**PARKINSON’S DISEASE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES**

**A project work Synopsis**

***Submitted By:***

GIRESH VENKATA TRINADH KANDRU (20BCS6536)

AYUSH SHARMA (20BCS6550)

MUDIT SHARMA (20BCS6542)

RISHABH RAI (20BCS6757)

Pooja jain (20bcs6551)

***in the partial fulfilment for the award of the degree of***

**BACHELOR OF ENGINEERING**

**IN**

Computer science engineering

***Under the supervision of:***

Ms. Merry K.P



##### DEC 2022

**Table of Contents**

1. **INTRODUCTION** 
   1. Project Overview
   2. Gantt Chart
   3. Problem Statement
   4. Problem Formulation
2. **HARDWARE/SOFTWARE SPECIFICATION** 
   1. Hardware Specification
   2. Software Specification
3. **LITERATURE SURVEY**
4. **METHODOOGY**
5. **RESULT**
6. **CONCLUSION AND FUTURE SCOPE**
   1. Conclusion
   2. Future Scope

**Abstract**

Parkinson’s disease (PD) among Alzheimer’s and epilepsy are one of the most common neurological disorders which appreciably affect not only live of patients but also their households.  However, traditional diagnostic approaches may suffer from subjectivity as they rely on the evaluation of movements that are sometimes subtle to human eyes and therefore difficult to classify, leading to possible misclassification. In the meantime, early non-motor symptoms of PD may be mild and can be caused by many other conditions. The symptoms of PD are often overlooked, making diagnosis of PD at an early stage challenging. To address these difficulties and to refine the diagnosis and assessment procedures of PD, machine learning methods have been implemented for the classification of PD and healthy controls or patients with similar clinical presentations (e.g., movement disorders or other Parkinsonian syndromes).  Even though there is no cure for PD, a proper medication at the early stage can help significantly in alleviating the symptoms. Since, the traditional method for identifying Parkinson disease is rather invasive, expansive and complicated for self-use, there is a high demand for using ML algorithms like classification method on PD detection. To solve these issues which we conducted a literature analysis of research papers. And most of the research papers were included in this study, with an examination of their targets, data sources and different types of datasets, ML algorithms, and associated outcomes. The results showed that ML approaches have a lot of promise for being used in clinical decision-making, resulting in a more systematic and informed way to predict Parkinson disease.

## 1. INTRODUCTION

Parkinson's disease is a progressive nervous system disorder that affects movement leading to shaking, stiffness, and difficulty with walking, balance, and coordination. Parkinson's symptoms usually begin gradually and get worse over time.

Parkinson's disease symptoms can be different for everyone. Early signs are mild that goes unnoticed. Symptoms usually begin on one side of your body and gets worsen on that side, afterwards it affects both the sides.

Parkinson's symptoms may include:

• Tremor

• Slowed movement

• Rigid muscles.

• Impaired posture and balance.

• Loss of automatic movements

• Speech changes

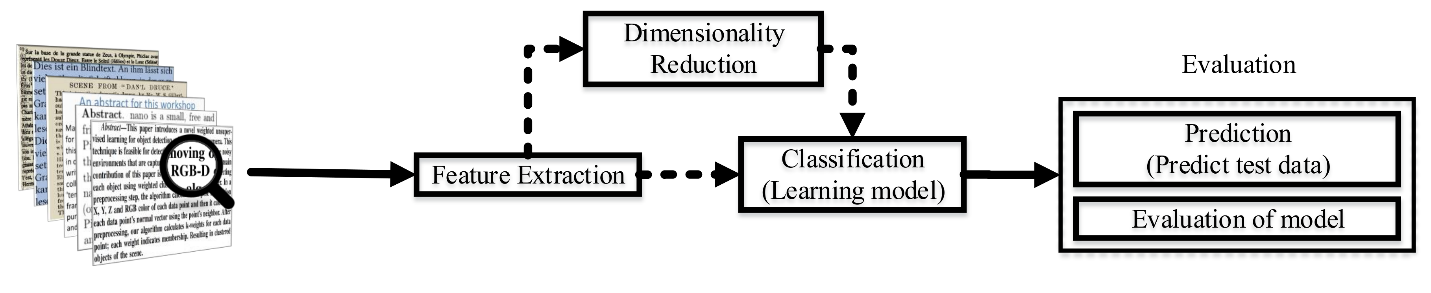
• Writing changes

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed task.

In addition to an informed, working definition of machine learning (ML), we detail the challenges and limitations of getting machines to ‘think,’ some of the issues being tackled today in deep learning (the frontier of machine learning), and key takeaways for developing machine learning applications for business use-cases. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

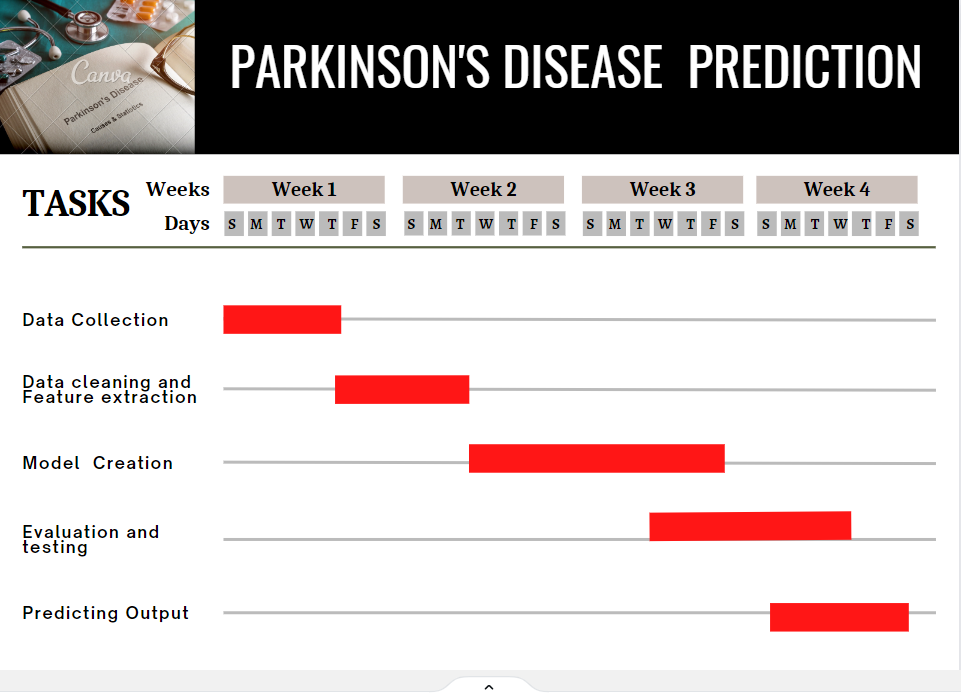
### Project Overview

Our project is based on ML classification algorithm and through this project we will detect whether the person is suffering from Parkinson Disease or not for that, first we have to collect the dataset of the patients. So, from Kaggle we will download Parkinson detection Datasets of PD affected and unaffected patients collected by neurologists are obtained from Machine Learning repository. These are stored into the python environment as Testing and Training datasets and imported using necessary packages. Python is an open source dynamic, high level, free and interpreted programming language. This supports object-oriented programming and procedural programming. Python is currently the most popular programming language for Machine Learning research and development. Jupyter notebook is an integrated development environment (IDE) primarily for the Python language, used in computer programming.



* 1. **GANTT CHART-**

A Gantt chart is a horizontal bar chart developed as a production control tool in 1917 by Henry L. Gantt, an American engineer and social scientist. It is a chart in which a series of horizontal lines shows the amount of work done or production completed in certain periods of time in relation to the amount planned for those periods.



### PROBLEM STATEMENT

The Parkinson's disease is due to a loss of neurons that produce a chemical messenger in the brain called dopamine. when there is a decrease in level of the amino acid named dopamine it leads to the abnormal brain activity, which leads to Parkinson’s disease. The cause of Parkinson's disease is still a question mark, but several factors appear to play a role, including:

• Genes

• Environmental

• Triggers

As a result, people suffer from this disease for many years before diagnosis. The estimated results have shown that there are 7-10 million people are affected by Parkinson’s disease (PD) worldwide. People with age above 50 are the one who has the higher possibility of getting Parkinson’s disease but still an estimated 4 percentage of people who are under the age 50 are diagnosed with Parkinson’s disease. There is no cure or prevention for PD. However, the disease can be controlled in early stage. The Machine Learning Classification Algorithm techniques is going to use as an effective way for early detection and diagnosis of the disease. ML techniques in medicine is a research area that combines sophisticated representational and computing techniques with the insights of expert physicians to produce tools for improving healthcare. ML is a statistical method for finding hidden patterns in datasets by constructing predictive or classification models that can be learned from past experience and applied in future cases, so there is a need for a more accurate, objective means of early detection, ideally one which can be used by individuals in their home setting.

* 1. **PROBLEM FORMULATION**

The project aims and objectives that will be achieved after completion of this project are discussed in this subchapter. The aims and objectives are as follows: The aim of this project is to compare various machine learning models in the successful prediction of the severity of Parkinson's disease and develop an effective and accurate model in order to help diagnose the disease accurately at an earlier stage which could in turn help the doctors to assist in the cure and recovery.

**2. HARDWARE/SOFTWARE SPECIFICATION**

### 2.1. HARDWARE SPECIFICATION

PC/LAPTOP:

A laptop, laptop computer, or notebook computer is a small, portable personal computer (PC) with a screen and alphanumeric keyboard Laptops typically have a clam shell form factor with the screen mounted on the inside of the upper lid and the keyboard on the inside of the lower lid, although 2-in-1 PCs with a detachable keyboard are often marketed as laptops or as having a "laptop mode".

|  |  |
| --- | --- |
| **Component** | **Minimum requirement** |
| Processor | 64-bit, four-core, 2.5 GHz minimum per core (If your dataset size is significantly larger than the medium dataset, we recommend 8 cores.) |
| RAM | 16 GB |
| Hard disk | 80 GB |

**2.2. SOFTWARE SPECIFICATION**

**2.2.1 Anaconda /Jupyter Notebook:**

Jupyter Notebook is a web-based open-source application that is used for editing, creating running and sharing documents that contain live codes, visualization, text, and equations. Its core supported programming languages are Julia, R, and Python. Jupyter notebook comes with anIPython kernel that allows the programmer to write programs in python. There are over 100 kernels other than IPython available for use.

**2.2.2 Python:**

A high-level, all-purpose programming language is Python. Code readability is prioritized in its design philosophy, which makes heavy use of indentation. Python uses garbage collection and has dynamic typing. It supports a variety of paradigms for programming, including functional, object-oriented, and structured programming.

**3. LITERATURE SURVEY**

This section describes the theoretical background of this project, starting with an explanation of Parkinson’s disease, followed by overviews of machine learning, deep learning, related work and finally Parkinson’s diagnosis (PD) problems. The detection of PD is extremely important at the first stage. The detection can be performed using ML technique.

Jie Mei et al. used all basic algorithms of Machine learning techniques for the detection of PD. Like SVM, RF, Decision Tree, ANN, KNN, Radial Basis Function Networks (RBF) and Deep Belief Networks (DBN) etc. The early identification of Parkinson's disease is critical. The identification can be performed with the use of a data mining technique. The techniques for detecting PD, such as Naive Bayes, support vector machine, multilayer perceptron neural network, and decision tree, are theoretically explained in this study. This study uses speech input from acoustic devices to predict Parkinson's disease. People from various areas and speech factors are investigated in this article in order to predict Parkinson's disease among patients.

Gabriel Solana-Lavalle et al . uses the algorithms such as Multilayer Perceptron (MLP), Random Forest (RF), K-Nearest Neighbor (KNN). For the prediction of Parkinson disease, three set of experiences were conducted to obtain the features with highest contribution to PD. This three sets are 1. a population with male and female subjects (balanced), 2 male subjects (balanced and unbalanced), and 3. Female subjects (balanced and unbalanced). In this study, the researchers used acoustic devices to collect speech parameters from 50 persons with Parkinson's disease and fifty healthy people. They employed the k-fold cross validation method for testing and claim that it can deliver good accuracy.

Kazi Amit Hasan et al . used different classification methods RF, KNN, Decision Tree, Logistic Regression (LR), SVM, and Naïve Bayes for detection of PD. The best result achieved by Decision Tree and Random Forest (RF) classification methods. The data mining techniques may be a more popular in many field of medical, business, railway, education etc. They are most commonly used for medical diagnosis and disease prediction at the early stage. The data mining is employed for healthcare sector in industrial societies.

Shail Raval et al. For the detection of PD they include all the aspects such as biological data, chemical data and genetic data. In this paper they mainly focused on the symptoms like rigidity, Tremor at rest, changing voice etc. The secure data transmission is proposed through authentication check, duplication check and faulty node detection. The proposed method is applicable to long ranges of transmission. It is also supporting a retransmission concept.

Mosarrat Rumman et al. based on Image Processing and Artificial Neural Network (ANN) classification algorithm According to ANN prediction, if value closer to 1 then suggests PD and value closer to 0 then suggest normal. Parkinson disease is a global public health issue. Machine learning technique would be a best solution to classify individuals and individuals with Parkinson's sickness (PD). This paper gives an entire review for the forecast of Parkinson disease by utilizing the machine learning based methodologies. A concise presentation of varied computational system-based methodologies utilized for the forecast of Parkinson disease are introduced. This paper likewise displays the outline of results acquired by different scientists from accessible information to predict the Parkinson disease

Prediction of Parkinson disorder is one of the most important problems that has to be detected in the early phases of the commencement of the disease so as to reduce the disease progression rate among the individuals. Various researches have been made to find the basic cause and some have reached to the heights by proposing a system which differentiates the healthy people from those with any ND’S (Neurodegenerative disorders) using various machine learning techniques. Lots of pre-processing feature selection and classification techniques have been implemented and developed in the past decades.

**4. METHODOLOGY**

In the prediction of diseases, ML techniques have proven to be effective This study uses both unsupervised and supervised learning techniques to diagnose PD through classification prediction. Several approaches that entail clustering, reducing dimensionality, and learning of prediction approaches are used to create the PD diagnoses method. depicts the proposed method with its main stages.

Data pre-processing, dimensionality reduction using random forest using ensembles of EM (Expectation -Maximization), and prediction using ensembles of all stages of the method that are utilized to predict disease through a set of real-world PD data

**5. RESULT**

ML has been used for medical disease detection lately and particularly Parkinson’s disease (PD) treatment. This can be explained by the convenient performance and accurate results of ML techniques. Classification of diseases is a significant type of predictive modelling. It is considered an important data mining approach because it clusters the population referring to a predetermined criterion. It is vital to compare the outcomes of various classification methods to decide which approach presents the best performance. Hence, the main goal of this research is to assess several approaches that are utilized for PD prediction and classification. Even though ML methods have been assessed in several studies separately, the evaluation of these methods based on various datasets makes it complex to perform an accurate comparison among the deployed methodologies. Hence, it is vital to evaluate these methods in one comparative study based on a chosen dataset

#### 6. CONCLUSION AND FUTURE WORK

### 6.1. Conclusion

### Parkinson’s disease affects the CNS of the brain and has yet no treatment unless it’s detected early. Late detection leads to no treatment and loss of life, so its early detection is significant. For early detection of the disease, we utilized machine learning algorithms such as Classification Random Forest. We checked our Parkinson disease data and find out Classification Random Forest will be the best Algorithm to predict the disease which will enable early treatment and save a life.

### 6.2. Future SCOPE

##### In future work, we can focus on different techniques to predict the Parkinson disease using different datasets. In this research, we using binary attribute (1- diseased patients, 0-non-diseased patients) for patient’s classification. In the future we will use different types of attributes for the classification of patients and also identify the different stages of Parkinson's disease.

**7. REFERENCES**

1. Templeton, J. M., Poellabauer, C. & Schneider, S. Design of a neurocognitive digital health system (NDHS) for neurodegenerative diseases. in Proceedings of the 2021 Workshop on Future of Digital Biomarkers 26–33, https://doi.org/10.1145/3469266.3471157 (2021)
2. Far, M. S., Eickhoff, S. B., Goni, M. & Dukart, J. Exploring test-retest reliability and longitudinal stability of digital biomarkers for Parkinson disease in the m-power data set: Cohort study. J. Med. Internet Res. 23, e26608. https://doi.org/10.2196/26608 (2021).
3. Waring, J., Lindvall, C. & Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif. Intell. Med. 104, 101822. https://doi.org/10.1016/J.ARTMED.2020.101822 (2020).
4. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A. & Escobar, G. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health 33, 1123–1131. https://doi.org/10.1377/HLTHAFF.2014.0041 (2017).
5. Marella, W. M., Sparnon, E. & Finley, E. Screening electronic health record-related patient safety reports using machine learning. J. Patient Saf. 13, 31–36. https://doi.org/10.1097/PTS.0000000000000104 (2017).
6. Deng, K. et al. Heterogeneous digital biomarker integration out-performs patient self-reports in predicting Parkinson’s disease. Commun. Biol.https://doi.org/10.1038/s42003-022-03002-x (2022).