect PD at a lower cost with higher accuracy. This model helps to detect PD at an early stage to prevent further growth of disease. The treatment methods available are at a higher cost with lower chance of cure at a final stage. If predicted in advance it can be treated with available resources. So PD prediction at an early stage is the main goal of this project. This can be secured using ML algorithm SVM which creates a hyperplane to classify data. It can also be done using other ML algorithms but SVM is more reliable.

CHAPTER 3

DEVELOPMENT OF THE PROJECT

3.1 METHODOLOGY OF THE PROJECT:

The methodology includes steps like Data collection, Data preprocessing, feature selection etc. Firstly, Data collection. It is an important step. We should acquire a comprehensive dataset which contains relevant features. To perform EDA, data should be preprocessed to understand dataset's features. Handling missing data, incorrect data and duplicate data is important step. To identify relevant attributes, feature selection is required. SVM is effective compared to other algorithms so ML model selection plays an important role. Model training and evaluation are important to know the accuracy of the ML model. Prediction on new instances simulates real-world scenarios. Next documentation and reporting is the last step. The methodology follows a systematic approach to develop an accurate and scalable ML model for detecting PD at an early stage.

Drug	Advantages	Disadvantages
Levodopa (L-dopa) + dopa decarboxylase inhibitor	• Probably the most potent dopaminergic drug for symptom relief • Generally well tolerated	• Motor complications (cumulative risk 10% per annum)
Catechol-O-methyl transferase inhibitors, for example, entacapone, tolcapone	• Increase levodopa half-life • Reduce 'off' time	• Tolcapone can cause liver damage. • Diarrhoea
Ergot dopamine agonists (for example, bromocriptine, pergolide, cabergoline Non-ergot dopamine agonists for example, pramipexole, ropinirole, rotigitine	• Good efficacy • Delay onset of motor complications • Generally well tolerated • Once-a-day preparations available with some • Transdermal patch for rotigitine • Theoretical neuroprotective action • Some antidepressant action with pramipexole	• Increased risk of somnolence, confusion, hallucinations, peripheral oedema and behavioural changes • Cardiac valve fibrosis with ergot drugs
Monoamine oxidase B inhibitor; selegiline; rasagiline	• Improve motor features in early and late disease • Easy to use, once-a-day • Well tolerated • Theoretical neuroprotective effect	• Relatively mild efficacy • Selegiline metabolized to amphetamines — potential cognitive effects
Amantadine	• Mild anti-Parkinsonian effect • Improves dyskinesias	• Cognitive disturbances • Peripheral oedema • Livedo reticularis

Table 3.1: Current drugs for Parkinsons Disease

3.2 DEVELOPMENT OF THE PROJECT:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import classification_report,accuracy_score
```

```
file_path = "C:\\Users\\khyat\\OneDrive\\Desktop\\parkinsons
dataset\\dataset\\parkinsons.csv"

p_data = pd.read_csv(file_path)

a = p_data.drop(columns=['name'])
b = p_data['status']

a_train, a_test, b_train, b_test = train_test_split(a, b, test_size=0.2, random_state=42)

classifier = SVC()

classifier.fit(a_train, b_train)

b_pred = classifier.predict(a_test)

print("Classification Report for dataset is :")
print(classification_report(b_test, b_pred))

accuracy = accuracy_score(b_test, b_pred)
print("Accuracy in decimals is :", accuracy)
print("Accuracy percentage is=", accuracy*100)

new_instance = pd.DataFrame([[119.992, 157.302, 74.997, 0.00784, 0.00007, 0.0037,
0.00554, 0.01109, 0.04374, 0.426, 0.02182, 0.0313, 0.02971, 0.06545, 0.02211, 21.033, 1,
0.414783, 0.815285, -4.813031, 0.266482, 2.301442, 0.284654]])

new_instance = new_instance.astype(float)

X = p_data.drop('name', axis=1)

predict = classifier.predict(new_instance)

if predict[0] == 1:
    print("It is better to consult the doctor because the person may have the Parkinsons disease
which is neurodegenerative disease and second most common disease like Alziemers")

else:
    print("The person is prognosticated not to have Parkinson's complaint.")
    print("The person may have some symptoms because of the age or some other factors")
```

3.3 RESULTS:

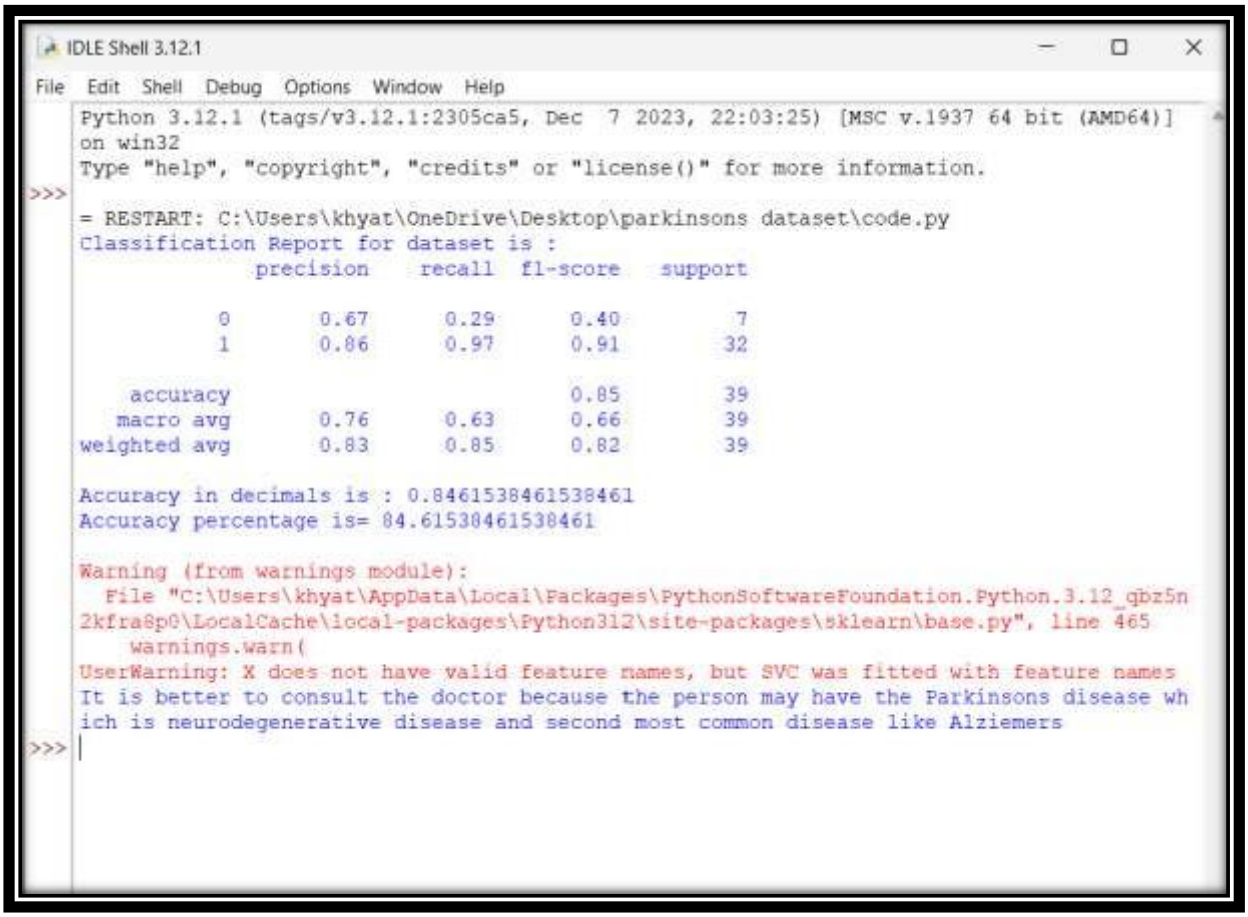

Fig 3.1: Result

CHAPTER 4

CONCLUSIONS

4.1 CONCLUSIONS:

In conclusion, early detection of this neurodegenerative disease prediction is a challenging task and significant for improving patients' health. The main reason for PD is lack of dopamine producing cells. The project objective is revolved around accuracy, scalability. SVM which creates a hyperplane plays a major role for classification. It uses python which is an object-oriented programming language which has libraries like pandas, scikit-learn. Collaboration with healthcare can create awareness to people. The journey towards early detection can improve the quality life of person procrastinated with PD.

4.2 RECOMMENDATIONS:

The recommendations are:

1. Collaboration with Healthcare Professionals

2. Continuous Data Monitoring and Updating

3. Incorporate Multiple Modalities

4. Integration with Electronic Health Records(EHR)

5. User-Friendly Interface for healthcare Professionals

6. Explainability and Interpretability

7. Cross-Validation across diverse populations

8. validation through clinical trials

9. public awareness and education

10. long-term monitoring and Follow-up

REFERENCES

- https://www.analyticsvidhya.com/blog/2021/07/parkinson-disease-onset-detection-using-machine-learning/

- https://ieeexplore.ieee.org/document/9752925

- https://www.frontiersin.org/articles/10.3389/fnagi.2021.633752/full

- Abiyev, R. H., and Abizade, S. (2016). Diagnosing Parkinson's diseases using fuzzy neural system. Comput. Mathe. Methods Med. 2016:1267919. doi: 10.1155/2016/1267919

  PubMed Abstract | CrossRef Full Text | Google Scholar

- Abos, A., Baggio, H. C., Segura, B., Campabadal, A., Uribe, C., Giraldo, D. M., et al. (2019). Differentiation of multiple system atrophy from Parkinson's disease by structural connectivity derived from probabilistic tractography. Sci. Rep. 9:16488. doi: 10.1038/s41598-019-52829-8

  CrossRef Full Text | Google Scholar

- Abujrida, H., Agu, E., and Pahlavan, K. (2017). "Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data," in 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT) (Bethesda, MD), 208–211. doi: 10.1109/HIC.2017.8227621

  CrossRef Full Text | Google Scholar

- Adams, W. R. (2017). High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. PLoS ONE 12:e0188226. doi: 10.1371/journal.pone.0188226

  PubMed Abstract | CrossRef Full Text | Google Scholar

- Adeli, E., Shi, F., An, L., Wee, C.-Y., Wu, G., Wang, T., et al. (2016). Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. NeuroImage 141, 206–219. doi: 10.1016/j.neuroimage.2016.05.054

  PubMed Abstract | CrossRef Full Text | Google Scholar

- Adeli, E., Thung, K.-H., An, L., Wu, G., Shi, F., Wang, T., et al. (2019). Semi-supervised discriminative classification robust to sample-outliers and feature-noises. IEEE Trans. Pattern Anal. Mach. Intell. 41, 515–522. doi: 10.1109/TPAMI.2018.2794470

  PubMed Abstract | CrossRef Full Text | Google Scholar

- Ahlrichs, C., and Lawo, M. (2013). Parkinson's disease motor symptoms in machine learning: a review. arXiv preprint arXiv:1312.3825. doi: 10.5121/hiij.2013.2401

  CrossRef Full Text | Google Scholar