# PARKINSON'S DISEASE PREDICTION USING MACHINE LEARNING IN PYTHON

[1] Dr.S.Kavitha M.Sc., M.Phil, Ph.D., Head of the Department,,

[2] M. Anusri, M.Sc (Computer Science),

[1] [2] Department Of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram.

## ABSTRACT

Around the globe, thousands of people worldwide are suffering by Parkinson's Disease (PD), a central nervous system degenerative condition. Early detection and diagnosis of PD is crucial for successful treatment and management of the disease. In past few years, Machine learning (ML) algorithms has shown great potential in predicting PD based on various physiological and neurological markers. In this disease prediction system, a system is proposed using ML-based approach to predict the presence of PD in patients. The system employs various machine learning models, including Gradient Boosted Tree, random forest, and logistic regression, to identify key markers and patterns associated with the disease. Overall, this disease prediction system provides a valuable tool for early detection and diagnosis of PD, which can lead to better management and treatment of the disease. The proposed approach can also be extended to other neurological disorders, providing a general framework for disease prediction and diagnosis.

## I INTRODUCTION

In a person who experiences health, bodily functions are working at peak performance, due not only to a lack of disease, but also to regular exercise, balanced nutrition, and adequate rest. We receive treatment, when necessary, to maintain the balance. Physical wellbeing involves pursuing a healthful lifestyle to decrease the risk of disease. Maintaining physical fitness, for example, can protect and develop the endurance of a person's breathing and heart function, muscular strength, flexibility, and body composition. Physical health and well-being also help reduce the risk of an injury or health issue. Examples include minimizing hazards in the workplace, practicing safe sex, practicing good hygiene, or avoiding the use of tobacco, alcohol, or illegal drugs.

## II LITERATURE SURVAY

Mohammad S Islam.Et Al Conducted A Comparative Analysis To Detect Parkinson's Disease Using Various Classifiers. Support Vector Machine (SVM), Feedforward Back-Propagation Based Artificial Neural Network (FBANN) And Random Tree (RT) Classifiers Were Used And A Comparison Between Them Is Made To Differentiate Between PD And Healthy Patients. The Study Has Utilized The UCI Machine Learning Repository. The Dataset Consists Of 195 Voice Samples From 31 Individuals Comprising Of Both Males And Females. From The Taken Subjects 23 Were Determined With PD And 8 Were Healthy. To Improve The Classification Accuracy With Minimal Error Rate A 10-Fold Cross Validation Which Was Repeated 100 Times Has Been Implemented For All The Three Classifiers. The FBANN Classifier Has Achieved A 97.37% Recognition Accuracy Thus Outperforming The Other Two Classifiers.

R. Arefi Shirvan.Et Al Proposed A System For Detect PD. The Data Classification Was Done Using Knn Method. Simplest Method In Grouping The Similarity Is KNN. Among Classification Method KNN Is Used Whenever The Facts For Data Distribution Are Not Enough. In This Method It Has Two Parts: A) Determine K Close Neighbors, B) Determining Class Type Using These Close Neighbors. It Was Shown That A 93.7% Of Accuracy Per 4 Optimized Features, An Accuracy Of 94.8% Per 7 Optimized Features And 98.2% Accuracy For 9 Optimized Features Is Achieved Which Is A Remarkable Result Compared To Other Studies. In This Research Data From [6] From UCI Repository Is Used. The Data Include 192 Voice Sample Recordings From 32 Male And Female. Each Subject Has Had 6 Voice Signal Recordings. 23 People Suffer From PD And The Rest Are Healthy. People Were About 46-85 Years Old Themain Disadvantage Of The KNN algorithm Is That It Is A Lazy Learner, I.E. Classification Is Done By Using Training Data And From The Training Data It Doesn't Learn Anything.

A Comparative Study on Tumor Classification research done by K. Srilatha and V. Ulagamuthalvi.Et Al, classification is an important task within the field of computer vision. Image classification refers to the labelling of images into one of a number of predefined categories that includes image sensors, image pre-processing, object detection, object segmentation, feature extraction and object classification. Many classification techniques have been developed for image classification. The highest concentration is on using various classifiers combined with several segmentation algorithms for detection of tumor using image processing.

S. R. Khonde & V. Ulagamuthalvi.Et Al proposed Feature selection used is based on average probability score of each feature. The features having less AP score are removed from the set used for training and testing classifiers. Performance parameters used by authors are true positive, true negative and accuracy. Authors make use of various semi-supervised classifiers for intrusion detection. All classifiers used NSL KDD dataset for intrusion detection. Resul Das .Et Al Used Various Classification Methods For Identifying PD. Four Different Classification Techniques Were Implemented And Analyzed And They Are Dmneural, Neural Networks, Regression, And Decision Tree Respectively. Various Evaluation Methods Were Used For Calculating The Performance Of The Classifiers. After Evaluation From The Results, The Neural Networks Classifier Yielded The Best Results. The Input Dataset Was Randomly Partitioned Into Train And Test Dataset. 65% Of The Input Dataset Was Used For Training And The Rest Of The Data-Set Was Used For Testing. The Adjustable Parameters Of Each Classifier Were Tuned. For The Neural Networks Classifier, The BPNN Algorithm Has Been Used In The Feed-Forward, Single Hidden Layer Neural Network. The Algorithm Used In The Study Is The Levenberg– Marquardt (Lm) Algorithms. The Neural Network Has 92.9% Accuracy.

Mercy Paul selvam.Et Al used a machine learning techniques for predicting student dropout using data mining. In this model decision tree was used to predict the dropout in student and they obtained an accuracy with 97.69% and the prediction was done by using various parapets, which are considered for every student.

Dr. R. Geetha Ramani.Et Al Has Proposed A System To Classify PD And Non- PD Patients By The Following Methods Binary Logistic Regression, Linear Discriminant Analysis (LDA), Partial Least Square Regression (PLS) , Random Tree (Rnd Tree) And Support Vector Machine (SVM). The Dataset Is The Parkinson’s Disease Data Acquired From The UCI Repository. The Training Dataset Comprises Of 197 Samples With 22 Features Extracted From The Patients. Fisher Filtering Feature Selection Algorithm Was Found To Be An Effective Feature Ranking System. The Rnd Tree Algorithm Achieved 100% Classification Accuracy While The Lda, C4.5, Cs-Mc4 And K-Nn Yielded Accuracy Results Greater Than 90%. Among All, The C-Pls Algorithm Achieved The Least Accuracy Of 69.74%.

# III THEORETICAL BACKGROUND

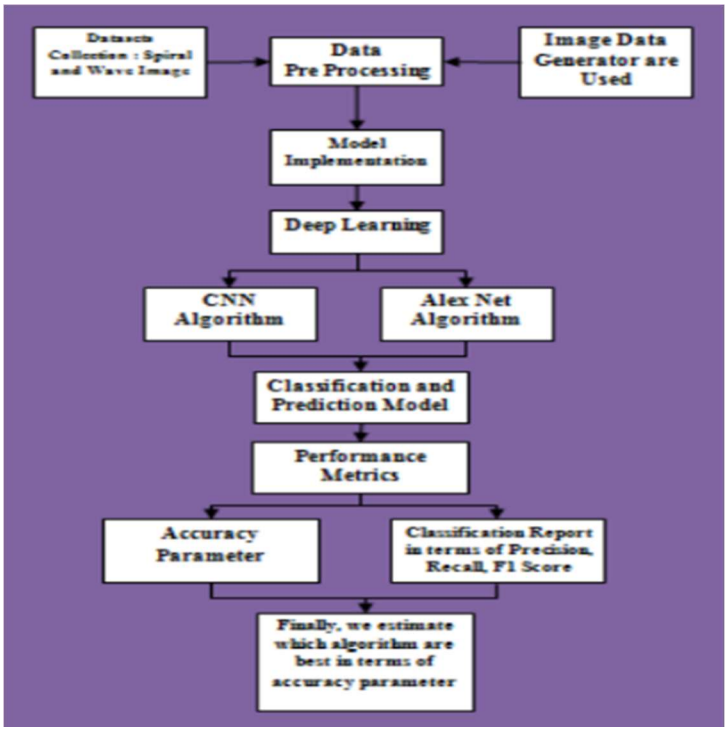
## 3.1 PROBLEM IDENTIFICATION

- Existing systems for predicting Parkinson's Disease primarily focus on leveraging machine learning techniques to analyze patient data and detect early symptoms of the disease. Uses speech signal features to detect changes in voice patterns commonly affected in Parkinson's Disease. Tools like digitized tablets capture these data points for analysis. Machine learning models classify these patterns to predict the likelihood of Parkinson's.

## 3.2 PROBLEM SOLVING

- The suggested methodology covered how our system would be implemented. In our proposed approach, we develop a deep learning method to identify Parkinson's disease and predict the disease's severity. The field of study for Parkinson's disease is important because systemic health of patients can be enhanced by early diagnosis. Here, depending on the patient's healthy and unhealthy conditions, we use the five modules to predict the disease. The first module is a dataset gathering founded on spiral as well as wave images of system patients who are healthy and unwell. The data preprocessing stage in the second module uses an image data generator technique. The deep learning algorithm's model implementation is covered in the third section. Here, CNN and Alex net are the two algorithms we use. The classification model, which is the module after that, uses the system's healthy and unhealthy conditions to forecast disease. The system's accuracy score is used to compute the performance metrics. In order to determine which algorithm is optimal to employ in terms of the accurate measure we can estimate for our system, we first evaluate both computational intelligence accuracy scores. Results of the experiments demonstrated the system's improved efficiency. Here, we use a few key components in our suggested approach to build our system.

### 3.3 SYSTEM ARCHITECTURE




## IV SYSTEM IMPLEMETATION

### 4.1. MODULE:

- Importing the dependencies
- Data collection and analysis
- Data processing
- Separating training and testing data
- Data standardisation
- Model training
- Model evaluation
- Building a predictive system

### 4.2 MODULE DESCRIPTION:

- **Dataset Collection:** In order to use analysis of information to discover recurring patterns, data collection enables you to record a record of prior occurrences. You can create forecasting models using deep neural network algorithms that search for trends and forecast future changes based on those patterns. Here, we employ Pictures as a dataset for this assignment. In the context of Category 0 and Category 1, there are two categories of data. A wave collection of healthy as well as unhealthy data is explained for mode 0. The spiral

collection of the system's both positive and negative data is explained for Mode 1. The system's Parkinson patient data can be used to create either Model 0 as well as Model 1 databases.

- **Data preprocessing:** Information pre-processing, which entails transforming unorganized data into a more organized structure, may be a part of information mining. Data pre-processing is a technique used in information extraction to transform unusable data into something useful and practical. The method known as the Image Database Generator technique is employed here. In the field of real-time data augmentation, Picture Data Generator is used to create groups that contain data from tensor images. By giving the appropriate settings and the necessary input to the Picture Data Generator resize class, we can use it. To modify the values of pixels from a possible range of 0 to 255 to the range 0-1 recommended for neural network models, use the Image Dataset Generator class. Normalization is the term used to describe scaling data to a 0–1 range.
- **Splitting of Dataset:** When database is separated into multiple categories, this is known as data splitting. A two-part split usually consists of developing the model in one part and assessing or analysing the data in the other. Data isolation is an essential component of data science, particularly when creating models from data. The simplest method to divide algorithmic knowledge into sets utilized for testing as well as training is to assign a majority of the information elements to the training collection and the remaining 1/3 to the testing set. As a consequence, we train the algorithm on the training set before deploying the model that was learned on the test set. This enables us to evaluate the performance of our algorithm.
- **Model implementation:** Here, we are utilizing deep learning models for this endeavour. In deep learning, the CNN algorithm and the the Alex net algorithm model are used to forecast the occurrence of Parkinson's disease based on the normal and abnormal states of two datasets, such as the spiral as well as wave datasets of acquired images.
- **Classification:** Here, we categorize the information based on two datasets of spiral as well as wave of Parkinson's disease patients using the CNN and Alex net algorithms. The user can input any type of image and based on the system's healthy and unhealthy conditions for Parkinson's disease, the end output will display which type of patient. It can be obtained from both method outputs based on two datasets, such as the system's spiral along with wave of Parkinson's data.
- **Prediction:** The forecast will be made using the Parkinson disease prognosis of a healthy or unwell patient. Finally, the system's classification report and accuracy score are used to compute the success metrics. As a conclusion to our article, we determined which algorithms are best in regard to accuracy score and evaluated them against the system as a whole.

# V CONCLUSION & FUTURE WORK

## 5.1 CONCLUSION

In order for predicting the severity of Parkinson's disease, we have employed a network of convolution neural networks in this project. When compared to other methods, the suggested Neural and Alex net models both had higher accuracy. Additionally, the categorization of Parkinson's disease based on two datasets the spiral and wave was discovered. We looked into using deep learning to recognize Parkinson's disease. The simulated workflows were created to examine how well the models trained with deep learning performed on a range of datasets using different deep learning techniques. The processing of some data is the first step in these workflows, which then move on to the classification job and the best results collection. As a result, we came to the conclusion that both algorithm accuracy values were high for the Alexnet algorithm, which demonstrated superior performance and high system performance.

## REFERENCE

[1] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman and L. O. Ramig, "Novel Speech Signal Processing Algorithms for HighAccuracy Classification of Parkinson's Disease," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264- 1271, May 2012.

[2] Abós, A. et al. Discriminating cognitive status in Parkinson’s disease through functional connectomics and machine learning. Scientific reports 7, 45347 (2017).

[3] B. Zhang, J. Ren, Y. Cheng, B. Wang and Z. Wei, "Health Data Driven on Continuous Blood Pressure Prediction Based on Gradient Boosting Decision Tree Algorithm," in IEEE Access, vol. 7, pp. 32423-32433, 2019.

[4] C. B. Levine, K. R. Fahrbach, A. D. Siderowf, R. P. Estok, V. M. Ludensky, and S. D. Ross, “Diagnosis and treatment of Parkinson’s disease: a systematic review of the literature,” Evid. Rep. Technol. Assess. (Summ.), no. 57, pp. 1–4, May 2003.

[5] Center for Machine Learning and Intelligent System, Website:http://Archive.Ics.UCI.Edu/ML/Datasets/Parkinsons,2009.

[6] Dobson, A. J. & Barnett, A. An introduction to generalized linear models. (CRC press, 2008).

[7] Dr. Geetha R Ramani, G Sivagami,”Parkinson Disease Classification Using Data Mining Algorithms”, International Journal Of Computer Applications 32(9):17-22, October 2011.

[8] Gladence, L. Mary, M. Karthi, and V. Maria Anu. "A statistical comparison of logistic regression and different Bayes classification methods for machine learning." ARPN Journal of Engineering and Applied Sciences 10, no. 14 (2015): 5947-5953.

[9] H. Ramezani and O. B. Akan, “Rate region analysis of multi-terminal neuronal nanoscale communication channel,” in 17th IEEE NANO Conf. IEEE, 2017.

[10] J. Jankovic, "Parkinson’s disease: clinical features and diagnosis" Journal of Neurology, Neurosurgery & Psychiatry, pp.368-376, 2008.