VISVESVARAYA TECHNOLOGICAL UNIVERSITY
"Jnana Sangama", Belgavi- 590018

AN INTERNSHIP REPORT ON

"PARKINSON DIESEASE DETECTION"

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE & ENGINEERING

Submitted by

GOVARDHAN G P (1RR19C040)

Internship carried out at

TEQUED LABS

Internal Guide

Mr. SHASHIDHAR V
Associate Professor
Dept of CSE, RRCE
Bangalore

External Guide

Mr. Aditya S K
CTO
Tequed Labs,

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

RAJARAJESWARI COLLEGE OF ENGINEERING

MYSORE ROAD, BENGALURU-560074

(An ISO 9001:2008 Certified Institute)

(2022-2023)

RAJARAJESWARI COLLEGE OF ENGINEERING

MYSORE ROAD, BENGALURU-560074

(An ISO 9001:2008 Certified Institute)

(Affiliated to Visvesvaraya Technological University, Belgavi)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that Internship entitled

"PARKINSON DISEASE DETECTION"

Carried out by

GOVARDHAN G P (1RR19CS040)

The student of **"Rajarajeswari College of Engineering"** in partial fulfillment for the award of the degree of **Bachelor Of Engineering** in **Computer Science & Engineering** of the Visvesvaraya Technological University, Belgaum during the year **2022–2023.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The Internship report has been approved as it satisfies the academic requirements in respect of Internship work prescribed for the Eighth semester.

…….....	……...	…….
Signature of guide	**Signature of HOD**	**Signature of Principal**
[Mr. SHASHIDHAR V] BALAKRISHNA] Associate Prof., Dept. of CSE RRCE, Bangalore	**[Dr. S Usha]** Prof. & HOD, Dept. of CSE RRCE, Bangalore	**[Dr. R.** Principal RRCE, Bangalore

External Viva-Voce

Examiners: ***Signature:***

1. _____
2. _____

DECLARATION

I, Govardhan G P (**1RR19CS040**), student of B.E in Computer Science and Engineering, **Rajarajeswari College of Engineering,** Bengaluru, hereby declare that the internship work entitled **"Parkinson Disease Detection"** submitted to the **Visvesvaraya Technological University** during the academic year 2021-2022 is record of an original work done by us under the guidance of Mr. Shashidhar V, Associate Prof., Department of Computer Science and Engineering, Rajarajeswari College of Engineering, Bengaluru. This internship work is submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Engineering** in **Computer Science and Engineering.** The results embodied in this report have not been submitted to any other university or institute for any degree.

(**Govardhan G P)**
(**1RR19CS005**)

ACKNOWLEDGEMENT

Any achievement, be it scholastic or otherwise does not depend solely on the individual efforts but on the guidance, encouragement and cooperation of intellectuals, elders and friends. A number of personalities, in their own capacities have helped me in carrying out this internship. I would like to take this opportunity to thank them all.

I wish to express my profound gratitude to **Dr. A.C Shanmugam,** R.R.C.E, Bengaluru for his moral support towards completion of my internship.

I wish to express my profound gratitude to **Dr. R Balakrishna**, Principal, R.R.C.E, Bengaluru for his moral support towards completion of my internship.

I would like to express my deepest gratitude to **Dr. Usha Sakthivel**, Head of Department, Computer Science and Engineering, R.R.C.E, Bengaluru, for her continuous support and encouragement.

I would like to acknowledge the unbridled enthusiasm of my internship guide **Mr. Shashidhar V**, Associate Prof., Department of Computer Science and Engineering, R.R.C.E, Bengaluru, for her encouragement and valuable guidance throughout my internship.

I thank my Parents, and all the teaching and non-teaching faculty members of Department of Computer Science and Engineering for their constant support and encouragement. Last, but not the least, I would like to thank my friends who provided me with valuable suggestions to my work during internship.

COMPANY CERTIFICATE

CONTENTS

LIST OF FIGURES

Chapter 1

Company Profile

1.1 Introduction

Tequed Labs Private Limited is a Private incorporated on 22 January 2018. It is classified as Non-government Company and is registered at Registrar of Companies, Bangalore. Tequed Labs is a research and development center and educational institute based in Bangalore. They are focused on providing quality education on latest technologies and develop products which are of great need to the society. They also involve in distribution and sales of latest electronic innovation products developed all over the globe to their customers. They run a project consultancy where they undertake various projects from wide range of companies and assist them technically and build products and provide services to them. They are continuously involved in research about futuristic technologies and finding ways to simplify them for their clients. This project was the world finalist in the international innovation challenge called MASTERPIECE in Dubai. It has been exhibited in NASSCOM Product Conclave and has received great appreciation from IT giants. This product has been patentedbearing a patent number -201741034208.

They have developed a women's safety device which sends the location of the woman in distress to the nearby police station. This product won the best ICT category project awardin a state level exhibit and was exhibited at NASSCOM PRODUCT CONCLVE 2017.

Their other research work includes development of a device for blind which can recognize objects and convert it into speech. This innovation has a lot of potential in helping the blind people.

Their other products include: -

- Automation of production line and remote quality control monitoring system.
- Development of mobile app and website for sales of artistic and antique products.
- Development of an energy conservation system for paper machineries.
- Development of an analytic tool for software based vehicle condition analysis for resales.

1.2 Mission and Vision

Mission: To up skill students with the latest and most advanced technologies which are highly valued in the industries in order to make them more creative, innovative and knowledgeable thereby improving their employability.

Vision: We are continuously involved in research about futuristic technologies and finding ways tosimplify them for our students.

1.3 Clients

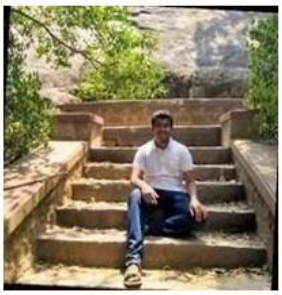
Figure: 1

1.4 Team

SUPREETH Y
Director and CEO at Tequed Labs Pvt Ltd.

ADITYA SK
CTO and Director at Tequed Labs and ASQR Technologies

RASHMI SWARAJ
Campus Ambassador at Tequed Labs



SRI KRISHNA VENKATESH
Intern Guide at Tequed Labs

1.5 Services

- Workshops
- Internships and Skill Development Programs
- Online Courses
- Faculty Development Programs

1.6 Training & Development

Tequed Labs are focused on providing quality education on latest technologies and develop products which are of great need to the society. They also involve distribution and sales of latest electronic innovation products developed all over the globe to our customers. They run a projectconsultancy where we undertake various projects from wide range of companies and assist themtechnically and build products and provide services to them. They are continuously involved inresearch about futuristic technologies and finding ways to simplify them for our students. Specialties involve internet of things, research and development, skill development, machine leaning, artificial intelligence, project consultancy, Software development, hardware design, and innovation.

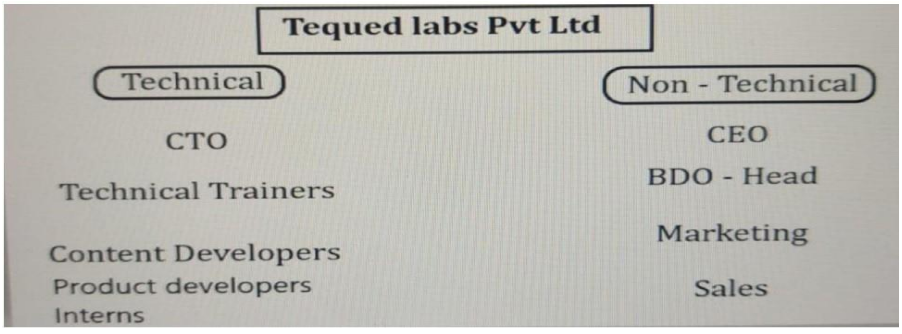
Tequed labs provides a weeklong training for the students for their relevant fields and gives them a small assessment to see whether they have understood the concepts or not.

1.7 Software Technologies

- Cyber Security and Ethical Hacking
- Internet of Things
- Artificial Intelligence
- Virtual and Augmented Reality
- IC Engine Design and Management

1.8 Organizational Structure:

Tequed Labs is a research and development center and educational institute based in Bangalore started by Mr. Aditya S K and Mr. Supreeth Y S. They are focused on providing quality education on latest technologies and develop products which are of great need to the society. They also involve distribution and sales of latest electronic innovation products developed all over the globe to our customers. They run a project consultancy where we undertake various projects from wide range of companies and assist them technically and build products and provide services to them. They are continuously involved in research about futuristic technologies and finding ways to simplify them for our students. Specialties involve internet of things, research and development, skill development, machine leaning, artificial intelligence, project consultancy, Software development, hardware design, and innovation. Tequed Labs assigns a group of 2-3 trainers for every class to overlook the training period and to assist the students regarding doubts.



Organization Structure

Chapter 2

Tasks Performed

2.1 System Design

System Design is the most important step for developing any model. depictsthe basic system architecture for our project. We created our own dataset in .csv file related to credit card details such as credit card number, transaction details fradualant details in tabular form which help us to carry out process of prediction. In the first step we import dataset and split them into two parts i.e Training data and Testing data where 80% of data is used for training model and 20% for testing the model.Pre-processing is an essential and mandatory step while creating model as it help us to remove unwanted noisy data, which not only helps in removing null values but adds on value to efficiency in prediction and accuracy. Then we apply Logistic regression algorithm which is one of the classification algorithms to predict the output.

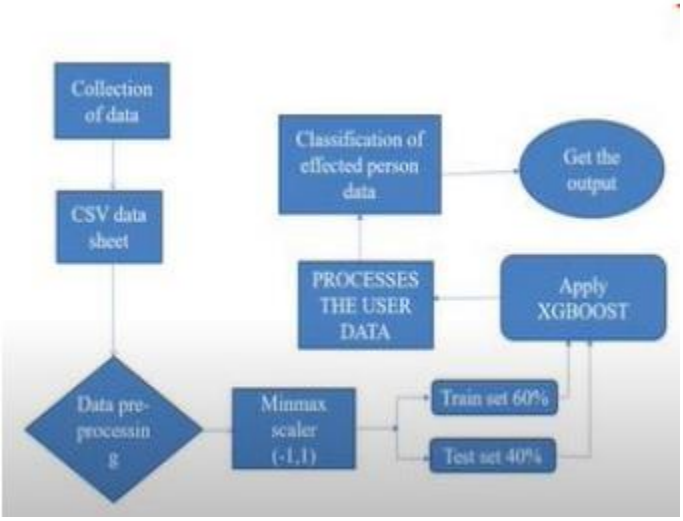


Fig 2: System design

2.2 Data acquisition and cleaning:

Data acquisition and cleaning are two essential steps in the data preprocessing phase of a machine learning project. Let's take a closer look at what each of these steps involves:

Data Acquisition:

Data acquisition is the process of collecting raw data from various sources such as databases, APIs, websites, or other data repositories. The quality of the data is critical as it can significantly impact the accuracy and performance of the machine learning model. Here are some steps involved in data acquisition:

- Identify Data Sources: Identify the data sources that are relevant to the problem you are trying to solve. This could be data from internal databases, public datasets, or data obtained through web scraping.
- Collect Data: Collect data from the identified sources using appropriate tools and techniques.
- Data Integration: Integrate the collected data from different sources into a single dataset.
- Data Transformation: Transform the data into a suitable format that can be analyzed and modeled.

Data Cleaning:

Data cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in the collected dataset. It is essential to perform data cleaning to ensure that the model is trained on high-quality data. Here are some steps involved in data cleaning:

- Identify Missing Values: Identify missing data points in the dataset and determine the best way to handle them. This could involve imputing missing values using mean, median, or mode, or removing the entire row or column with missing values.
- Identify Outliers: Identify outliers in the data and determine if they need to be removed or kept in the dataset.
- Handle Duplicate Data: Identify and remove duplicate data points in the dataset.
- Handle Inconsistencies: Identify inconsistencies in the dataset, such as typos or misspellings, and correct them.
- Handle Irrelevant Data: Identify irrelevant data points that do not contribute to the problem being solved, and remove them from the dataset.

2.3 Data Visualization:

Data visualization refers to the use of visual elements such as charts, graphs, maps, and other graphical representations to communicate complex data in an easy-to-understand manner. The main goal of data visualization is to enable people to easily and quickly interpret and comprehend large amounts of data, patterns, and relationships within the data.

Effective data visualization uses various design techniques to display data in a clear and intuitive manner. For example, the use of colors, shapes, sizes, and positioning of visual elements can help emphasize important points or trends in the data. Good data visualization also takes into account the target audience, purpose of the visualization, and the type of data being presented.

Data visualization has many practical applications, including business intelligence, scientific research, finance, marketing, and many other fields. It helps decision-makers gain insights and identify patterns and trends that might otherwise be hidden in large volumes of data.

2.4 Feature Extraction:

Feature extraction is the process of studying the behavior and pattern of the analyzed data and draw the features for further testing and training. Finally, our models are trained using the Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

2.5 Evaluation Model:

Model Evaluation is an essential part of the model development process. It helps to find the best model that represents our data and how well the selected model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can effortlessly generate overoptimistically and over fitted models. To avoid overfitting, evaluation methods such as hold out and cross-validations are used to test to evaluate model performance. The result will be in the visualized form. Representation of classified data in the form of graphs. Accuracy is well-defined as the proportion of precise predictions for the test data. It can be calculated easily by mathematical calculation i.e. dividing the number of correct predictions by the number of total predictions.

Chapter 3

Technical and Non-Technical Activities

3.1 Non-technical activities:

- **Research:** First and foremost we tried to learn more about Parkinson's disease, its causes, symptoms, and treatments. This knowledge helped us to better understand the dataset we were working with and provide context for our analysis. This knowledge also helped u s analyze the Parkinson's dataset more effectively and identify key insights that could inform future research.

- **Data visualization:** Create visually appealing and easy-to-understand graphs and charts to present your findings. This can be an effective way to communicate complex information to non-technical audiences. Creating clear and concise data visualizations can help you communicate complex information more effectively. We have used tools like charts, graphs, and info graphics to present your findings in a visually appealing and easy-to-understand way. This can be particularly helpful when presenting our results to non-technical audiences.

- **Communication:** Practice your communication skills by presenting your findings to your colleagues or supervisors. This will not only help you improve your presentation skills but also allow you to receive feedback and suggestions for future research.

- **Collaboration:** Work with other interns or colleagues to create interdisciplinary projects that combine technical and non-technical skills. For example, you could collaborate with a graphic designer to create data visualizations, or work with a writer to create compelling content about Parkinson's disease.

- **Oral Communication Skills:** Making a speech or presentation before a large group of people such as manager, customer, and your peers. The need to effectively communicate technical information about our particular discipline to others and the need to maintain a friendly work environment. We must have a good communication skill in any

organization today. My internship taught me of how to communicate effectively both within the team and outside the team.

- **How to Work with Co-Workers:** Everyone is different and getting along with co-workers isn't always easy. We often come across unusual characters at work. My internship helped me learn how to interact with different types of people in a professional setting, this can take some getting used to but will help me to settle in at new jobs faster.

- **Time Management Skills:** To be an effective employee and successful we must be able to manage our time efficiently. Most people have fairly good time management skills from school or university experience. In the workplace however, it can be quite different, often with a much higher tempo of tasks needing to be completed and usually unforeseen variations during the day. My internship taught me techniques of how to manage time better.

3.2 Technical Activities:

Software:

Python:

Python is a powerful, versatile programming language that is widely used in the data science and machine learning communities. It has many libraries and frameworks that make it easy to build and train machine learning models, such as scikit-learn, TensorFlow, and Keras. Python's popularity is due in part to its ease of use, flexibility, and the fact that it is an open-source language with a large and active community of developers. Another advantage of Python is its strong community and ecosystem. Python has a large and active community of developers who contribute to the language and its libraries, provide support, and share knowledge and resources. This community has created many open-source libraries and tools that make it easy to build and deploy Python applications.

Google Collab :

Google Collab (short for Collaboratory) is a free cloud-based service from Google that provides a Jupyter notebook environment for running Python code. It is widely used by researchers, students, and developers for data analysis, machine learning, and deep learning tasks.

Some key features of Google Collab are:

- Free access: Anyone with a Google account can access Google Collab for free.
- Collaborative: Google Collab allows multiple users to collaborate on the same notebook in real-time.
- Pre-installed libraries: Google Collab comes pre-installed with popular Python libraries such as NumPy, pandas, and Matplotlib.
- GPU and TPU support: Google Collab allows users to use GPUs and TPUs for machine learning tasks, which can significantly speed up computation time.
- Cloud storage: Google Collab provides cloud storage for users to store and access their notebooks and data.
- Easy sharing: Google Collab allows users to share their notebooks with others easily.

Libraries:

Numpy:

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

Pandas:

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data. The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

Key Features of Pandas

- Quick and efficient data manipulation and analysis.
- Tools for loading data from different file formats into in-memory data objects.
- Label-based Slicing, Indexing, and Subsetting can be performed on large datasets.
- Merges and joins two datasets easily.
- Pivoting and reshaping data sets.

Algorithm:

XGBoost Classifier:

- XGBoost is a new Machine Learning algorithm designed with speed and performance in mind.
- XGBoost stands for eXtreme Gradient Boosting and is based on decision trees.
- In this project, we will import the XGBClassifier from the xgboost library
- This is an implementation of the scikit-learn API for XGBoost classification.

Chapter 4

OUTCOME OF INTERNSHIP

4.1 Learning from Internship Program:

The Internship program was beneficial for us. It helped us in improving our various technical skills and enhanced my knowledge in new areas.

- We gained knowledge in the domain of AIML, the various issues involved and mechanisms in the systems etc.
- By studying Machine leaning, we also learnt that how machine learning functions and what are the various issues one need to be aware of while analyzing data.
- We worked on a data analysis project for the first time we got introduced to this important area of data analysis and the importance of the same.
- We brushed up our knowledge on Python as a coding language and also the required libraries for the purpose of analyzing data and representing the same by using python libraries like Matplotlib.

Work Experience

Our Internship was quite satisfactory in terms of work environment. The team with which we worked with was very friendly and helped us in solving our problems. New experiences include

- **Team work**

In this project 3-4 people work together thus providing enough opportunity for team work and coordination. This was a good experience for us as the team was very cooperative and understanding.

- **Responsibility and Keeping Commitments**

The importance of keeping commitments and the time of others was an important thing, which I learnt as an intern.

4.2 Technical Outcomes:

4.2.1 Knowledge of Data Science

Data science is an interdisciplinary field that combines computer science, statistics, and domain expertise to extract insights and knowledge from data. The field has grown in importance in recent years due to the explosion of digital data and the increasing availability of powerful computing resources. Data scientists work with large and complex data sets, often requiring them to clean, preprocess, and transform the data before analysis. One important aspect of data science is data visualization. Data visualization is the process of creating visual representations of data to help people understand patterns and relationships. Data scientists often use tools like Matplotlib, Seaborn, and Tableau to create interactive and informative data visualizations. Data visualization can be used to identify trends, anomalies, and outliers in data, as well as to communicate findings to stakeholders. Another important aspect of data science is machine learning. Machine learning is a type of artificial intelligence that involves the use of statistical models and algorithms to make predictions or classifications based on data. Machine learning can be used for a wide range of applications, from fraud detection and recommendation engines to image recognition and speech processing. Popular machine learning frameworks include Scikit-Learn, TensorFlow, and Keras.

4.2.2 Data Science

Data science has significant applications across industries, including finance, healthcare, marketing, and transportation. In finance, data science is used for fraud detection, risk management, and trading strategies. In healthcare, data science is used to develop predictive models for disease diagnosis, drug discovery, and. In marketing, data science is used for customer segmentation, targeting, and personalized recommendations. In conclusion, data science is a vital field that involves the extraction of knowledge and insights from data using a combination of tools and techniques from several domains. Data science has significant applications across industries, including finance, healthcare, marketing, and transportation.

4.2.3 Using profession specific terminology appropriately

One of the most important things is that writing must be appropriate for others to understand. I have learnt to use the terminology that makes sense to my readers. When we become a

professional in a particular field, we have to learn language of that field. As a professional, people hire us to work on their behalf because we have the knowledge and skills that they don't possess. Therefore, we find our self-communicating in writing with people who don't know the language of our field; our writing must make sense to people to read it.

4.3 Non-Technical Outcomes

4.3.1 Verbal and Written Communication Skills

We have demonstrated and learnt our strong communication skills, both verbal and written, throughout the internship, while working for training and development.

4.3.2 Personality Development

Personality development plays an imperative role at workplace as it decides the way an individual interacts with his fellow workers and responds to various situations. How an individual behaves at the workplace depends on his/her personality. Personality development helps in polishing and grooming individuals and makes them better and efficient resources for the organization. Personality development also reduces stress levels and teaches an individual to face even the worst situations with a smile. Personality reflects how one conducts himself/herself in the professional environment. Never carry your personal problems to work. Personality development helps an individual to keep his personal life separate from his professional life.

Differences in opinions and views often lead to conflicts and arguments among employees. Employees with different attitudes and mindsets find it extremely difficult to adjust with each other and work in unison. Personality development sessions motivate an individual to think positively and eventually reduce stress at the workplace. Individuals as a result of personality development tend to behave in a mature way; making the organization a much better place to work. Personality development is essential to bring a change in an individual's attitude, thinking, behavior and mindsets. It also strengthens the relationship among co -workers.

4.3.3 Time Management

Proper time management at work place has a number of positive effects, ranging from making you more focused and valuable employee of reducing the stress of your job. Employers appreciate employees who can get the maximum amount of good works done in minimum amount of time. It also strengthens the relationship among co- workers.

4.3.4 Resource Utilization Skills

Effective management of resources is essential tasks of companies that are managing different projects. It is important for them to efficiently organize allocate personal as well as equipment for different projects, same time avoiding idle resources. Having the information of availability of resource and have those available at the right time for the activities plays the vital role in managing the costs and smoothly executing the project activities.

4.3.5 Time Managing Skills:

To be an effective employee and successful we must be able to manage our time efficiently. Most people have effectively good tome management skill from school or university experience. We had an opportunity to improve our time management skills such that we could apply it both in our professional and personal life. Oftentimes we were tested for time management with tasks to be performed within schedule that helped us to complete the tasks within scheduled time.

4.3.6 Teamwork:

Teamwork is such an important aspect of running a successful company and our internship has thought us how to do this on a business level. Key techniques adopted in our internship for effective team work:

- Communication Process: exchange contact details, establishing rules and regulations and limits on communication forms (Email, Texts, Mobile Facebook), maximum length of time responses, when face-to-face meetings were held.
- Task Analysis and distribution of workload: dividing up questions or parts of questions into equal proportions, going off to work individually to return the day before the task is due to put it all together.
- Meeting Schedule: Meetings were conducted frequently (once in 4 days) to know the status of the task being completed. This allowed us maintain our pace completing the assigned work.

Chapter 5

Screenshots

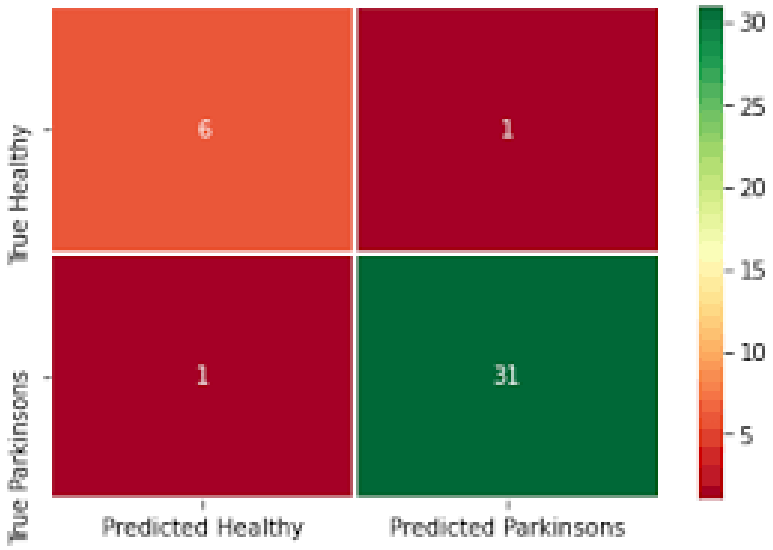


Fig 3 : Correlation in heatmap

Training data accuracy:	**96%**
Testing data accuracy:	94.87%

Fig 4: Accuracy Score

CONCLUSION

Parkinson's is the second most neurodegenerative disease which has no cure. It results in difficulty of body movements, anxiety, breathing problems, loss of smell, depression, and speech. In this paper, the machine learning algorithms used to measure the performance XGB Classifier applied on the dataset. The author chose the voice features of patients as the dataset contains more than 700 features and finally took the ten important features that are useful to evaluate the system. The author compared the machine learning methods accuracies and based on this one prediction model is generated. Hence, the aim is to use various evaluation metrics like confusion matrix, accuracy, precision, recall, and f1-score which predicts the disease efficiently. This gives us an accuracy of 94.87%.

REFERENCES

[1] Arvind Kumar Tiwari, "Machine Learning based Approaches for Prediction of Parkinson's Disease", Machine Learning and Applications: An International Journal (MLAU) vol. 3, June 2016.

[2] Carlo Ricciardi, et al, "Using gait analysis' parameters to classify Parkinsonism: A data mining approach" Computer Methods and Programs in Biomedicine vol. 180, Oct. 2019.

[3] Dr. Anupam Bhatia and Raunak Sulekh, "Predictive Model for Parkinson's Disease through Naive Bayes Classification" International Journal of Computer Science & Communication vol. 9, Dec. 2017, pp. 194- 202, Sept 2017 - March 2018. [4] Dragana Miljkovic et al, "Machine Learning and Data Mining Methods for Managing Parkinson's Disease" LNAI 9605, pp 209-220, 2016.

[5] M. Abdar and M. Zomorodi-Moghadam, "Impact of Patients' Gender on Parkinson's disease using Classification Algorithms" Journal of AI and Data Mining, vol. 6, 2018.

[6] M. A. E. Van Stiphout, J. Marinus, J. J. Van Hilten, F. Lobbezoo, and C. De Baat, "Oral health of Parkinson's disease patients: a case-control study," Parkinson's Disease, vol. 2018, Article ID 9315285, 8 pages, 2018.

[7] Md. Redone Hassan et al, "A Knowledge Base Data Mining based on Parkinson's Disease" International Conference on System Modelling & Advancement in Research Trends, 2019.

[8] Mehrbakhsh Nilashi et al, "A hybrid intelligent system for the prediction of Parkinson's Disease progression using Machine Learning techniques" Biocybernetics and Biomedical Engineering 2017.

[9] R. P. Duncan, A. L. Leddy, J. T. Cavanaugh et al., "Detecting and predicting balance decline in Parkinson disease: a prospective cohort study," Journal of Parkinson's Disease, vol. 5, no. 1, pp. 131–139, 2015.

[10] Ramzi M. Sadek et al., "Parkinson's Disease Prediction using Artificial Neural Network" International Journal of Academic Health and Medical Research, vol. 3, Issue 1, January 2019.