https://doi.org/10.46335/IJIES.2024.9.1.2 **e-ISSN: 2456-3463**
Vol. 9, No. 1, 2024, PP. 9-12

International Journal of Innovations in Engineering and Science, www.ijies.net

Parkison's Disease Prediction Using Machine Learning: A Review

Krutik Dharange[1], Amod Polkamwar[2], Yash Tarale[3], Om kute[4], Akhil Bisen[5]

[1, 2,3,4,5] *Department of Computer Technology, Priyadarshini College of Engineering, Nagpur.*

krutikdharange383@gmail.com

***Received on**: 15 February, 2024* ***Revised on**: 20 March, 2024* ***Published on**: 24 March, 2024*

***Abstract –** The majority of studies in this decade of rapid advancements in the medical sciences neglect to address age-related illnesses. These are illnesses that almost never fully recover and only show symptoms at a much later period. The second most frequently identified neurodegenerative brain illness is Parkinson's disease (PD). It may be argued that the patients suffer excruciating pain and that the condition is nearly incurable. All of them demonstrate how urgently effective, trustworthy, and broad diagnostic methods for Parkinson's disease are needed. For a problem this serious, the diagnosis must be auto mated in order to produce accurate and trustworthy results. Speech measurements and indicators are crucial in predicting Parkinson's disease (PD) because the majority of PD patients exhibit some form of speech impairment or dysphonia. The goal of the paper is to compare different machine learning model to predict the Parkinson's disease. This will enable the development of an accurate and efficient model that will aid in the early diagnosis of the condition, which will in turn enable medical professionals to aid in the treatment and recovery of PD patients. We intend to use the Parkinson's dataset, which was obtained from Kaggle, for the previously described reason.*

Keywords- Parkinson's disease, PD , Parkison's dataset, SVM

I- INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder that causes tremors, bradykinesia, and rigidity, as well as non-motor symptoms such as cognitive impairment and autonomic dysfunction. Early diagnosis and prediction of Parkinson's disease are critical for initiating appropriate therapies, tracking disease progression, and establishing tailored treatment methods. Traditional diagnostic procedures rely on clinical evaluation and observation, which may lack sensitivity and specificity, especially in the early stages of the illness. In recent years, there has been a growing interest in applying machine learning (ML) approaches to Parkinson's disease prediction. ML has the capability to integrate and analyze massive and heterogeneous datasets containing clinical, genetic, and neuroimaging information, allowing for the building of prediction models with greater accuracy and dependability. These models can help doctors identify people who are at risk of developing Parkinson's disease before they exhibit overt symptoms, allowing for earlier intervention and disease management.

We will identify significant issues and limitations in the field by critically examining existing literature, such as dataset heterogeneity, small sample numbers, and model interpretability. This review paper provides a complete summary of the present status of research on machine learning algorithms for Parkinson's disease prediction.It also presents methodology used, including ML algorithms like support vector machines (SVMs), artificial neural networks (ANNs), and random forests (RFs), as well as data pre-treatment and feature selection methods . This intends to shed light on the potential of machine learning technologies to revolutionize Parkinson's disease diagnosis

https://doi.org/10.46335/IJIES.2024.9.1.2 e-ISSN: 2456-3463
Vol. 9, No. 1, 2024, PP. 9-12

and prediction, ultimately leading to enhanced patient care and treatment strategies in neurodegenerative disorders.

II- LITERATURE REVIEW

In recent years, there has been interest in the application of machine learning in healthcare, namely in the areas of disease detection and prediction. Numerous investigations have looked into the prediction of different diseases, including Parkinson's disease, using machine learning techniques.

Sakar et al. (2018) classified Parkinson's disease based on voice attributes using multiple machine learning methods. According to their analysis, out of all the models that were examined, Support Vector Machines (SVM) offered the best accuracy.

Wang, Wu, et al. investigate early detection of Parkinson's disease (PD) using deep learning (DL) and machine learning (ML) techniques.[1] They underline the importance of early diagnosis and suggest a strategy for analyzing different clinical and neuroimaging data using DL and ML algorithms. Using multimodal data fusion approaches, they attain high accuracy rates in differentiating Parkinson's disease patients from healthy controls. The study shows encouraging findings for improving PD diagnosis, but more validation is required for real-world use.

Nissar, Iqra, Mir, and Shaikh [2] provide an overview of machine learning (ML) techniques for detecting and diagnosing Parkinson's disease. The study addresses various machine learning techniques used in Parkinson's disease diagnosis, stressing their significance in leveraging heterogeneous datasets to increase accuracy. The authors' detailed examination highlights the potential of machine learning algorithms in improving Parkinson's disease diagnosis and management. Their evaluation provides useful insights into the present landscape of machine learning-based techniques for Parkinson's disease identification, thereby contributing to ongoing research efforts in the field.

Perumal et al.'s [3] thorough analysis included the function of machine learning in early diagnosis of Parkinson's disease. The author highlighted the importance of feature selection and the choice of machine learning model in achieving high prediction accuracy.By investigating the application of several machine learning algorithms in the integrated data prediction of Parkinson's disease, this study seeks to close this gap.

In order to propose an effective diagnosis system, Chen et al.'s work uses the Fuzzy K-Nearest Neighbor algorithm .[4] The best feature set is first created using Principal Component Analysis, and the FKNN model is then constructed using this feature set. Next, the model is contrasted with methods based on support vector machines. The paper concluded that KNN-based systems perform better than the SVM-based systems,

The study by Hariharan et al. stands out from the others since it suggests a hybrid intelligent system for Parkinson's disease identification and subsequent diagnosis. [5] Model-based clustering (the Gaussian Mixture Model) is used for preprocessing in the proposed system, and Principal Component Analysis, Linear Discriminant Analysis, Sequential Forward Selection, and Sequential Backward Selection are used for feature reduction. Three supervised classifiers—the least square support vector machine, the probabilistic neural network, and the general regression neural network—were used for the classification process.

In the field of machine learning, Resul Das [6] evaluates four distinct categories of classification algorithms to facilitate the diagnosis of Parkinson's disease. A variety of classifiers that can identify the existence of Parkinson's disease (PD) have been modeled in the study using SAS-based software. Decision Tree, Neural Network, Regression, and NN are the classifiers used. The effectiveness of the classifiers was assessed using a variety of evaluation techniques, and the accuracy of NN produced a 92.9% correct classification rate.

A parallel feed-forward neural network structure4 is used in Astrom and Koker's work to predict PD. This was due to the fact that a group of nine neural networks operating in parallel only produced an 8.4% improvement over a single neural network. A rule-based system is utilized to assess the ultimate output of every neural network. Every NN's unlearned data is gathered and sent to the subsequent NN's training set. [7]

The work of Challa K. [8] and collaborators uses non-motor symptoms such sleep disorder, olfactory loss5, and rapid eye movement to predict Parkinson's disease.
Agarwal K et al.'s[9] second study uses extreme machine learning techniques, dubbed Extreme Learning Method, to forecast Parkinson's disease. The prediction scheme is

https://doi.org/10.46335/IJIES.2024.9.1.2 **e-ISSN: 2456-3463**
Vol. 9, No. 1, 2024, PP. 9-12

International Journal of Innovations in Engineering and Science, www.ijies.net

highly dependable due to its straightforward architecture and built-in data learning.

One of the main problems with all the research mentioned above is how tiresome it is to assess PD patients. By allowing cellphones to gather data on Parkinson's disease patients and using that data for prediction, the paper by S.Arora et al addresses precisely that domain. [10]

Sakar et al. did a study to compare several speech signal processing algorithms for Parkinson's disease categorization. [11] They investigated the efficacy of several approaches and proposed the tunable Q-factor wavelet transform as a possible solution. Their comparison research sheds light on the most effective ways for PD classification based on speech signals.

III -PROPOSED METHODOLOGY

The proposed system (Fig 1) analyses and predicts the Parkinson's disease from the Kaggle's dataset.

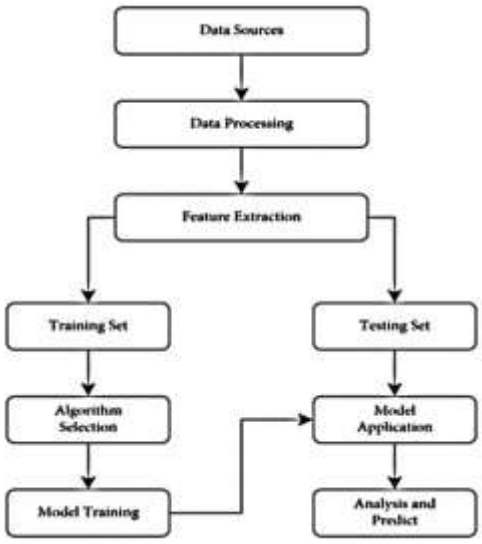

Fig (1):- Proposed Methodology

A. Data Collection

We have collected the Parkinson's disease dataset from the Kaggle website, originally generated by Oxford named as Oxford Parkinson's Disease Detection Dataset which contains 195 records and 24 attributes. The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The original study published the feature extraction methods for general voice disorders

Some Attribute Information

Matrix column entries (attributes):
name - ASCII subject name and recording number
MDVP:Fo(Hz) - Average vocal fundamental frequency
MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
MDVP:Flo(Hz) - Minimum vocal fundamental frequency
MDVP:Jitter(%) MDVP:Jitter(Abs) , MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency
MDVP:Shimmer, MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude

B. Data Processing

The data that we collected, can't be used directly for performing the analysis as this data may be unorganized and may contain a lot of missing vales, duplicates, noisy data and extreme values.The data pre-processing and tuning involves dealing with missing values replacing these values 'NaN', removing the duplicates, correcting outliers if present and handling data, etc. This phase actually involves removing useless and incomplete data and data dimensionality.

C. Feature Extraction

The data will be split into 80% training and 20% testing data as our applied ML algorithms first train themselves from the given training dataset and the use remaining testing dataset to test and predict the outcome.

D. Selection & Training the Model

The next step will be to select the machine learning algorithm and train the model using data collected from step [A].The most commonly used is SVM.

Support Vector Machines (SVM) is a popular machine learning algorithm used in various domains, including medical research, for predictive modeling tasks, such as Parkinson's disease (PD) prediction. SVM is a supervised learning algorithm that can be applied to both classification and regression problems.

E. Evaluation of Model

There are various ways to check the performance of the machine learning algorithms. The model is evaluated based on its accuracy and the confusion matrix. A confusion matrix is a summary of prediction in table format that is used to describe the performance of the model. It is a table with combination of predicted and actual values.

Performance Measures

For the comparison of predictive performance of the applied algorithms 4 evaluation metrics namely accuracy, precision, recall, f1-score are generally used which is given in confusion matrix

Accuracy=(TP+TN)/(TP+FP+TN+FN)
Precision=(TP) / (TP+FP)
Recall= (TP) / (TP+FN)
F1-score: (2*Precision*Recall)/(Precision + Recall)

IV -CONCLUSION

In conclusion, this review emphasizes how important machine learning is to improve Parkinson's disease prediction, especially with regard to Support Vector Machines and ensemble approaches. Notwithstanding certain obstacles, such as restricted datasets and moral dilemmas, these algorithms exhibit potential for early detection. Future advancements in the subject could come from incorporating deep learning and investigating new biomarkers. In order to fully realize the potential of artificial intelligence in Parkinson's disease prediction, this study emphasizes the necessity for ethical implementation and ongoing collaboration, providing a succinct guidance for academics, clinicians, and policymakers.

REFERENCES

[1] *Wang, Wu, et al. "Early detection of Parkinson's disease using deep learning and machine learning." IEEE Access 8 (2020): 147635-147646.*

[2] *Nissar, Iqra, Waseem Ahmad Mir, and Tawseef Ayoub Shaikh. "Machine Learning Approaches for Detection and Diagnosis of Parkinson's Disease-A Review." 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2021.*

[3] *Perumal, M., Annadurai, M., Kaliaperumal, S., & Mahesh, K. (2020). Design of an intelligent Parkinson disease prediction system using machine learning approaches. Expert Systems with Applications, 146 ,113192.*

[4] *Chen, H. L., Huang, C. C., Yu, X. G., Xu, X., Sun, X., Wang, G., and Wang, S. J. (2013). An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert systems with applications, 40(1), 263-271.*

[5] *Hariharan, M., Polat, K., and Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease.*

[6] *Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568-1572.*

[7] *Åström, F., and Koker, R. (2011). A parallel neural network approach to prediction of Parkinson's Disease. Expert systems with applications, 38(10), 12470-12474.*

[8] *Challa, K. N. R., Pagolu, V. S., Panda, G., and Majhi, B. (2016, October). An improved approach for prediction of Parkinson's disease using machine learning techniques. In Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference on (pp. 1446-1451). IEEE.*

[9] *Agarwal, A., Chandrayan, S., and Sahu, S. S. (2016, March). Prediction of Parkinson's disease using speech signal with Extreme Learning Machine. In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on (pp. 3776-3779). IEEE.*

[10] *Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K. M., Dorsey, E. R., and Little, M. A. (2015). Detecting and monitoring the symptoms of Parkinson's disease using smart phones: A pilot study. Parkinsonism & related disorders, 21(6), 650-653.*

[11] *Sakar, B. E., Serbes, G., Sakar, C. O., Nordehn, G., & Ozer, M. (2018). A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing, 62, 387-397.*