INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 754-758

EARLY-STAGE PREDICTION OF PARKINSON DISEASE USING MACHINE LEARNING APPROACHES

Lamkhade Anil[1], Malgunde Mahesh[2], Nachan Rushikesh[3], Unde Suvarna[4]

[1,2,3]Student, Computer Engg., RGCOE Ahmednagar, India.

[4]Asst. Prof., Computer Engg., RGCOE Ahmednagar, India.

ABSTRACT

Parkinson's disease (PD) is a neurodegenerative movement disorder in which symptoms develop gradually, starting with a mild tremor in one hand and a feeling of body stiffness and worsening over time. It affects more than 6 million people worldwide. Currently, there is no conclusive result for this disease from non-specialist doctors, especially in the early stage of the disease, where the identification of symptoms in its early stages is very difficult. The proposed predictive analytics framework is a combination of K-means clustering and a decision tree that is used to extract information from patients. Using machine learning techniques, the problem can be solving minimum error rate. Voice data files obtained from the UCI Machine learning repository, if provided as input for voice data analysis. Our proposed system also provides accurate results by integrating spiral drawings of normal and Parkinson's patients. A random forest classification algorithm is used from these drawings which converts these drawings into pixels for classification and the extracted values are compared with a trained database to extract different features and the results are produced with maximum accuracy. also OpenCV (Open Source Computer Vision Library) a library of programming functions focused primarily on real-time computer vision was built to provide an infrastructure for computer vision applications and accelerate the use of real-time machine perception. Our output will thus show early detection of the disease and may be able to prolong the life of the sick patient with proper treatment and medication leading to a peaceful life. Artificial intelligence (AI) has played a promising role in PD diagnosis. However, it introduces bias due to insufficient sample size, poor validation, clinical evaluation, and lack of big data configuration.

Keywords: Parkinson’s disease; exploratory data analysis; coefficient of variation; t-SNE; REF; machine learning

1. INTRODUCTION

Parkinson's disease (PD) is an idiopathic disease of the nervous system characterized by both motor and non-motor manifestations. It is a chronic progressive neurodegenerative disorder that occurs mostly in the elderly, but can also occur in much younger patients. It is the second most common neurodegenerative disease (1). Other neurodegenerative disorders can mimic idiopathic PD. These include dementia with Lewy bodies (DLB), corticobasal degeneration (CBD), multiple system atrophy (MSA), and progressive supranuclear palsy (PSP). The main focus of this review will be idiopathic PD and not these other parkinsonian syndromes. A population-based study of US Medicare beneficiaries found an average prevalence of 1.6% of PD in persons aged 65 years and older. Fewer blacks and Asian Americans are affected than whites. Higher rates of PD exist in the Midwest/Great Lakes region northeast coast of USA.

2. LITERATURE SURVEY

Timothy J. Wroge et al. [9] Data were collected through mPower, a clinical observational study conducted by Sage Bionetworks using an iPhone application. The raw audio is cleaned using VoiceBox before being fed into the feature extraction algorithms Voice Activation Detection (VAD) method. Scikit-Learn was used to build the decision tree and support vector machine classifiers. Several decision tree classifiers, including additional trees, random forests, gradient-boosted decision trees, and normal decision trees, were used to categorize the dataset.

Muhtasim Shafi Kader in el.[4] Mushtasim Shafi Kader at el. [] selected 195 datasets related to Parkinson's disease from the UCI machine learning library to identify Parkinson's disease. There were 24 attributes in the given dataset. After training data, they were able to identify the machine learning algorithms that were most accurate. Naive Bayes, Adaptive Boosting, Bagging Classifier, Decision Tree Classifier, Random Forest Classifier, XBG Classifier, K Nearest Neighbor Classifier, Support Vector Classifier and Gradient Boosting Classifier were nine machine learning algorithms that were used to predict the disease. The analysis of evaluation metrics and the analysis of confusion metrics (Precision, Recall, F measure and Accuracy) were used for the calculation. study results. The algorithm with the highest accuracy was found using the above metric analysis.

Mohesh T et al.[5] The input is the Parkinson's disease voice dataset from the UCI Facility Management Library. In addition to that By combining spiral drawings of healthy individuals and Parkinson's patients, this gadget provides accurate findings. It can be deduced that the hybrid approach accurately reads disabled individuals' spiral drawings

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 754-758

and voice data. This model aims to do that method of expertise in the case of Parkinson's disease, so the goal is to use numerous machines to learn strategies like SVM, tree selection, to purchase the most accurate result.

Basil K Varghese at el. [10] used the UCI ML repository to access the dataset. Data dimensionality was then reduced using principal component analysis. They used SVM (Support Vector Machine), decision trees, linear regression and neural networks to predict values from the test data set. Accuracy was then determined by using a training model to predict values from a test data set. The dataset used in this study consists of characteristics extracted from voice recordings of 42 individuals who they were diagnosed with early-stage Parkinson's disease.

Srishti Grover et al. [11] He proposed a method of applying deep learning to predict the severity of Parkinson's disease. In the first phase, voice recordings from PD patients are obtained for analysis. The obtained data are then normalized using min-max normalization. After obtaining the data, the model performs training, evaluation and prediction. A deep neural network containing an input layer, hidden layers and an output layer is created. And finally, an evaluation is performed on the resulting DNN classifier.

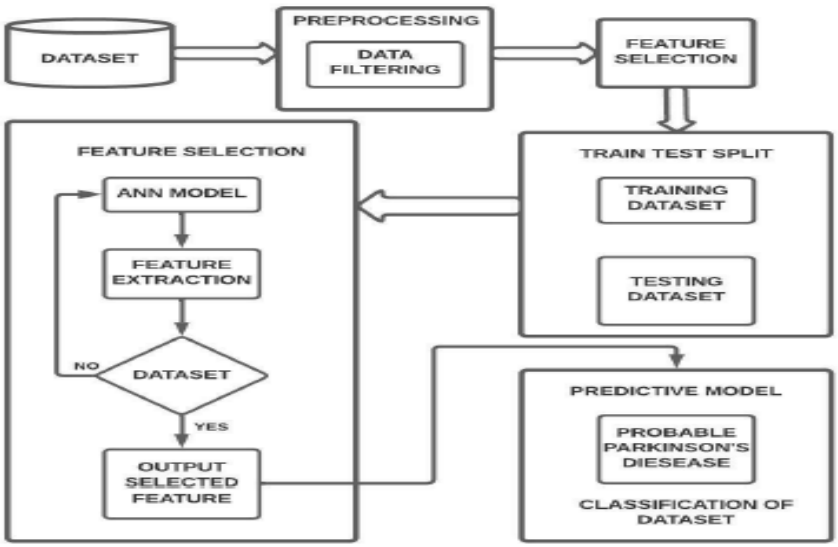


Fig 1. Block Diagram of Proposed Methodology

Anitha R at el. [9] proposed a predictive analytics system that uses K-means clustering and a decision tree to extract information from patients. This specific study uses the Parkinson's UCI Machine Learning library as input. The proposed approach also produced accurate results by combining the spiral drawings of Parkinson's patients and healthy subjects. They proposed a hybrid methodology that created a technique that detects after evaluating the patient's speech and spiral drawing data. The drawings were converted to pixels using the Random Forest classification technique, and the extracted values were compared with the training database to generate different characteristics.

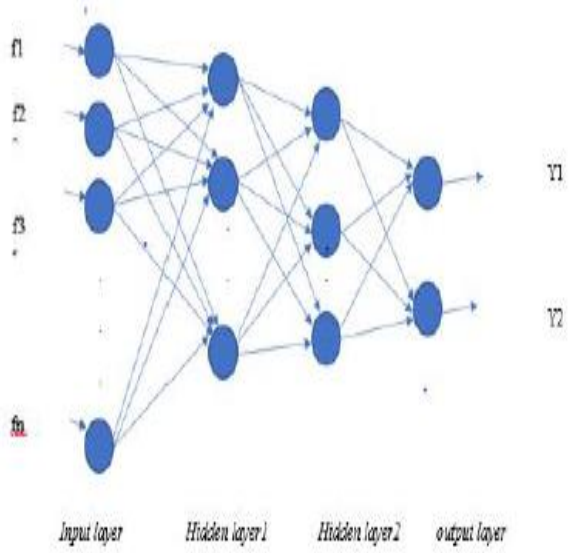


Fig.2 ANN

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

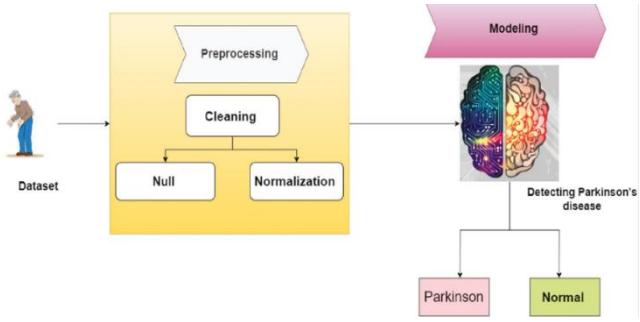
Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 754-758

3. PROPOSED MODEL

Input: The first step is data collection. This step is extremely important because the standard and amount of information you collect will directly affect the scope of your predictive model. So we took the data of various voice recordings of the patient.



Data Preprocessing: In this step, the information is well visualized to identify the relationship between the parameters present in the data and also use to get the data imbalance. With this we need to split the information into two parts. The first part for training the model as in our model we used 70 percent knowledge for training and 30 percent for testing.

Feature Selection: The next step in our workflow is feature selection. There are various models that have been used by researchers and scientists so far. Some are intended for image processing, some for sequences such as text, numbers or patterns. In our case, we defined samples of PD patients from different patients, so we chose such models that will classify or differentiate an unhealthy patient from a healthy one.

Training: Training a dataset is one of the main tasks of machine learning. We will apply the data to Gradual improvement of the ability of the selected model to predict better, i.e. the actual result should be approx. predict one.

Neural Network Model:

ANN BLACK PROMOTION: Artificial neural networks are a special kind of machine learning algorithms that are modelled after the human brain. This means that a bit like how the neurons in our nervous system are primed to learn from past data, an ANN is able to learn from information and provide answers in the form of predictions or classifications. ANNs are non-linear statistical models that depict a fancy relationship between inputs and outputs to obtain a surrogate pattern. Various tasks such as image recognition, speech recognition, MT also as diagnostics use these artificial neural networks.

Feature Extraction: The metrics we have calculated are ROC, Accuracy, Specificity, Precision etc. which will highlight the simplest algorithm of all.

Prediction: In this phase, we will finally prepare a model to detect the prediction of Parkinson's disease based on the given data set.

Matrix of Confusion

The permutation matrix is the most accurate matrix used to determine the precision of a accuracy of models. Used for binary categories and multilevel problems. Describes the performance of class models where the true values are already known. The a confusion matrix is a table with two dimensions, one of the actual target value and one of the predicted values. To explain the concept of confusion matrix, consider the binary partitioning problem where the classes 1 and 0 are shown in the figure.

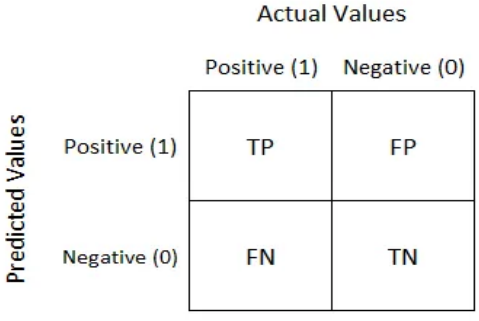


Fig. 3 Confusion matrix

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 03, March 2024, pp: 754-758

4. MODULE DESCRIPTION

The PD voice dataset is collected from the UCI machine learning repository and stored in the RStudio environment as test and training datasets. These are saved in the RStudio environment as test and training datasets. R is a programming language and software environment for statistical analysis, graphical representation, data analysis, and also machine learning. It includes the following steps and procedures.

1.Import data into RStudio - arrange the data in the CSV FILE to include the column names in the first row (i.e. a person's voice collected in different time zones) and each subsequent line contains all the information (i.e. a set of 22 parameters is considered and the range of a person's voice for these parameters is tested and then recorded), finally the status column shows two values 0 (healthy) and 1 (affected). Import the data into RStudio using the Import data command.

2. Clustering– an unsupervised learning algorithm that attempts to cluster data based on their similarity and only tries to find patterns in the data. Here we have to specify the number of clusters we want to group the data into, and then the algorithm randomly assigns each observation to a cluster, finds the centroid of each cluster, and then iterates by reordering the data points to the cluster whose centroid is closest and calculates the new centroid of each cluster.

3. Classification - also called prediction tree. It uses structure to specify sequences of decisions and consequences, the goal being to predict a response or output. Prediction can be done by creating a decision tree with test points and branches. At each checkpoint, it decides to select a specific branch and cross trees, and can be used in different disciplines based on individual characteristics.

4.Predicted output – Predicted output for voice analytics based on clustering and classification is 88% accurate.

Performance metrics

The suggested methods for detecting PD is evaluated using performance measures, including accuracy, recall, F1-score, and precision.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$\text{Fscore} = \frac{2*preision*\text{Sensitivity}}{preision + \text{Sensitivity}} \times 100\%$$

5. CONCLUSION

In this paper, the concepts of deep learning are discussed, while the application in the prediction of Parkinson's disease is focused. A literature survey was conducted on Parkinson's disease.

To avoid the disadvantages of conventional methods, new deep learning techniques are proposed in this review. The included studies showed that deep learning techniques have a significant impact on the early detection of Parkinson's disease with a high degree of accuracy. However, most of the proposed methods are still under development and have not been tested in a clinical setting. In this paper, the work mainly focuses on the development of predictive models to achieve good accuracy in predicting valid disease outcomes using deep learning methods such as artificial neural network (ANN) based prediction. In this paper, deep learning techniques are proposed to predict early-stage Parkinson's disease. The main goal of this work is diagnostics using the analysis of voice signals. For many years now, speech processing has had incredible potential as a voice measurement. This work is intended to analyse performance classification algorithms. Innovation (ETJRI) disease research detection in the patient's life better.in results. As part of our identification is mastered using learning models is to show that PDs in PD detection are non-invasive.

6. REFERENCES

[1] Jie Mei, Christian Desrosiers and Johannes Frasnelli, “Machine Learning for the Diagnosis of Parkinson’s Disease” Front Aging Neuroscience, 2021.

[2] Amreen Khanum D, Prof. Kavitha G and Prof. Mamatha H S, “Parkinson’s Detection using MachinLearning Algorithms”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321- 9653 vol 10 Issue X, pp 786-790, Oct 2022.

[3] Gabriel Solana-Lavalle, “Analysis of voice as an assisting tool for detection of Parkinson’s disease and its subsequent clinical interpretation”, ELSEVIER, Biomedical Signal Processing and Control, 2021.

[4] Muhtasim Shafi and Fizar Ahmed, “Parkinson’s Disease Detection Analysis through Machine Learning Approaches” https://www.researchgate.net/publication/359711136, 2022.

[5] Mohesh T,Gowtham K,Vijeesh P and Arun Kumar S, “Parkinson’s Disease Prediction Using Machine Learning”,IJRASET44075,ISSN:2321-9653,vol 10 Issue VI,2022.

[6] Oduntan Ifeoma, “Prediction of Parkinson's Disease Using Biomedical Voice Measurements Dataset”, https://www.researc hgate.net/publication/35725620,2021.

[7] Sonia Singla, “Parkinson disease onset detection Using Machine Learning”, https://www.analyticsvidhya.com/blog/2021/07/parkinson-disease-onset detection-using-machine-learning/,2022.

[8] Pramanik Anik and Sarker Amlan, “Parkinson's Disease Detection from Voice and Speech Data Using Machine Learning”, https://www.researchgate.net/publication/347520593,2020.

[9] Anitha R,Nandhini T, Sathish Raj S and Nikitha V, “Early Detection Of Parkinson’s Disease Using Machine Learning”,IJARIIE,ISSN(O):2395-4396, Vol 6 Issue 2, pp 505-511,2020. Journal of Neuro Engineering and Rehabilitation, ISSN: 1743-0003, vol 17, Issue 1,2020.

[10] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi, “Parkinson’s Disease Diagnosis Using Machine Learning and Voice”,IEEE Signal Processing in Medicine and Biology Symposium (SPMB), ISSN: 2372- 7241, pp. 1-7, 2018.