International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 24 Issue 05, MAY, 2024

Prediction of Parkinson's disease Using Machine Learning Techniques

P.Punnarao [1], Mr.P.Ramesh Babu[2], N.Sarada Kumari [3]
Student[1, 3], Associate Professor[2]
Amritha Sai Institute of Science and Technology Paritala-521180
Autonomous NAAC with A Grade, Andhra Pradesh, India

ABSTRACT

Various data repositories contain extensive medical datasets that are utilized for disease identification. Parkinson's disease, a progressive and deadly nervous system disorder that affects movement, is the second most common neurological disorder ausing disability, reducing lifespan, and currently has no cure. Approximately 90% of individuals with Parkinson's suffer from speech disorders. In real-world applications, data is generated and analyzed using various machine learning techniques, which help derive meaningful insights. These algorithms are crucial for detecting diseases in their early stages, thereby potentially extending the lifespan of elderly individuals. Speech features play a significant role in Parkinson's disease diagnosis.

In this paper, the author employs several machine learning techniques, such as K-Nearest Neighbors (KNN), Naïve Bayes, and Logistic Regression, to predict Parkinson's disease based on user input and a relevant dataset. The study aims to determine which algorithm provides the highest accuracy. The results show that KNN achieves an accuracy of 80%, Logistic Regression 79%, and Naïve Bayes the highest at 81%, making it the preferred method for predicting the presence of Parkinson's disease in a frontend application. Early-stage prediction is essential for patient recovery, and this process can be effectively carried out using machine learning techniques.

Keywords: Parkinson's, Machine Learning, Speech disorders, KNN, Naïve Bayes, Logistic regression.

I. INTRODUCTION

A recent report from the World Health Organization indicates a significant increase in the number and health burden of Parkinson's disease patients, with the situation worsening rapidly. In China, the disease is spreading so quickly that it is estimated to affect half the population within the next decade. Classification algorithms are commonly used in the medical field to categorize data based on various characteristics. Parkinson's

disease, the second most dangerous neurological disorder, leads to symptoms such as shaking, stiffness, and difficulty with walking and balance, primarily due to the breakdown of nerve cells. This disease manifests both motor symptoms (e.g., slowness of movement, rigidity, balance issues, tremors) and non-motor symptoms (e.g., anxiety, breathing problems, depression, loss of smell, speech changes). Early detection of these symptoms is crucial for effective management.

In this paper, the author focuses on the speech features of patients to predict Parkinson's disease using machine learning techniques. Neurodegenerative disorders, including Parkinson's, result from the progressive loss of neurons, which are the functional units of the brain. Healthy neurons have extensions called dendrites and axons, and they contain a cell body and nucleus with DNA. When neurons deteriorate, they lose their extensions and accumulate waste, eventually losing their functionality.

This study aims to predict Parkinson's disease, a highly prevalent and incurable disorder named after James Parkinson, who initially described it as paralysis agitans. Parkinson's primarily affects neurons responsible for body movements, with key chemicals like dopamine and acetylcholine playing significant roles. Various environmental and genetic factors contribute to the disease's onset and progression. Environmental factors include exposure to heavy metals, poor air and water quality, unhealthy lifestyles, psychological stress, brain injuries, aging, and genetic predispositions. Speech articulation issues are also prevalent in Parkinson's patients, with symptoms such as a breathy and softer voice, smeared speech, and slower speech due to difficulty finding words.

Parkinson's disease symptoms are categorized into motor and non-motor types. Motor symptoms involve movement-related disorders like tremors, rigidity, freezing, and bradykinesia. Non-motor symptoms encompass mood and cognitive dysfunctions, complex behavioral disorders, and apathy. Additionally, primary symptoms (rigidity, tremor, slowness of movement) and secondary symptoms (impacting quality of life and varying widely among individuals) are identified. Other related symptoms include micrographia, decreased olfaction, postural instability, digestive issues, constipation, fatigue, weakness, and hypotension. Speech difficulties, such as dysphonia (impaired speech production) and dysarthria (speech articulation issues), are also common.

II. LITERARURE SURVY

The increasing prevalence of Parkinson's disease (PD) and its substantial impact on health burden have driven significant research into predictive modeling using machine learning techniques. Ozcift (2012) explored the use of Support Vector Machine (SVM) feature selection combined with rotation forest ensemble classifiers to enhance the accuracy of computer-aided diagnosis of Parkinson's disease, highlighting the potential of ensemble methods in improving diagnostic performance by selecting the most relevant features from medical datasets [1]. Several studies have utilized artificial neural networks (ANN) for PD diagnosis. Anila and Pradeepini (2020) employed ANN to diagnose Parkinson's, demonstrating its efficacy in handling complex patterns in medical data [2], while SoleimanianGharehehopogh and Mohammadi (2013) conducted a case study using ANN, underscoring its accuracy in diagnosing PD [8]. Tiwari (2016) provided a comprehensive review of various machine learning approaches for predicting PD, emphasizing the importance of algorithm selection in improving prediction accuracy [3]. Similarly, Miljkovic et al. (2016) discussed the application of machine learning and data mining methods in managing Parkinson's disease, highlighting the role of these techniques in early detection and patient management [7]. Ricciardi et al. (2019) utilized gait analysis parameters in a data mining framework to classify Parkinsonism, illustrating how biomechanical data can distinguish PD from other neurological disorders using machine learning techniques [4].

Bhatia and Sulekh (2017) developed a predictive model for Parkinson's disease using Naive Bayes classification, demonstrating its applicability in predicting the disease based on patient records [5]. GeethaRamani et al. (2012) performed feature relevance analysis on telemonitoring data, using data mining techniques to classify Parkinson's disease, emphasizing the importance of relevant feature selection in predictive modeling [6]. Abdar and Zomorodi-Moghadam (2018) studied the impact of patients' gender on Parkinson's disease using classification algorithms, finding that demographic factors significantly influence disease prediction and should be considered in model development [10]. Wroge et al. (2018) and Swapna and Devi (2019) focused on using voice attributes and machine learning algorithms to diagnose Parkinson's disease, underscoring the importance of speech features in PD diagnosis and the potential of deep learning techniques in extracting

relevant patterns from vocal data [24][23]. Alissa (2018) further explored deep learning methods, highlighting their superior performance in complex pattern recognition tasks associated with Parkinson's disease [14]. Bind et al. (2015) provided a survey of machine learning approaches for Parkinson's disease prediction, offering insights into various algorithms and their effectiveness [20], while Reddy and Ramanadham (2020) discussed the use of big data analytics in early-stage prediction of Parkinson's disease, emphasizing the role of large-scale data processing in enhancing prediction accuracy [21]. Research by Van Stiphout et al. (2018) and Duncan et al. (2015) highlighted the influence of genetic and environmental factors on Parkinson's disease progression, advocating for the integration of these variables into predictive models to improve accuracy [11][16]. Lastly, Sriram et al. (2013) and Mandal and Sairam (2014) investigated the use of intelligent systems and novel machine learning algorithms for Parkinson's disease prediction, demonstrating the potential of advanced computational techniques in medical diagnostics [22][13].

III. PROBLEM STATEMENT

EXISTING SYSTEM:

The current methods for diagnosing and predicting Parkinson's disease (PD) utilize various machine learning techniques, each with distinct advantages. Support Vector Machines (SVM) with feature selection enhance diagnostic accuracy by identifying key features from medical datasets. Artificial Neural Networks (ANN) are effective in handling complex patterns, as demonstrated in multiple studies showcasing their accuracy in PD diagnosis. Ensemble methods, like the rotation forest ensemble classifier, combine multiple learning algorithms to improve performance. Gait analysis parameters, integrated with data mining, effectively classify Parkinsonism by distinguishing PD from other neurological disorders.

Predictive models using Naive Bayes classification have shown promise in predicting PD from patient records. Speech features are also crucial, with machine learning algorithms leveraging voice attributes to detect PD. Deep learning techniques excel in recognizing complex patterns associated with the disease. Additionally, demographic factors, such as gender, significantly influence disease prediction models, while research on genetic and environmental factors aims to enhance predictive accuracy.

These methods demonstrate promising results but highlight the need for further integration and refinement of machine learning approaches to improve early detection and management of Parkinson's disease across diverse populations.

PROPOSED SYSTEM:

The proposed system aims to advance Parkinson's disease (PD) diagnosis and prediction through an integrated approach leveraging machine learning techniques. Building upon the strengths of existing methods, the proposed system will focus on optimizing feature selection and algorithm selection to enhance diagnostic accuracy and predictive capabilities. Feature selection methods, such as SVM-based feature selection and ensemble techniques, will be employed to identify the most relevant features from medical datasets, ensuring that the predictive models capture essential information for accurate diagnosis.

Artificial Neural Networks (ANN) will be further explored and optimized for handling complex patterns in medical data, with a focus on improving scalability and efficiency. Deep learning techniques will be integrated to extract intricate patterns associated with PD from diverse datasets, enhancing the system's ability to detect subtle disease markers. Ensemble methods will be employed to combine multiple learning algorithms and improve overall predictive performance.

Moreover, the proposed system will incorporate advancements in speech analysis, leveraging voice attributes and speech patterns to enhance PD detection. Integration of demographic factors, genetic predispositions, and environmental influences will be prioritized to develop more comprehensive predictive models. The system will be designed to be adaptable to diverse patient populations, ensuring its applicability across different demographics and geographic regions.

Overall, the proposed system aims to advance the state-of-the-art in PD diagnosis and prediction by integrating cutting-edge machine learning techniques, optimizing feature selection, and considering a holistic approach that accounts for various factors influencing the disease. Through these efforts, the proposed system seeks to improve early detection, facilitate personalized treatment strategies, and ultimately enhance the quality of life for individuals living with Parkinson's disease

Methodology:

The utilization of machine learning algorithms, including Logistic Regression, KNN, and Naïve Bayes, empowers

computer systems to autonomously learn from data without explicit programming. The architectural framework outlines a systematic process to refine raw data for predicting Parkinson's disease, encompassing five key steps. Firstly, the architecture diagram delineates the flow of processing, elucidating the refinement of raw data for Parkinson's prediction. Subsequently, the collected data undergoes preprocessing to render it into an interpretable format. The dataset is then partitioned into training and testing subsets to facilitate algorithm training.

Following data preparation, the Parkinson's data undergoes evaluation utilizing machine learning algorithms, specifically Logistic Regression, KNN, and Naïve Bayes. The classification accuracy of each model is ascertained

through this evaluation phase. Upon completion of training with the respective algorithms, testing is conducted to validate the model's predictive efficacy. Finally, the results obtained from each algorithm are juxtaposed based on their classification accuracy, enabling comparative analysis and informed decision-making regarding algorithm selection for Parkinson's disease prediction.

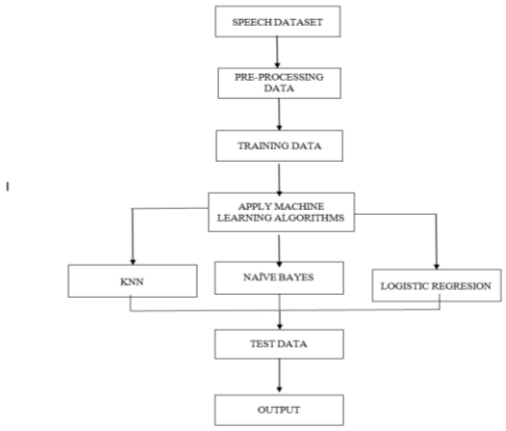

Fig-1. Architecture

IV. RESULTS & DISCUSSION

The KNN algorithm undergoes a systematic training process with the designated training dataset, followed by subsequent testing with the remaining test data. As depicted in below a screenshot from our notebook illustrates the step-by-step execution of the KNN algorithm, showcasing the process flow and the resultant accuracy achieved by the model, which stands at 80%. This accuracy metric serves as a quantitative assessment of the model's performance in predicting Parkinson's disease.

Fig-2. KNN Test Accuracy

Naïve Bayes algorithm is trained with the training dataset and later it was tested with the remaining test data. In Fig 3, a screenshot of our notebook is showing that how the process of Naïve Bayes

algorithms is done and the accuracy the model returns and it isof more with 81%.

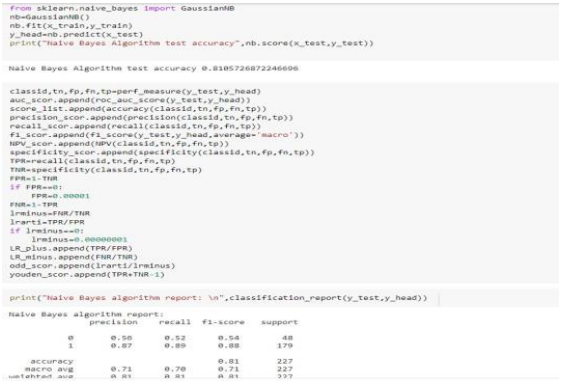

Fig-3. Navie Bayes Test Accuracy

Logistic Regression algorithm is trained with the training dataset and later it was tested with the remaining test data. In Fig 4. a screenshot of our notebook is showing that how the process of Logistic Regression algorithms is done and the accuracy the model returns and it is of more with 79%.

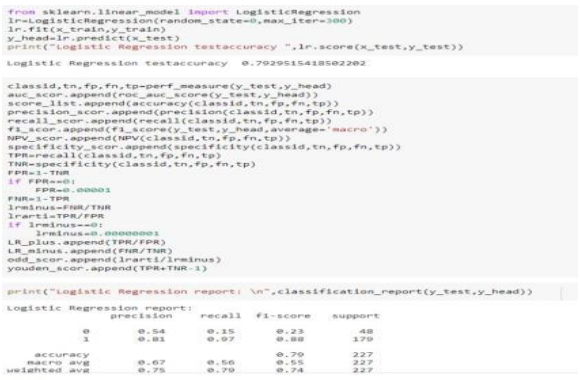

Fig-4. Logistic Regression Test Accuracy

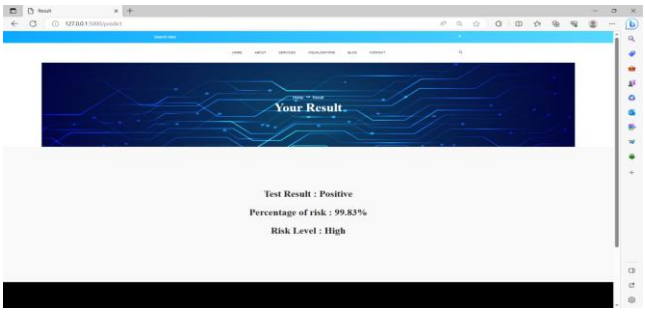

Fig-5. Predicted Result

VI. CONCLUSION

Parkinson's disease, ranking as the second most perilous neurodegenerative ailment without a known cure, necessitates accurate prediction for effective mitigation. This project employs three distinct prediction models – KNN, Naïve Bayes, and Logistic Regression – all falling under the purview of Machine Learning Techniques. Utilizing a dataset sourced from Kaggle containing voice features of over 750 patients across 700+ attributes, the models are trained and compared, aiming to identify the most suitable predictor. Feature selection techniques refine the dataset to five optimal features, enhancing model performance. Evaluation metrics including Accuracy, Precision, Recall, Specificity, F1-score, LR+, LR-, and Youden score are employed to assess prediction efficiency.

Among the models, Naïve Bayes emerges as the standout performer, boasting an accuracy of 81%. This system demonstrates the capability to effectively predict Parkinson's disease, offering valuable insights into disease prognosis and facilitating timely intervention.

VII. REFERENCE

[1] A. Ozcift, "SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of

Parkinson disease" Journal of medical systems, vol-36, no. 4, pp. 2141-2147, 2012.

[2] Anila M Department of CS1, Dr G Pradeepini Department of CSE, "DIAGNOSIS OF PARKINSON'S DISEASE USING ARTIFICIAL NEURAL NETWORK", JCR, 7(19): 7260-7269, 2020.

[3] Arvind Kumar Tiwari, "Machine Learning based Approaches for Prediction of Parkinson's Disease" Machine Learning and Applications: An International Journal (MLAU) vol. 3, June 2016.

[4] Carlo Ricciardi, et al, "Using gait analysis' parameters to classify Parkinsonism: A data mining approach" Computer Methods and Programs in Biomedicine vol. 180, Oct. 2019.

[5] Dr. Anupam Bhatia and Raunak Sulekh, "Predictive Model for Parkinson's Disease through Naive Bayes Classification" International Journal of Computer Science & Communication vol-9, Dec. 2017, pp. 194-202, Sept 2017 - March 2018.

[6] Dr. R.GeethaRamani, G.Sivagami, ShomonaGraciajacob "Feature Relevance Analysis and Classification of Parkinson's Disease TeleMonitoring data Through Data Mining" International Journal of Advanced Research in Computer Science and Software Engineering, vol-2, Issue 3, March 2012.

[7] Dragana Miljkovic et al, "Machine Learning and Data Mining Methods for Managing Parkinson's Disease" LNAI 9605, pp. 209-220, 2016.

[8] FarhadSoleimanianGharehehopogh, PeymenMohammadi, "A Case Study of Parkinson's Disease Diagnosis Using Artificial Neural Networks" International Journal of Computer Applications, Vol-73, No.19, July 2013.

[9] Heisters. D, "Parkinson's: symptoms, treatments and research". British Journal of Nursing, 20(9), 548–554. doi:10.12968/bjon.2011.20.9.548, 2011.

[10] M. Abdar and M. Zomorodi-Moghadam, "Impact of Patients' Gender on Parkinson's disease using Classification Algorithms" Journal of AI and Data Mining, vol-6, 2018.

[11] M. A. E. Van Stiphout, J. Marinus, J. J. Van Hilten, F. Lobbezoo, and C. De Baat, "Oral health of Parkinson's disease patients: a case-control study" Parkinson's disease, vol-2018, Article ID 9315285, 8 pages, 2018.

[12] Md. Redone Hassan, et al, "A Knowledge Base Data Mining based on Parkinson's Disease" International Conference on System Modelling & Advancement in Research Trends, 2019.

[13] Mandal, Indrajit, and N. Sairam. "New machine-learning algorithms for prediction of Parkinson's disease" International Journal of Systems Science 45.3: 647-666, 2014.

[14] Mohamad Alissa," Parkinson's Disease Diagnosis Using Deep Learning", August 2018.

[15] PeymanMohammadi, AbdolrezaHatamlou and Mohammed Msdaris "A Comparative Study on Remote Tracking of Parkinson's Disease Progression Using Data Mining Methods" International Journal in Foundations of Computer Science and Technology(IJFCST), vol-3, No.6, Nov 2013.

[16] R. P. Duncan, A. L. Leddy, J. T. Cavanaugh et al., "Detecting and predicting balance decline in Parkinson disease: a prospective cohort study" Journal of Parkinson's Disease, vol-5, no. 1, pp. 131–139, 2015.

[17] Ramzi M. Sadek et al., "Parkinson's Disease Prediction using Artificial Neural Network" International Journal of Academic Health and Medical Research, vol-3, Issue 1, January 2019.

[18] Satish Srinivasan, Michael Martin & Abhishek Tripathi, "ANN based Data Mining Analysis of Parkinson's Disease" International Journal of Computer Applications, vol-168, June 2017.

[19] Shahid, A.H., Singh, M.P. A deep learning approach for prediction of Parkinson's disease progression, https://doi.org/10.1007/s13534-020-00156-7, Biomed. Eng. Lett. 10, 227–239, 2020.

[20] Shubham Bind, et al, "A survey of machine learning based approaches for Parkinson disease prediction" International Journal of Computer Science and Information

Technologies vol-6, Issue 2, pp. 1648- 1655, 2015.

[21] Siva Sankara Reddy Donthi Reddy and Udaya Kumar Ramanadham "Prediction of Parkinson's Disease at Early Stage using Big Data Analytics"ISSN: 2249 – 8958, Volume- 9 Issue-4, April 2020

[22] Sriram, T. V., et al. "Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms" International Journal of Engineering and Innovative Technology, vol-3, Issue 3, September 2013.

[23] T. Swapna, Y. Sravani Devi, "Performance Analysis of Classification algorithms on Parkinson's Dataset with Voice Attributes". International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 2 pp. 452-458, 2019.

[24] 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), pp.1-7, doi: 10.1109/SPMB.2018.8615607, 2018.