WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Early Diagnostic of Parkinson's Disease Using Machine Learning

Vadde Veeresh[1] ,Anugu Shreshna[2], Gumpula Swapna[3], Uppala Rahul Raj[4], Mukkara Devi Sree[5]

[1] Assoc.Professor CSE Dept ACE Engineering College Hyderabad, India vaddeveeresh@gmail.com
[4] Student CSE ACE Engineering College Hyderabad, India rahulraj55944@gmail.com
[2] Student CSE ACE Engineering College Hyderabad, India shreshnaanugu@gmail.com
[5] Student CSE ACE Engineering College Hyderabad, India msree519@gmail.com
[3] Student CSE ACE Engineering College Hyderabad, India gumpulaswapna2002@gmail.com

ABSTRACT:

Parkinson's disease is a degenerative neurological disorder that impairs movement. Patients with this condition require expensive and time-consuming periodic monitoring. The relationship between Parkinson's disease (PD) and speech impairment has been extensively studied in recent years in an effort to develop a system for remote patient monitoring as well as an early diagnosis of the condition. Many studies have used signal and speech processing techniques to convert acoustic signals into vector of features which are then mapped into different machine learning algorithms. The result obtained in PD telemedicine studies have shown that the choice of feature extraction techniques and classification algorithms directly influence the accuracy and realiabity of the proposed system. This provides a system to assess the speech disorders in the context of PD using features extracted.

Keywords: decision tree, random forest algorithm, bootstrap samples, overfitting, non-invasive approach

I.INTRODUCTION :

Parkinson’s Disease (PD) is a degenerative neurological disorder marked by decreased dopamine levels in the brain. It manifests itself through a deterioration of movement, including the presence of tremors and stiffness. There is commonly a marked effect on speech, including dysarthria (difficulty articulating sounds), hypophonia (lowered volume), and monotone (reduced pitch range). Additionally, cognitive impairments and changes in mood can occur, and risk of dementia is increased. Traditional diagnosis of Parkinson’s Disease involves a clinician taking a neurological history of the patient and observing motor skills in various situations. Since there is no definitive laboratory test to diagnose PD, diagnosis is often difficult, particularly in the early stages when motor effects are not yet severe. Monitoring progression of the disease over time requires repeated clinic visits by the patient. An effective screening process, particularly one that doesn’t require a clinic visit, would be beneficial. Since PD patients exhibit characteristic vocal features, voice recordings are a useful and non-invasive tool for diagnosis. If machine learning algorithms could be applied to a voice recording dataset to accurately diagnosis PD, this would be an effective screening step prior to an appointment with a clinician.

II. OBJECTIVES :

The objectives of the project on early diagnosis of Parkinson's disease using machine learning are to develop accurate predictive models capable of identifying the disease at its initial stages based on patient data, including clinical observations, demographic information, and possibly biomarkers. These models aim to improved diagnostic accuracy, facilitate timely intervention, and ultimately enhance patient outcomes by enabling early detection and personalized treatment strategies. Additionally, the project seeks to explore the effectiveness of decision tree and random forest algorithms in this context, aiming to optimize their performance through rigorous experimentation and validation against clinical standards.

III. LITERATURE SURVEY

1.Al-Fatlawi A.H., Jabardi M. H., Ling S. H. (2016) proposed an efficient diagnosis system for Parkinson's disease using a deep belief network, achieving an accuracy of 94%, but with a limitation of a small dataset comprising only 195 voice recordings.

2.Lorente, A., & Garcia-Serrano, A. (2018) explored machine learning approaches for identifying Parkinson's disease using voice signal features.

3.Kadam V. J., Jadhav S. M. (2019) developed a feature ensemble learning system based on sparse autoencoders for Parkinson's disease diagnosis, achieving high accuracy (up to 98%) with voice features. However, the study's reliance on a small dataset necessitates further validation with larger and more diverse populations for clinical application.

4.Senturk Z.K. (2020) developed a Parkinson's diagnosis system using voice analysis and machine learning algorithms, achieving an accuracy of 93.84%. However, further validation is needed to confirm its reliability.

5.Fothergill Misbah N. et al. (2020) highlighted the rapidly growing impact of Parkinson's disease globally, emphasizing the lack of a cure and the reliance on symptomatic treatments for management.

IV. PROPOSED SYSYTEM

In this project we are going to use Decision tree and Random Forest.

1. Early Detection and Screening: There is a need for an effective screening process that can accurately detect PD in its early stages, before motor symptoms become severe. This would facilitate timely intervention and management, potentially improving outcomes for individuals with PD.

2. Remote Monitoring and Management: Current methods for monitoring disease progression require frequent clinic visits, which can be burdensome for patients. There is a need for a non-invasive and accessible approach to remotely monitor disease progression over time, enabling personalized interventions and optimizing treatment efficacy. Addressing these challenges requires the development of a machine learning-based system that can analyze voice recordings to accurately detect PD and track disease progression. Such a system would offer a novel and non-invasive approach to screening and monitoring PD, potentially revolutionizing the way the disease is diagnosed and managed.

V. TECHNOLOGY DESCRIPTION

The technology utilized for this project revolves around Python, a versatile programming language renowned for its rich ecosystem of libraries and tools conducive to machine learning and data analysis tasks. Leveraging Python's extensive libraries such as scikit-learn, pandas, and NumPy facilitates efficient data preprocessing, model development, and evaluation. Additionally, visualization libraries like matplotlib and seaborn aid in exploring and communicating insights from the data. Python's flexibility allows seamless integration of various machine learning algorithms, including decision trees and random forests, enabling comprehensive experimentation to optimize model performance. Its open-source nature and vibrant community support further enhance accessibility and collaboration, making Python an ideal choice for advancing research and development in early Parkinson's disease diagnosis.

VI. IMPLEMENTATION

In the implementation phase, decision tree algorithms are deployed to create a hierarchical structure of nodes, where each node represents a feature and each branch represents a decision rule. This structure is trained using patient data to classify individuals as either Parkinson's disease positive or negative based on their feature values. The tree is recursively split at each node based on the feature that maximizes information gain or minimizes impurity, resulting in a series of decision rules that collectively form a predictive model.
Concurrently, random forest algorithms are implemented to construct an ensemble of decision trees. Multiple decision trees are trained on bootstrap samples of the data, with

feature subsets randomly selected at each node to promote diversity among the trees. During prediction, each tree in the
forest independently classifies instances, and the final prediction is determined by a majority voting scheme. By leveraging the collective wisdom of multiple trees, random forests enhance predictive accuracy and robustness, mitigating overfitting and capturing complex patterns within the data.

VII. EXPERIMENTAL WORK

The experimental work focuses on optimizing decision tree and random forest models for early Parkinson's disease diagnosis. It involves extensive hyperparameter tuning, feature engineering, and ensemble techniques. Through rigorous evaluation using cross-validation and relevant performance metrics, the goal is to achieve the highest accuracy possible. This iterative process aims to enhance the models' predictive capabilities, ultimately improving diagnostic accuracy and patient care.

VIII. SYSTEM STUDY

The system study involves a thorough examination of components and interactions in the early diagnosis of Parkinson's disease using machine learning. It aims to ensure reliability and effectiveness through testing, validation, and optimization of the system's functionality and performance.

IX. SYSTEM TESTING

System testing for the project focusing on the early diagnosis of Parkinson's disease using machine learning involves comprehensive steps. These include unit testing for individual components, integration testing for seamless interaction, data quality assessment, model performance evaluation using various metrics, robustness testing against noise and errors, scalability testing for large datasets, cross-validation for stability, user acceptance testing for usability, security testing for data protection, and documentation review for completeness and accuracy. These measures ensure the system's reliability, accuracy, and readiness for real-world deployment.

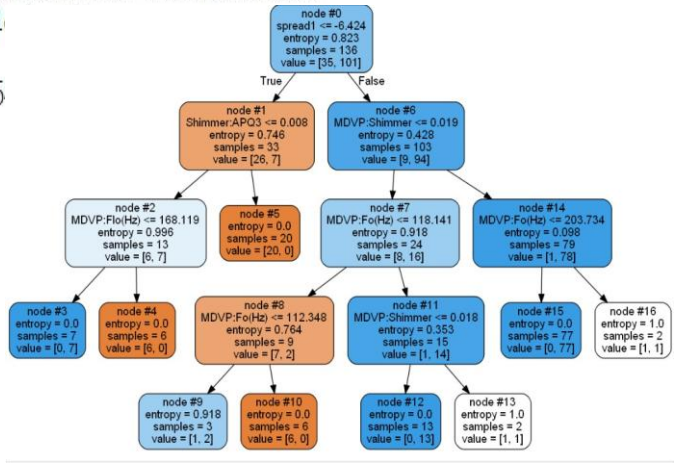
X. OUTPUT

1.Using the regularization parameters of max_depth, min_sample_leaf to recreate the model and the model accuracy

Model Accuracy for Regularized decision tree: 0.864406779661017

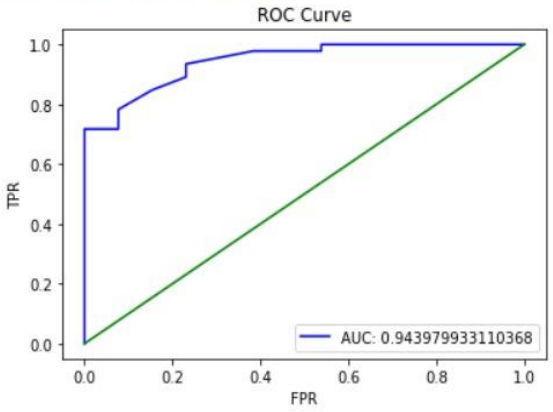
(max_depth = 4 and min_samples_l

Model accuracy has increased to 0.86



2.Implementing a Random Forest model and the optimal number of trees that gives the best result.

AUC Score: 0.943979933110368



Accuracy of Random Forest Model with optimal number of trees (51): 0.864406779661017

Confusion Matrix:

```
 [[10  3]
 [ 5 41]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.67	0.77	0.71	13
1	0.93	0.89	0.91	46
accuracy			0.86	59
macro avg	0.80	0.83	0.81	59
weighted avg	0.87	0.86	0.87	59

XI. CONCLUSION

In conclusion, employing machine learning algorithms on voice recordings offers a promising avenue for enhancing the screening and early diagnosis of Parkinson's Disease (PD) by analyzing characteristic vocal features like dysarthria, hypophonia, and monotone. This non-invasive method could serve as an initial screening step, enabling earlier detection and intervention, particularly in the disease's early stages when motor symptoms may not be as evident. However, ensuring accuracy and reliability across diverse populations through rigorous validation and testing is essential, along with addressing challenges such as variability in speech patterns and the need for representative datasets. The development of an automated screening process based on voice analysis could alleviate the burden on patients and healthcare systems, enabling remote monitoring and timely adjustments to treatment plans, ultimately improving patient outcomes and quality of life. Despite challenges, the application of machine learning to voice recordings represents a promising advancement in the early detection and management of Parkinson's Disease.

XII. REFERENCES :

[1] Rayan et al.Machine learning approaches in smart health
Procedia Comput Sci (2019)

[2] K.R. Wan et al.
A review on microelectrode recording selection of features for machine learning in deep brain stimulation surgery for Parkinson's disease Clin Neurophysiol (2019)

[3] T.Í. Pedrosa et al. Machine learning application to quantify the tremor level for Parkinson's disease patients
Procedia Comput Sci (2018)

[4] R. Prashanth et al. Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning Neurocomputing (2018)

[5] F. Cavallo et al. Upper limb motor pre-clinical assessment in Parkinson's disease using machine learning
Park Relat Disord (2019)

[6] J.S. Almeida Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques Pattern Recogn Lett (2019)

[7] L. Parisi et al. Feature-driven machine learning to improve early diagnosis of Parkinson's disease Expert Syst Appl (2018)

[8] M. Nilashi et al. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques Biocybern Biomed Eng
(2018)

[9] R. Das A comparison of multiple classification methods for diagnosis of Parkinson disease Expert Syst Appl (2010)
[10] Y. Wang et al. An adaptive kernel-based weighted extreme learning machine approach for effective detection of Parkinson's disease Biomed Signal Process Control (2017)

[11] G. Singh et al. Machine learning-based framework for multi-class diagnosis of neurodegenerative diseases: a study on Parkinson's Disease IFAC-Papers Online (2016)

[12] E. Abdulhay et al. Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease Fut Gener Comput Syst (2018)

[13]. Stratification of Parkinson Disease using python scikit- learn ML library, Ashish kolte, Bodireddy Mahita, Dr. NV Ganapathi Raju. Vellore Institute of Technology Tamil Nadu, India (2020)

[14]. Features Selection for Building an Early Diagnosis Machine Learning Model for Parkinson's Disease, Abu Bakr Soliman, Mohamed Fares, Mohamed M. Elhefnawi, Mahmoud Al- Hefnawy National Research Centre Giza, Egypt. (2018)

[15] Comparison of Classification Methods to Detect Parkinson disease,A.BOUROUHOU,A.JILBAB,C.NACIR,A.HAMMOUCH.MohammedV University Rabat, MOROCCO.(2017)

[16] Machine Learning Based Approaches for Parkinson Disease Prediction, Bind S., Tiwari A. K., Sahani A. K., Koulibaly P., Nobili F., Pagani M., et al. (2015)

[17] Early diagnosis of Parkinson's disease using machine learning algorithms, Almeida, T.A., & Hidalgo, J.M.G. (2011). Using Machine Learning to Combat Spam: An Evaluation. IEEE Intelligent Systems, 26(4), 80-84

[18] Machine Learning for the Diagnosis of Parkinson's Disease, Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

[19] The diagnosis of Parkinson’s disease, Tolosa E., Wenning G., Poewe W. (2006)

[20] A comparison of multiple classification methods for diagnosis of Parkinson disease, Das R. (2010)

[21] Diagnosing Parkinson by using artificial neural networks and support vector machines, Gil D., Manuel D. J. (2009)

[22] Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study, Harel B., Cannizzaro M., Snyder P. J. (2004)