A SUPERVISED MACHINE LEARNING APPROACH TO CLASSIFY PARKINSON DISEASE

BY

R.I.M. ABDULLAH AL FARABI
ID: 182-15-11729

AND

MST. SONIATUZ ZHOHRA
ID: 182-15-11451

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Md. Sadekur Rahman

Assistant Professor

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER 2022

APPROVAL

This Project/internship titled **"Predicting Password Strength Based on Natural Language Processing Technique"**, submitted by **Ahsan kabir**, ID No: 161-15-7301 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *September 13,2022*.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Nazmun Nessa Moon (NNM)
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mr. Faisal Imran (FI)
Assistant professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

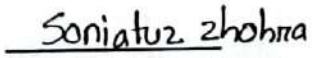
We hereby declare that, this project has been done by us under the supervision of **Md. Sadekur Rahman, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

R.I.M. Abdullah Al Farabi
ID: 182-15-11729
Department of CSE
Daffodil International University



Mst. Soniatuz Zhohra
ID: 182-15-11451
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Supervisor Md. Sadekur Rahman**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude **Md. Sadekur Rahman and Dr. Touhid Bhuiyan Sir, Head,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Parkinson's disease is a degenerative brain disease. In most cases, over the age of 50 people are affected by this disease, but nowadays, young people are also affected by this disease due to genetic reasons. Due to reduced brain function in patients suffering from this disease, the patient suffers from imbalance along with difficulty in walking and speaking. Our present research work was done to assess classification algorithms in machine learning language. Here we identify Parkinson's disease from the data set that this disease is Parkinson's disease YES or No. This research work also focused on the medical field for the confirmation of clinical trials and identifying Parkinson's disease.

In our study, we used six types of machine learning classification analysis algorithms (LR, KNM, DT, RF, XGB, SVM) to identify PD datasets. From the Logistic Regression classifiers algorithm, we gain 76% accuracy from the accuracy calculation. We archive 77% accuracy from KNM, from SVM we achieve 77% accuracy, from DT we get 73% accuracy, from XGB we found 78 % accuracy. But we archive the highest accuracy 81% from Random Forest (RF) in our research study where the best accuracy was found to identify Parkinson's disease. In the future, using this model in the medical field will help to classify Parkinson's disease more easily.

Keywords—Parkinson’s Disease, Machine Learning, Identify, Classifications, Analysis, Logistic Regression, K Neighbors, Decision Tree, Random Forest, XGB, SVM, Impact.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Parkinson's disease is a neurodegenerative disorder that affects approximately 5 million people worldwide. It usually occurs in people over the age of 50, but is now affecting younger people as well. Because PD is usually classified as a movement disorder. Parkinson's disease area unit is the foremost essential reason for death and disability worldwide in keeping with the paralysis against foundation, the affected peoples worldwide by Parkinson's syndrome disease is projected that one million folks Living by 2020 within the USA [1].

PD is a brain disease. which can cause uncontrolled movements, such as tremors, difficulty walking, finger curling, and imbalance. Symptoms of this disease usually start slowly and worsen over time. Parkinson's disease is a chronic neurological disorder. We all know that the number of patients with this disease is increasing day by day in Asian countries. But most people don't seem to be aware of this illness. This disease is very dangerous for individuals. However, they treat this disease once it reaches the last stage. According to new World Health Organization (WHO) data published in 2020, the number of Parkinson's disease deaths in Asian countries has reached 3,782.

There is a reasonable health information processing system to identify patients with Parkinson's syndrome. This method additionally serves as a simulation to reduce the employment of physicians. The diagnosis of Parkinson's syndrome is supported by neuropathology and histopathology [2] [3].

Medical Diagnostic Identification Board of Parkinson's Disease Impact of the trademark Parkinson's syndrome illness highlights and will support quality. In this approach, Parkinson's disease is predicted by analyzing clinical, pathologic, and medical studies based on the incidence, characteristics, and regression of risk factors of the test. [4][5]. Parkinson's syndrome typically affects an outsized part of worldwide patients over the age of fifty, which has affected up to currently [6]. Still, currently, there's no acknowledged reason for paralysis against, however, it's terribly doubtless doable to assuage symptoms knowingly within the early stage of the subjective patients [7]. Approximate ninetieth of

the patients affected with voiced injury a study appealed this [8]. In general, the treatment of Parkinson's syndrome is likely to be very expensive. That is why most patients cannot afford the cost of anti-paralysis.

Nowadays, paralysis against prediction is the most crucial matter for clinical practitioners to require accurate choices concerning such a malady. It's a good exercise at present time, machine learning-based in-depth platform will detect paralysis against. Medical information has matured a massive scale volume from totally different clinical areas as well as health care services. To handle this information and attain insights from this information there's a requirement for a giant information analysis through.

As Parkinson's disease is a degenerative brain disease and mostly, older people are affected by this disease, and nowadays, due to genetic reasons, young people are also affected by this disease. Due to reduced brain function in patients with this disease, the patient suffers from difficulty in walking and speaking as well as imbalance.

Parkinson's syndrome treatment is probably going very expensive. This causes most of the patients cannot manage the fetched of loss of motion against. Nowadays, paralysis against prediction is the most crucial matter for clinical practitioners to require accurate choices concerning such a malady. It's a good exercise at present time, machine learning-based in-depth platform will detect paralysis against. Medical information has matured a massive scale volume from totally different clinical areas as well as health care services. To handle this information and attain insights from this information there's a requirement for a giant information analysis through.

In our study, we used six types of machine learning classification analysis algorithms (LR, KNM, DT, RF, XGB, SVM) to detect PD. As if this disease can be easily identified and quickly treated, people can lead a normal life again.

1.2 Motivation

The prevalence of PD patients is increasing in the present world. Although it is a very common disease, we see that many people do not recognize this disease easily and get more affected if not treated at the right time.

Nowadays people are becoming increasingly dependent on technological solutions. Our main work is to develop a system that can early predict Parkinson's Disease among humans as well as assist the Doctor to make decisions. So we use machine learning by developing

a system that can detect Parkinson's disease from datasets and initiate treatment quickly and help doctors make their decisions at the medical field.

In today's technology world, People are becoming increasingly reliant on technical solutions. Bangladesh is currently adopting new technologies and its citizens are found of Smart salutation to access for good treatment.

1.3 Rational Questions

To develop a research study, we would need to go over a research question, which is a concern that a study or research project intends to resolve. This question typically answers a problem addressed in the review' s judgment through the analysis and translation of data. As a result, below are a few fundamental research questions that are the focus of this research.

- What are the features used for identifying Parkinson disease?
- What are the specific Parkinson’s Class?
- How to collect data for Parkinson disease prediction?
- Which state of the art classifier perform best in classifying Parkinson disease?

1.4 Research Output

When we go outside and through social media, we can see that there are lots of patients who are affected by this disease but they actually don't know about this. As the days go we can see the percentage of this disease is going up although it's old.

In our research paper, we apply classifier algorithms to detect Parkinson’s disease by using machine learning language which results generated by a decision. From the dataset table, we know that this disease is Parkinson’s disease. We also find that it can help in the medical field.

1.5 Expected Output

This examination-based projects consolidate with sections. Every part addresses an alternate perspective. Every one of part integrates an assortment of explicit sub-areas that are displayed in an unmistakable and succinct way. In our report, we have constructed our contents into chapters. Using machine learning model to classify label class and deploy the model into e-commerce Business website or can use all kind of people free we have several

advantages. The proposed model can be integrated into an e-commerce website to classify the patient and provide the best service based on illness labels.

1.6 Project Management and Finance

When we began our work, we are totally unknown about all the process because it`s our first time. So, we arrange meeting and communicate with traders through social media such as Facebook, Messenger, WhatsApp, etc. We have decided to do this research after discussing it with them. We have to collect the information from online platform after lots of searching. And run to the method we have used Collab. So, we have not expensed any cost for this research.

1.7 Report Layout

This examination-based projects consolidate with six sections. Every part addresses an alternate perspective. Every one of part integrates an assortment of explicit sub-areas that are displayed in an unmistakable and succinct way. In our project report, we have constructed our contents in chapters such as.

Chapter 1: We have alluded to the 1.1 Introduction, 1.2 Motivation, 1.3 Rational Questions, 1.4 Research Output, 1.5 Expected Outcome 1.6 Project Management, and finance and 1.7 Report Layout.

Chapter 2: 2.1Preliminaries/Terminologies, 2.2 Related Works, 2.3 Comparative Analysis and Summary 2.4 Scope of the Problem and 2.5 Challenges.

Chapter 3: 3.1 Research Subject and Instrumentation, 3.2 Data Collection Procedure/Dataset Utilized 3.3 Statistical Analysis, 3.4 Proposed Methodology/Applied Mechanism and 3.5 Implementation Requirements.

Chapter 4: 4.1 Experimental Setup ,4.2 Experimental Results & Analysis and 4.3 Discussion.

Chapter 5: 5.1 Impact on Society, 5.2 Impact on Environment,5.3 Ethical Aspects and 5.4 Sustainability Plan.

Chapter 6: 6.1 Summary of the Study, 6.2 Conclusions and 6.3 Implication for Furth

CHAPTER 2

Background

2.1 Preliminaries/Terminologies

Parkinson's disease is a degenerative brain disease. In most cases, over the age of 50 people are affected by this disease, but nowadays, young people are also affected by this disease due to genetic reasons.

2.1.1 Parkinson

Parkinson's Disease is a neurodegenerative clutter that influences predominately the dopamine-producing (“dopaminergic”) neurons in a particular region of the brain called the substantia nigari. It’s best known for causing moderated developments, tremors, adjusted issues and more. Most cases happen for obscure reasons, but a few are acquired.

2.1.2Machine Learning

The brain affects the CNS and there's no treatment for Parkinson's disease until it is identified early. It affects the CNS of the brain. Delay in detection and timely treatment can lead to death. So early detection is very important. For the early discovery of this disease, we used machine learning algorithms such as XGB, and Random Forest.

2.1.3 Diagnosis

Age is the foremost strong hazard for PD [9, 10] with an ordinary age of onset of generally 50 to 60 a long time. Two other danger components have shown up to be basic: family history (a innate interface) and pesticide introduction. Additional chance factors have been recognized in show disdain toward the truth that how they may differentially impact men vs women is still unclear [11]. Numerous other risk factors have been proposed in spite of the fact that the epidemiologic confirmation isn't as solid. These consolidate: Utilize of well water, depletion utilization, plenitude body weight, presentation to hydrocarbon solvents, living in common ranges, developing or agrarian work, living in urban districts

or industrialized ranges with presentation to copper, manganese and lead, tall dietary affirmations of press, history of press lack and higher levels of instruction [12].

In show disdain toward of decades of request, the conclusion and organization of Parkinson's disease are hampered by flawed procedures for revelation and speculating. In other words, endorsed biomarkers (tests or screening components) with tall affectability and specificity for the illness are essentially required but are right presently missing. This deficiency constitutes a major investigation reroute since the clinical trial arrangement demands a target or biomarker to test neuroprotective medications [13]. In expansion, no single marker is by and by able to anticipate the PD movement with extraordinary faithful quality and authenticity [14].

Highlights that increase the likelihood of Parkinson's disease conclusion join those related to bradykinesia, such as micrographic, a modifying walk, and challenges performing motor errands such as turning in bed, rising from a chair and controlling objects. Then again, other side effects diminish the probability of PD counting falling early within the illness, symmetric tremor at the start, fast malady movement, small reaction to dopamine treatment, etc. [15].

Interestingly a prodromal of non-motor highlights may go before engine indications of PD by numerous a long time. These incorporate stoppages, hyposmia (modified sense of scent), REM rest clutter, orthostatic hypotension, sadness, encourage urinary incontinence, and erectile brokenness [16]. Since no biomarkers exist for PD, neuroprotective masters (on the off chance that they were open) cannot be utilized to expect empowering neurodegeneration.

2.1.4 Differential Diagnosis

The differential conclusion is challenging given the reality that the classic PD signs (e.g., rest tremor, rigid nature etc.) can be show in other neurodegenerative disarranges. Cautious history taking and intelligent physical assessment coupled with introductory therapeutic treatment (e.g., the individual's response to pharmacotherapy) are principal to recognize idiopathic PD from Fundamental Tremor, DLB, CBD, MSA, PSP, or assistant Parkinsonism due to drugs, harms and head damage [17,18].

2.2 Related Works

Parkinson's disease Researchers for symptom detection The initial symptoms are variously mentioned. Indira et al [19] ...as an early symptom of PD The voice of patients use Other authors have analyzed mechanistically whether it is affected by metal content Creates a model to explain.

The model requires Fuzzy C-Means (FCM) to use their datasets and of a pattern is needed. For sensitivity they found 68.04%, 75.34% accuracy and 45.83% for metallic money stability.

Also Amit et al. [20] adopted a method to bring such patients into a single class based on their physical handicap. He uses L2 as the standard metric with vector machines to implement this method. In [21], the authors applied this method to odor identification from a dataset (UPSIT-40). They can even do 16-item identification. This study was conducted on a Brazilian population.

They used logistic regression to identify the disease. Use the model. They validated 89 sniffing sticks and found a sensitivity of 81%, they found that the UPSIT-40 had a specificity of 83-3.5% and a sensitivity of 3.5%. and 82.1%. [22] also identified sleep behavior disorders from the 40-item UPSIT with Locative Loss Feature Loss and Sleep Eye Movement Sleep Behavior Disorder Screening Form (RBDSQ).

Also they achieved 90.40% and 85.48% accuracy for sensitivity using vector machine and classification tree techniques respectively. [23]

Researchers use machine learning methods to diagnose Parkinson's disease. where from the dataset through the use of machine learning algorithms Parkinson's disease Identify symptoms (Sriram, Rao, Narayan, Kaladhar, and Vital, 2013).

They presented their work with a comparative result. Where they as algorithms Bayes, Random Forest, Logistic Regression, Support Vector models are used. They are the result SVM got 88.9% and NB got 69.23%. But the maximum value obtained from RF is 90.26%. Which helps in accurate diagnosis compared to other models.

We used machine learning to accurately diagnose Parkinson's disease in our study.They have increased our interest in this paper. Our current paper uses machine learning to accurately diagnose Parkinson’s disease.

2.3 Comparative Analysis and Summary

Parkinson's disease is the first common neurodegenerative disease. The disease is anticipated to be twofold within another 30 long times. The exact conclusion of Parkinson's disease is still a challenging assignment and the early stages of the infection are continuous. Later advancements in the past 5 long time incorporate the approval of clinical symptomatic criteria, the presentation of inquiries about criteria for prodromal Parkinson's infection, the distinguishing proof of testing and hereditary subtypes, and the expanding number of hereditary variations related to Parkinson's malady hazard.

2.4 Scope of the Problem

Parkinson's disease is a neurological disease that causes inattentive or uncontrollable movements, like shaking, stiffness, and problems with balance and coordination. Symptoms sometimes begin step by step and worsen over time. because the illness progresses and could have problems walking and talking. Parkinson’s disease is also a progressive neurodegenerative condition resulting in death of the dopamine-containing cells of the nucleus Niger. The 'cardinal signs' of the malady area unit rest in tremor, rigidity, and hyperkinesia. Bodily property instability and falls occur later throughout the course of the condition.

2.5 Challenges

The major challenge, indeed for developmental disorder specialists, is the early symptomatic separation of Parkinson’s disease from atypical parkinsonian disorders. The term atypical parkinsonism is an umbrella term for an assortment of neurodegenerative clutters in which parkinsonian disorder may be a conspicuous clinical highlight, but the complete clinical range, basic pathology, movement, and forecast are on a very basic level diverse from those of Parkinson's disease [30,31]. Clinical-pathological studies have revealed error rates in the clinical assignment of patients with these different syndromes ranging from 7–35% of the cases. [32,33].

Biomechanical sensors such as accelerometers, gyroscopes, and magnetometers are well suited for the location of tremors, bradykinesia, stride impairment, and engine complications, such as dyskinesia. However, data collected in the home and community settings using these sensors do not always provide sufficient information to achieve a reliable clinical assessment of motor symptoms. For instance, it is troublesome to induce from the sensor information alone in case the gradualness of movement (as recognized utilizing biomechanical sensors) can be utilized as an intermediary of bradykinesia or is the result of weariness or other variables related to the setting in which an engine the errand is performed (eg, moderate strolling since of fear of falling). Too, the determination of biomechanical sensors is confined to the anatomical zone to which they are connected, which may surrender low quantitative understanding with the more extensive range of motor inability, quality of life, and other quantifiable patient-relevant endpoints [34, 35].

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

We inquire about subject is Parkinson disease. First of all, we have to know about this disease. What is Parkinson disease? Parkinson's disease is a brain disorder that causes automatic or uncontrolled movements, such as tremors, stiffness, and trouble with adjusting and coordination. Symptoms more often than not begin gradually and get more awful over time. As the illness advances, individuals may have trouble strolling and talking. To do this work and make it successful we have to take help from some instrument. We can divide those instruments into three-part Hardware, Software and Development Tools. In the Hardware part we use a pc with configuration Intel core i5 8th Generation (Installed 4GB RAM) and one TB hard disk. In the software part we use Google Colab, GPU, Weka and MS PowerPoint. In part we use development tools here we use windows 10, Python3.9, Matplotlib, NumPy and Pandas. Without use of those tools we are unable to finish our work.

3.2 Data Collection Procedure/Dataset Utilized

When it comes to developing a machine learning model, data is the most important thing. Since machine learning with data, accurate and good amounts of data can further enhance the performance of the model correct classification.

Table 3.1: Dataset Collection

Attribute	Full Form	Data Type	Value/ Range
GP	Gender of Patient	Nominal	Male, Female
PPE	Pitch Period Entropy	Numerical	0.04-0.90
DFA	Detrended Fluctuation Analysis	Numerical	0.54-0.85
RPDE	Recurrence period density entropy	Numerical	O.15-0.87
NP	Number Pulse	Numerical	2-907
NPP	NumPeriodsPulse	Numerical	1-905
MPP	MeanPeriodPulses	Numerical	0.00-0.01
SDPP	StdDevPeriodPulses	Numerical	0.00-0.00
LPJ	LocPctJitter	Numerical	0.00-0.02
LAJ	LocAbsJitter	Numerical	0.00-0.00
RJ	RapJitter	Numerical	0.00-0.01
PQJ	Ppq5Jitter	Numerical	0.00-0.01
DPJ	DdpJitter	Numerical	0.00-0.03
LS	LocShimmer	Numerical	0.00-0.25
LDS	LocDbShimmer	Numerical	0.05-2.11
AQS	Apq3Shimmer	Numerical	0.00-0.13
APQS	Apq5Shimmer	Numerical	0.00-0.19
APQSM	Apq11Shimmer	Numerical	0.00-0.27
DDS	DdaShimmer	Numerical	0.00-0.39
MACH	MeanAutoCorrHarmonicity	Numerical	0.58-0.99
MWHH	MeanNoiseToHarmHarmonicity	Numerical	0.00-0.76
MHTWH	MeanHarmToNoiseHarmonicity	Numerical	1.65-33.19

3.3 Statistical Analysis

We take this data set from an online platform. In this data set we found seven hundred and fifty-seven data. In this data we found two types of data Numerical and Nominal. Numerical data can take 2 different forms, namely; discrete data, which represents countable items and continuous data, which represents data measurement. Nominal data are used to label variables without any quantitative value.

3.4 Proposed Methodology/Applied Mechanism

We apply Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), and Support Vector Machine (SVM). Then, using the training dataset, we tested our model. The training and validation accuracy graph of our project is the best. Ultimately, we identify that us develop Random Forest Classifier (RFC)

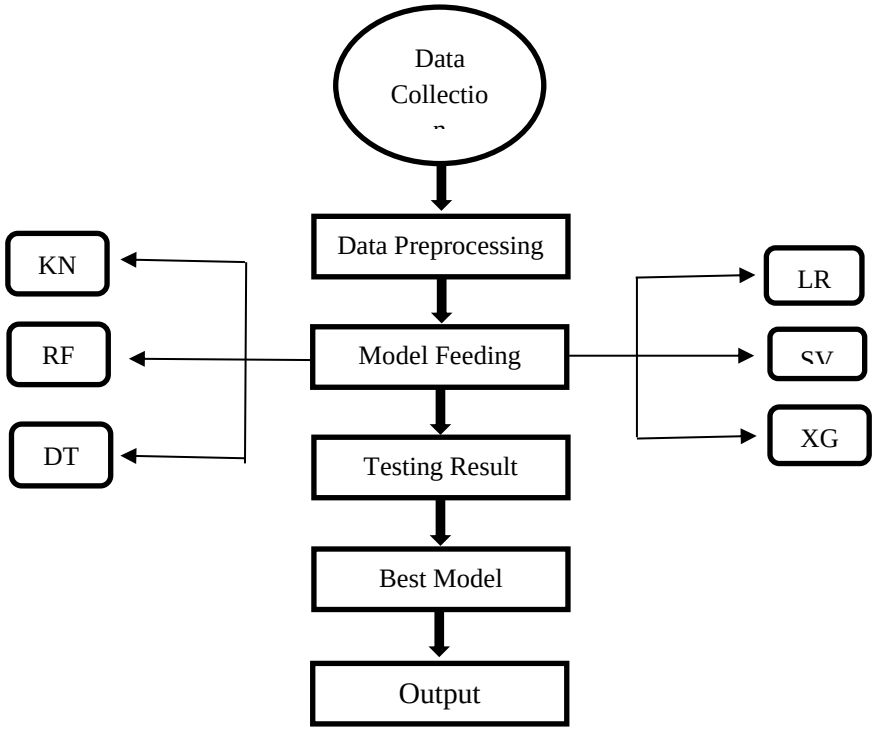


Figure 3.1: Proposed Methodology

3.5 Implementation Requirements

In order to conduct our research, we have used the following tools and technologies:

1. Google colab
2. Python verson 3.9
3. Windows 10
4. Pandas
5. NumPy
6. MS PowerPoint etc

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup

We came to the taking overview and quickly developed a model to recognize them. Our proposed framework is built using AI computation. Working in AI area will be really difficult as it requires a high-setup PC, GPU and other devices. Next is a summary of the tools and innovations expected to enhance and drive our model.

Software and Hardware:

- A PC with configuration: Intel Core i5 8th generation (Installed 4GB RAM)
- 1TB HDD
- MS PowerPoint
- Google Colab and GPU
- Weka

Development Tools:

- Windows 10
- Python 3.9
- Pandas
- Matplotlib
- NumPy
- Seaborn

4.2 Experimental Results & Analysis

Best Accuracy

Here is the best accuracy graph of classification for Parkinson's disease.

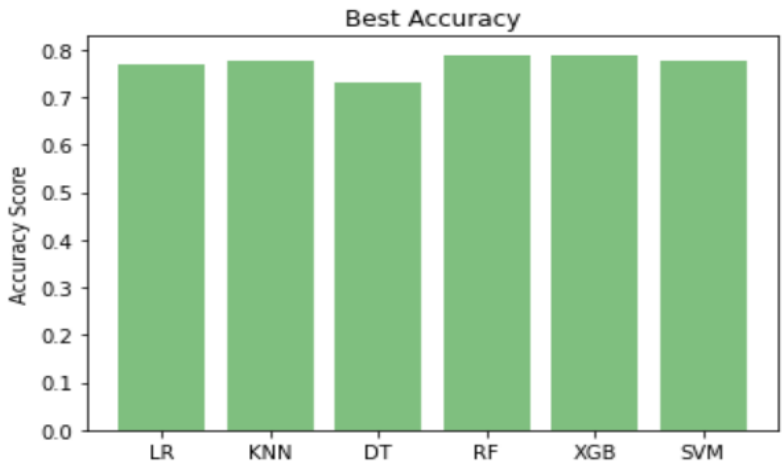


Figure 4.1: Best Accuracy

Given below the table of classification report of six algorithms which are LR, KNM, DT, RF, XGB, SVM

Table 4.1: Classification Report

Model	TP	FN	FP	TN
LR	5	29	6	112
KNN	12	22	12	106
DT	15	19	23	95
RF	26	21	8	110
XGB	15	19	13	105
SVM	1	33	1	117

Given below the table 4.2 is a performance table where we calculate accuracy, TPR, FNR, FPR, TNR, PRECISIO, FI SCORE

Table 4.2: Performance Table

ACCURACY	TPR	FNR	FPR	TNR	PRECISION	F1 SCORE
76.97368	14.70588	85.29412	5.084746	94.91525	45.45455	22.22222
77.63158	35.29412	64.70588	10.16949	89.83051	50	41.37931
72.36842	44.11765	55.88235	19.49153	80.50847	39.47368	41.66667
89.47368	55.31915	44.68085	6.779661	93.22034	76.47059	64.19753
78.94737	44.11765	55.88235	11.01695	88.98305	53.57143	48.3871
77.63158	2.941176	97.05882	0.847458	99.15254	50	5.555556

Correlation

Here we found Correlation about Parkinson's Disease where No Missing Values of Parkinson's Data set.

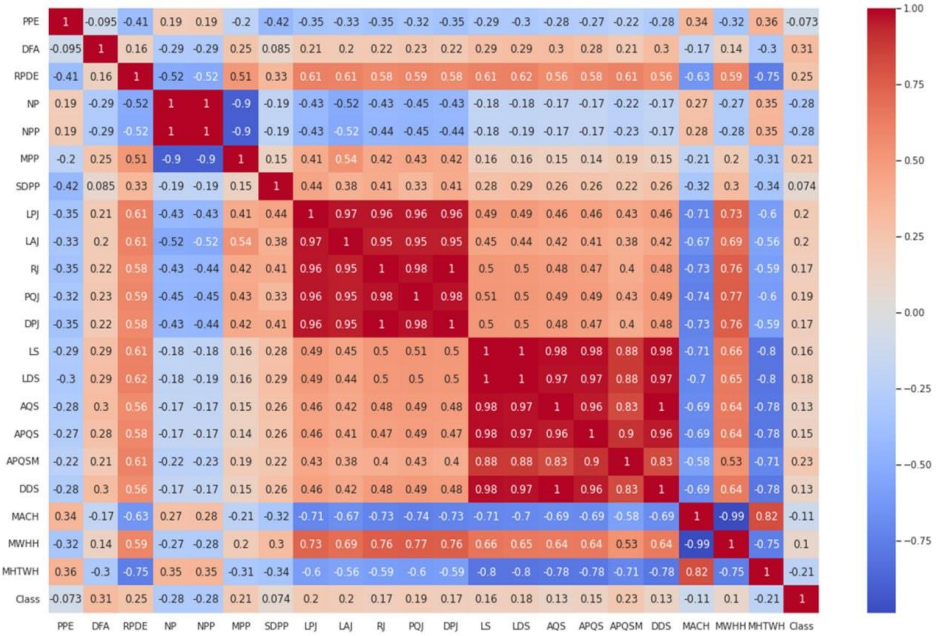


Figure 4.2: Correlation

Confusion matrix

A confusion matrix is a tractor for detruncating the performance of a classification algorithm.

- **LR confusion matrix experiment**

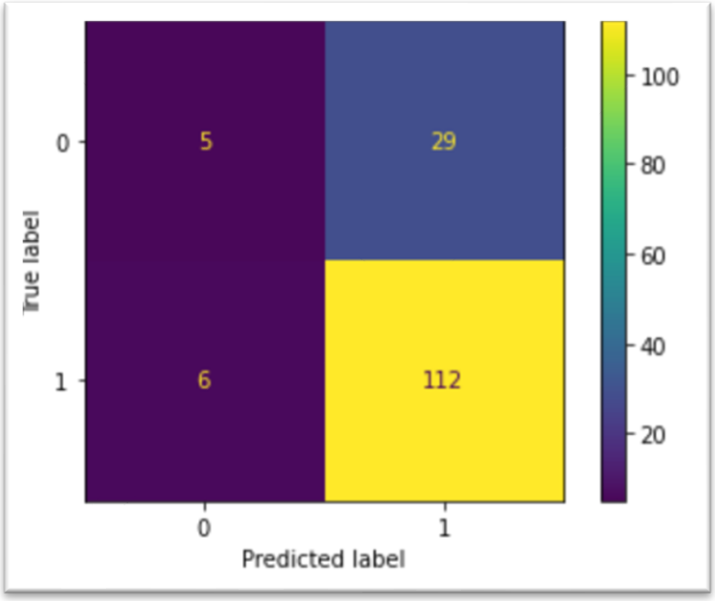


Figure 4.3: Logistic Regression confusion matrix

- **KNM Confusion matrix**

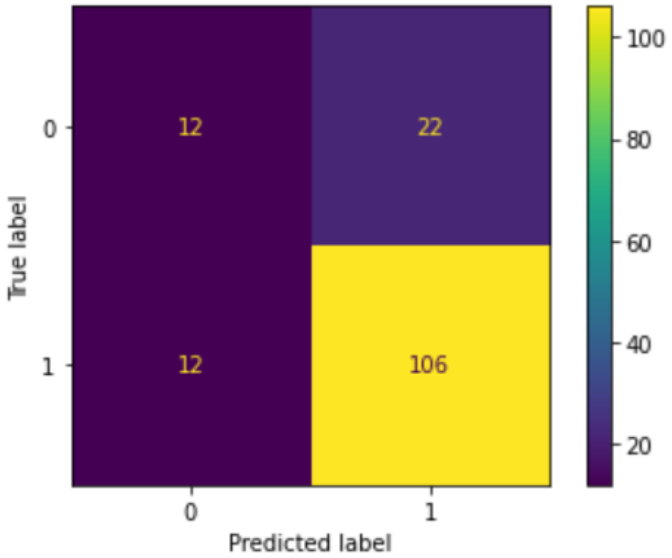


Figure 4.4: K-Nearest Neighbors

- **DT Confusion Matrix**

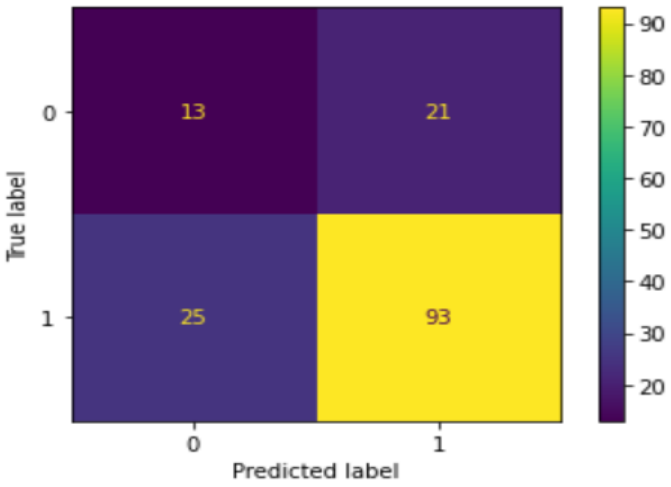


Figure 4.5: Decision Tree

- **RF Confusion Matrix**

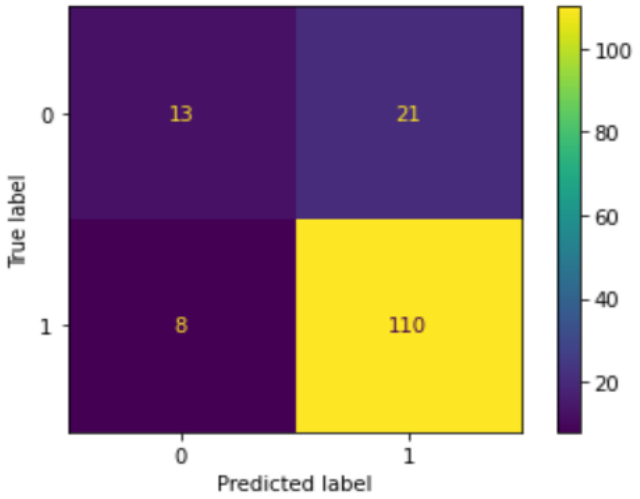


Figure 4.6: Random Forest

- **XGB Confusion Matrix**

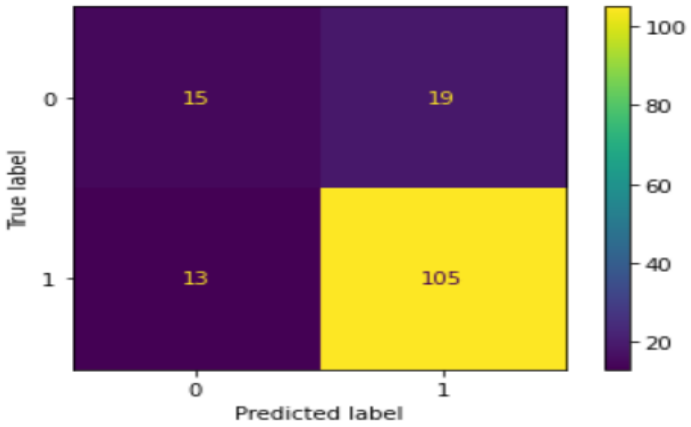


Figure 4.7: Extreme Gradient Boosting

- **SVC Confusion Matrix**

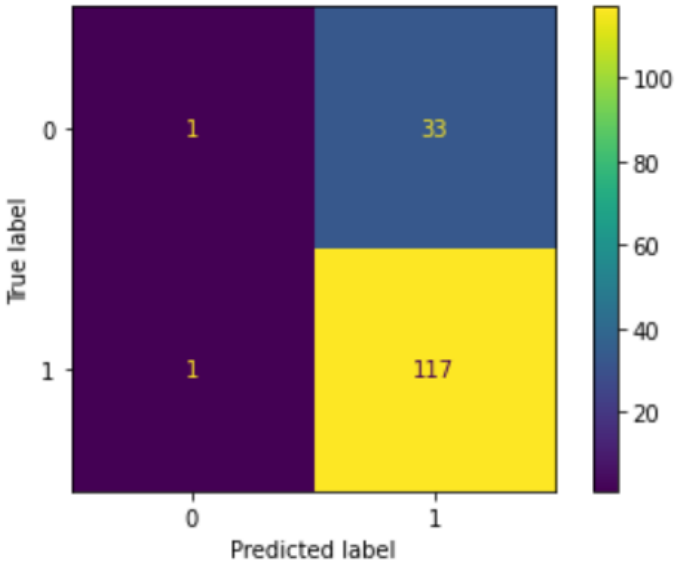


Figure 4.8: Support Vector Machine

Classification Report of Confusion Matrix

- **LR Curve**

From the given Logistic regression curve we see that the True positive rate increase and false positive rate decrees. In Logistic Regression, we don't directly fit a straight line to our data like in linear regression

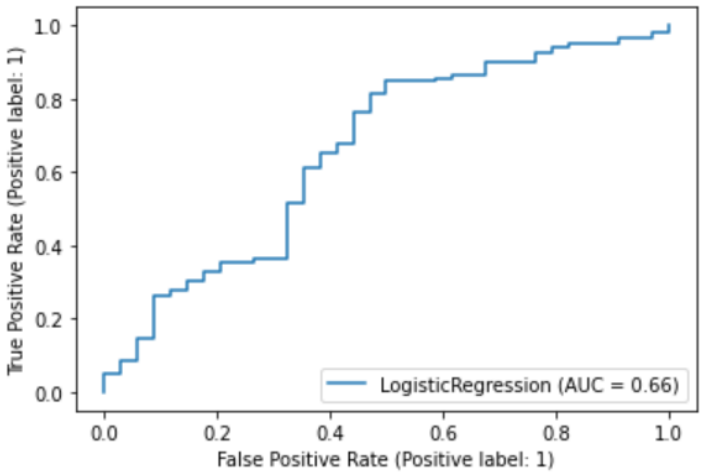


Figure 4.9: Logistic Regression

- **KNM Curve**

KNM Confusion matrix is one such important tool which helps us evaluate our model's performance. As the name suggests it is a matrix of size n x n. where 'n' is the number of class labels in our problem.

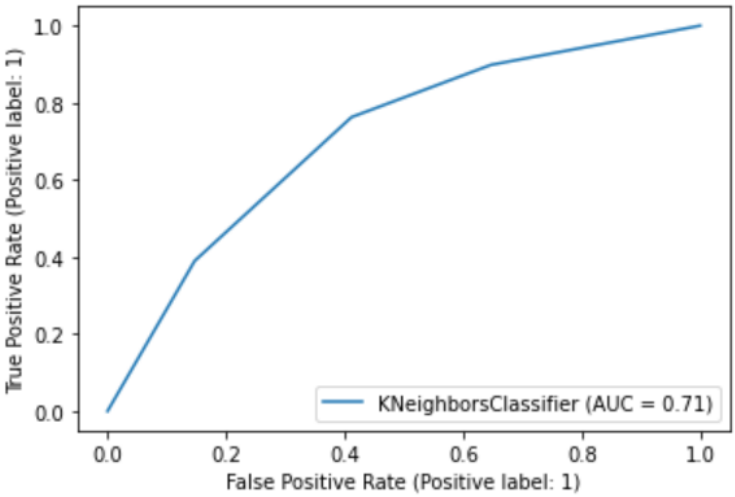


Figure 4.10: K-Nearest Neighbors

❖ **DT Curve**

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks.

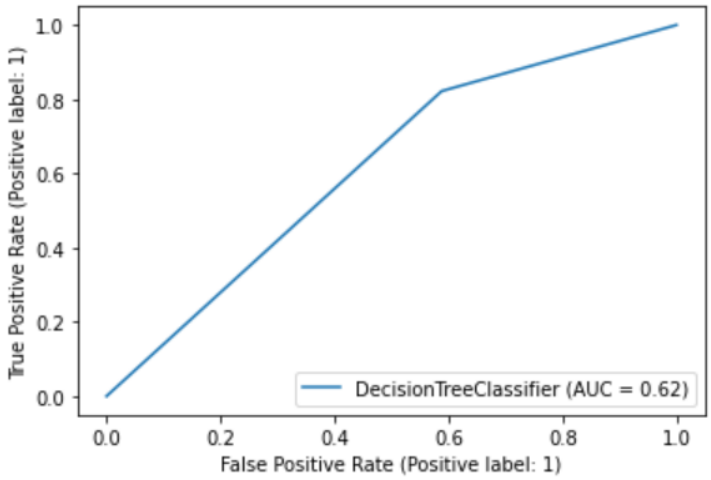


Figure 4.11: Decision Tree

❖ **RF Curve**

Random Forest (RF) algorithm is one of the best algorithms for classification. RF is able for classifying large data with accuracy.

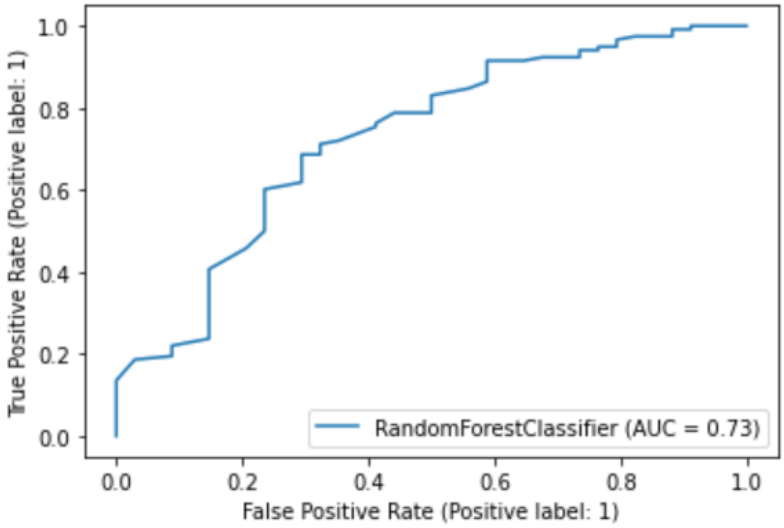


Figure4.12: Random Forest

❖ **XGB Curve**

XGBoost is an open-source software library that implements optimized distributed gradient boosting machine learning algorithms under the Gradient Boosting framework.

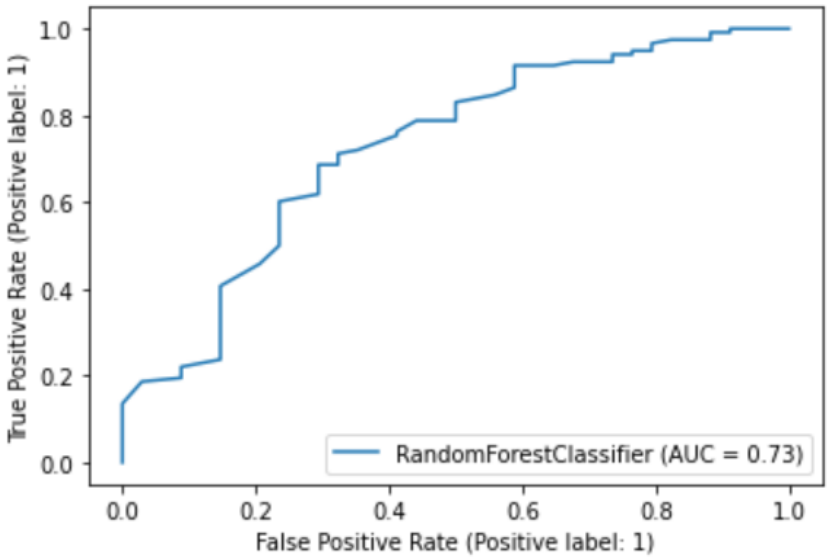


Figure 4.13: Extreme Gradient Boosting

❖ **SVC Curve**

The confusion matrix is a table that is used to show the number of correct and incorrect predictions on a classification problem when the real values of the Test Set are known.

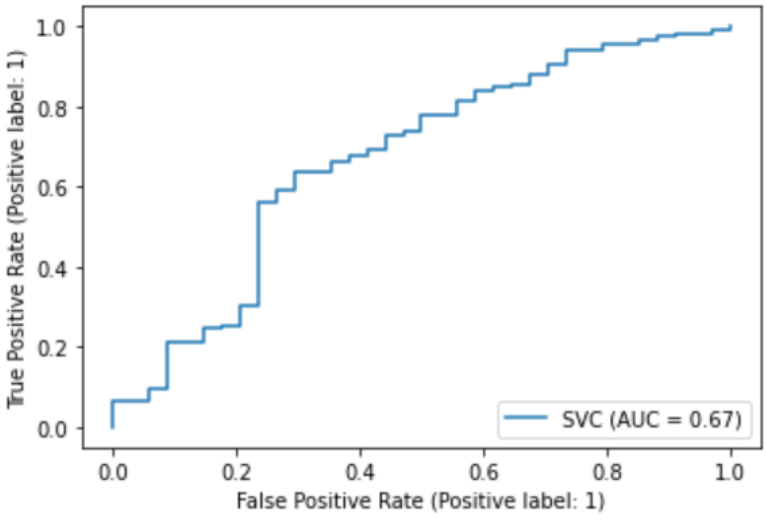


Figure4.14: Support Vector Machine

4.3 Discussion

Our present research work was done to assess classification algorithms in machine learning language. Here we identify Parkinson's disease from the data set that this disease is Parkinson's disease YES or No. This research work also focused on the medical field for the confirmation of clinical trials and identifying Parkinson's disease.

In our study, we used six types of machine learning classification analysis algorithms (LR, KNM, DT, RF, XGB, SVM) to identify PD datasets. From the Logistic Regression classifiers algorithm, we gain 76% accuracy from the accuracy calculation. We archive 77% accuracy from KNM, from SVM we achieve 77% accuracy, from DT we get 73% accuracy, from XGB we found 78 % accuracy. But we archive the highest accuracy 81% from Random Forest (RF) in our research study where the best accuracy was found to identify Parkinson's disease.

In the future, using this model in the medical field will help to classify Parkinson's disease more easily.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The main motive of our work is to help every kind of person. We can see in our society there is lots of people who are not aware about their health and they are unable to take proper treatment although it's a basic right. We work on Parkinson Disease which is enhancement day by day but there are lots of people who are not aware about this disease. This is the main thing we want to do in this work to raise awareness among all kinds of people in society. Village people are very poor and illiterate; they don't know about this disease. This disease can be identified by its syndromes and give the patient proper treatment.

5.2 Impact on Environment

People are becoming smarter day by day and they want smart greetings. It`s make a huge impact on the Environment. We saw that there is lots of people are invaded by this disease. Those patients are maximum over 60 years. This disease directly makes a huge impact on the patient 's brain. This disease damages the brain's working process. The people who are affected by this disease are unable to remember lots of things. It's an extensive effect in their daily life. If we grasp that the patient is invaded by this disease, we can take better care of them. Patients must have to visit a specialist doctor for that.

5.3 Ethical Aspects

Our work practices or procedures are ethical and do not violate anyone's privacy or human rights. We collect data from online platforms. Consequently, no one can be injured as a result of this study because we use this data only for our research. We did not perform any hazardous work on humans or animals during this investigation. While working, we utilized our own computers. During our investigations, we have maintained honesty, legal compliance, integrity, legality and transparency.

5.4 Sustainability Plan

We use machine learning algorithms to identify consumers and classify patients with Parkinson's disease. To develop our model, we use a specific dataset. As a result, a large number of datasets are linked Patients and doctors will be required to progress and maintain this study in future. Using our approach, you can make many long-term changes to any medical center entity. The online disease dataset with more features will be expanded and improved in the future. This concept can be employed in a variety of settings like a medical center. The medical fields will be improved by adopting and upgrading this model. Our newly proposed framework to integrate with medical website or mobile application will be useful in the future when dealing with another large dataset or others.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

First, we discuss our idea with our respected supervisor sir. Then do some basic research on what to do and we search some online sites about our ideas. Then our honorable teacher praised us. Then, we were concerned about our data collection. After much searching, we will collect our data within a few days. Finally, we got seven hundred and fifty-seven data. Then we pre-processed our data. After preprocessing the data, our data is flat. Then we apply Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM). Then, using the training dataset, we tested our model. The training and validation accuracy graph of our project is the best. Ultimately, we identify that our developed Random Forest Classifier (RFC) shows the result is 81%, which is the highest level of accuracy.

6.2 Conclusions

In this analysis, we've got six diagrammatic supervised learning machine learning approaches. A moment later, the performance of the six classifiers that are utilized within the prediction of paralysis against and assessed their exhibition utilizing various applied mathematics ways. The tentative performance demonstrates that the Random methodology has achieved the very best performance of the opposite 5 classifiers among the Parkinson datasets. It is 100%. This analysis has utilized six machine learning ways for the exposure of brain disorder disease visible of some parameters. In accumulation, this work is a component element} of a Thesis that has the aim to cultivate an automated application to allow a lot of correct action to traditional occurrences and build larger choices in many-sided situations. The applying is able to notice brain disorder disease in only a few minutes and apprize the damaging probability of getting the illness. This application is outstandingly useful for individuals, wherever could be a lack of medical institutes and additionally as explicit physicians. In my experiments, every classification rule was ready

and assessed on a coaching set that features each positive and negative samples. Moreover, the work is a certificate of Parkinson's disease detection by aggregation information from totally different clinical and medical centers and may offer a lot of correct results for illness prediction and diagnosing. In our analysis goal, there are many directions for future add this space of research. We've got solely investigated six standard supervised algorithms; it is most well-liked a lot of algorithms for developing the precise model of those paralysis against predictions and performance is improved. In abstract, our study painted the analysis objective besides chance with respect to paralysis against space by machine learning approaches that has associate degree arising impression in health fields.

6.3 Implication for Further Study

Right now, one of the biggest areas in terms of research and imaging is the inclusion of artificial intelligence and especially machine learning and deep learning algorithms to actually look at this when we're thinking of imaging. Today we're also dealing with numbers and these numbers need to be analyzed and when we're dealing with all of multi-modal. We're also looking at multimodal data and big data as getting more and more experience and sharing. We will end up with big collections of data and deal with all of that. We think artificial intelligence algorithms can actually help in terms of how we manage all this information. We don't think right now it's a clinical practice kind of application in terms of imaging analysis. We think we're not there yet. If we're in a regular in hospital that's not involved in a lot of research and we're only doing standard. we don't think it's something that we use every day but in future this will really help us in terms of addressing all this information and actually helping in terms of deciding which combination of parameters will be better to include in the differential diagnosis of this or that so we think in we're in the process of actually having these algorithms helping us to address all this big data that we're generating and it will be in the future very helpful.

References

[1] C. Marras et al., "Prevalence of Parkinson's disease across North America", *npj Parkinson's Disease*, vol. 4, no. 1, 2018. Available: 10.1038/s41531-018-0058-0.

[2] D. Gelb, E. Oliver and S. Gilman, "Diagnostic Criteria for Parkinson Disease", *Archives of Neurology*, vol. 56, no. 1, p. 33, 1999. Available: 10.1001/archneur.56.1.33.

[3] R. Khaskhoussy and Y. Ayed, "Speech processing for early Parkinson's disease diagnosis: machine learning and deep learning-based approach", *Social Network Analysis and Mining*, vol. 12, no. 1, 2022. Available: 10.1007/s13278-022-00905-9.

[4] T. Sherer, S. Chowdhury, K. Peabody and D. Brooks, "Overcoming obstacles in Parkinson's disease", *Movement Disorders*, vol. 27, no. 13, pp. 1606-1611, 2012. Available: 10.1002/mds.25260.

[5] T. Sherer, S. Chowdhury, K. Peabody and D. Brooks, "Overcoming obstacles in Parkinson's disease", *Movement Disorders*, vol. 27, no. 13, pp. 1606-1611, 2012. Available: 10.1002/mds.25260.

[6] D. Aarsland, K. Andersen, J. Larsen and A. Lolk, "Prevalence and Characteristics of Dementia in Parkinson Disease", *Archives of Neurology*, vol. 60, no. 3, p. 387, 2003. Available: 10.1001/archneur.60.3.387.

[7] T. Sherer, S. Chowdhury, K. Peabody and D. Brooks, "Overcoming obstacles in Parkinson's disease", *Movement Disorders*, vol. 27, no. 13, pp. 1606-1611, 2012. Available: 10.1002/mds.25260.

[8] T. Sherer, S. Chowdhury, K. Peabody and D. Brooks, "Overcoming obstacles in Parkinson's disease", *Movement Disorders*, vol. 27, no. 13, pp. 1606-1611, 2012. Available: 10.1002/mds.25260.

[9] A. Tarakad and J. Jankovic, "Recent advances in understanding and treatment of Parkinson's disease", *Faculty Reviews*, vol. 9, 2020. Available: 10.12703/b/9-6.

[10] A. Ho, R. Iansek, C. Marigliani, J. Bradshaw and S. Gates, "Speech Impairment in a Large Sample of Patients with Parkinson's Disease", *Behavioural Neurology*, vol. 11, no. 3, pp. 131-137, 1999. Available: 10.1155/1999/327643.

[11] T. Sherer, S. Chowdhury, K. Peabody and D. Brooks, "Overcoming obstacles in Parkinson's disease", *Movement Disorders*, vol. 27, no. 13, pp. 1606-1611, 2012. Available: 10.1002/mds.25260.

[12] T. Sherer, S. Chowdhury, K. Peabody and D. Brooks, "Overcoming obstacles in Parkinson's disease", *Movement Disorders*, vol. 27, no. 13, pp. 1606-1611, 2012. Available: 10.1002/mds.25260.

[13] R. Savica, B. Grossardt, J. Bower, J. Ahlskog and W. Rocca, "Risk factors for Parkinson's disease may differ in men and women: an exploratory study", *Hormones and Behavior*, vol. 63, no. 2, pp. 308-314, 2013. Available: 10.1016/j.yhbeh.2012.05.013.

[14] A. Schapira, "Etiology and Pathogenesis of Parkinson Disease", *Neurologic Clinics*, vol. 27, no. 3, pp. 583-603, 2009. Available: 10.1016/j.ncl.2009.04.004.

[15] E. Saracchi, "Emerging Candidate Biomarkers for Parkinson's Disease: a Review", *Aging and Disease*, 2014. Available: 10.14336/ad.2014.050027.

[16] R. Postuma, J. Gagnon and J. Montplaisir, "Clinical prediction of Parkinson's disease: planning for the age of neuroprotection", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 81, no. 9, pp. 1008-1013, 2010. Available: 10.1136/jnnp.2009.174748.

[17] S. Yao, A. Hart and M. Terzella, "An evidence-based osteopathic approach to Parkinson disease", *Osteopathic Family Physician*, vol. 5, no. 3, pp. 96-101, 2013. Available: 10.1016/j.osfp.2013.01.003.

[18] S. Lim and A. Lang, "The nonmotor symptoms of Parkinson's disease-An overview", *Movement Disorders*, vol. 25, no. 1, pp. S123-S130, 2010. Available: 10.1002/mds.22786.

[19] S. Yao, A. Hart and M. Terzella, "An evidence-based osteopathic approach to Parkinson disease", *Osteopathic Family Physician*, vol. 5, no. 3, pp. 96-101, 2013. Available: 10.1016/j.osfp.2013.01.003.

[20] K. Chou, "Parkinson Disease", *Annals of Internal Medicine*, vol. 157, no. 9, p. ITC5, 2012. Available: 10.7326/0003-4819-157-9-201211060-01005.

[21] I. Rustempasic and M. Can, "Diagnosis of Parkinson's Disease using Fuzzy C-Means Clustering and Pattern Recognition", *Southeast Europe Journal of Soft Computing*, vol. 2, no. 1, 2013. Available: 10.21533/Sc journal. v2i1.44.

[22] Z. Guo and X. Guan, "Nonlinear generalized predictive control based on online least squares support vector machines", *Nonlinear Dynamics*, vol. 79, no. 2, pp. 1163-1168, 2014. Available: 10.1007/s11071-014-1733-2.

[23] L. Silveira-Moriyama et al., "The use of smell identification tests in the diagnosis of Parkinson's disease in Brazil", *Movement Disorders*, vol. 23, no. 16, pp. 2328-2334, 2008. Available: 10.1002/mds.22241.

[24] R. Prashanth, S. Dutta Roy, P. Mandal and S. Ghosh, "High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning", *International Journal of Medical Informatics*, vol. 90, pp. 13-21, 2016. Available: 10.1016/j.ijmedinf.2016.03.001.

[25] R. Prashanth, S. Dutta Roy, P. Mandal and S. Ghosh, "High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning", *International Journal of Medical Informatics*, vol. 90, pp. 13-21, 2016. Available: 10.1016/j.ijmedinf.2016.03.001.

[26] S. Fahn, "Parkinson's disease: 10 years of progress, 1997-2007", *Movement Disorders*, vol. 25, no. 1, pp. S2-S14, 2010. Available: 10.1002/mds.22796.

[27] T. Fritsch, K. Smyth, M. Wallendal, T. Hyde, G. Leo and D. Geldmacher, "Parkinson Disease", *Southern Medical Journal*, vol. 105, no. 12, pp. 650-656, 2012. Available: 10.1097/smj.0b013e318273a60d.

[28] R. Dabby et al., "Skin biopsy for assessment of autonomic denervation in Parkinson's disease", *Journal of Neural Transmission*, vol. 113, no. 9, pp. 1169-1176, 2006. Available: 10.1007/s00702-005-0431-0.

[29] V. Sternweiler, "Wound-Dermatological Management/Issues", *Journal of Wound, Ostomy & Continence Nursing*, vol. 36, no. 3, pp. S23-S24, 2009. Available: 10.1097/01.won.0000351951.92716.14.

[30] V. Sternweiler, "Wound-Dermatological Management/Issues", *Journal of Wound, Ostomy & Continence Nursing*, vol. 36, no. 3, pp. S23-S24, 2009. Available: 10.1097/01.won.0000351951.92716.14.

[31] G. Höglinger et al., "Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria", *Movement Disorders*, vol. 32, no. 6, pp. 853-864, 2017. Available: 10.1002/mds.26987 [Accessed 16 September 2022].

[32] C. Adler et al., "Low clinical diagnostic accuracy of early vs advanced Parkinson disease: Clinicopathologic study", *Neurology*, vol. 83, no. 5, pp. 406-412, 2014. Available: 10.1212/wnl.0000000000000641 [Accessed 16 September 2022].

[33] S. Koga et al., "When DLB, PD, and PSP masquerade as MSA", *Neurology*, vol. 85, no. 5, pp. 404-412, 2015. Available: 10.1212/wnl.0000000000001807 [Accessed 16 September 2022].

[34] T. Ellis, J. Cavanaugh, G. Earhart, M. Ford, K. Foreman and L. Dibble, "Which measures of physical function and motor impairment best predict quality of life in Parkinson's disease?", *Parkinsonism & Related Disorders*, vol. 17, no. 9, pp. 693-697, 2011. Available: 10.1016/j.parkreldis.2011.07.004.

[35] K. Berganzo et al., "Motor and non-motor symptoms of Parkinson's disease and their impact on quality of life and on different clinical subgroups", *Neurología (English Edition)*, vol. 31, no. 9, pp. 585-591, 2016. Available: 10.1016/j.nrleng.2014.10.016 [Accessed 16 September 2022].

A SUPERVISED MACHINE LEARNING APPROACH TO CLASSIFY PARKINSON DISEASE

ORIGINALITY REPORT

22%	13%	8%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Malaysia Sabah Student Paper	5%
2	www.researchgate.net Internet Source	3%
3	Submitted to Daffodil International University Student Paper	3%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
5	Kamal Nayan Reddy Challa, Venkata Sasank Pagolu, Ganapati Panda, Babita Majhi. "An improved approach for prediction of Parkinson's disease using machine learning techniques", 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 Publication	2%
6	docplayer.net Internet Source	1%
	Submitted to Bridgepoint Education	

7	Student Paper	1%
8	onlinelibrary.wiley.com Internet Source	1%
9	www.mdpi.com Internet Source	1%
10	Submitted to Vientiane International School Student Paper	<1%
11	www.Formpl.Us Internet Source	<1%
12	www.worldlifeexpectancy.com Internet Source	<1%
13	knowledgecommons.lakeheadu.ca Internet Source	<1%
14	Submitted to Macquarie University Student Paper	<1%
15	Submitted to Prairie View A&M University Student Paper	<1%
16	Submitted to University College London Student Paper	<1%
17	www.tandfonline.com Internet Source	<1%
18	academic-accelerator.com Internet Source	<1%

19	Md. Hasan Imam Bijoy, Anik Pramanik, Md. Sadekur Rahman, Mehedi Hasan, Sumiya Alam Akhi, Md. Mahbubur Rahman. "MKRF Stacking-Voting: A Data Mining Technique for Predicting Educational Satisfaction Level of Bangladeshis Student During Pandemic", 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022 Publication	<1%
20	ebin.pub Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		