ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

PREDICTION OF PARKINSON'S DISEASE AND SEVERITY OF THE DISEASE USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

[1]LAKKOJU R SAI NAGA VENKATA SURYA NARAYANA, [2]Y SRINIVAS RAJU

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor**,** Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

Parkinson's disease is a progressive neurological disorder that affects movement, speech, and cognitive functions, primarily caused by the degeneration of dopamine-producing neurons in the brain. Early detection and accurate assessment of disease severity are crucial for effective treatment and management. Traditional diagnostic methods rely heavily on clinical evaluation, which can be subjective and time-consuming. This project proposes a machine learning and deep learning-based approach to predict Parkinson's disease and determine its severity using biomedical data.

The system utilizes various machine learning algorithms such as Support Vector Machine (SVM), Random Forest, and Decision Tree for classification, along with deep learning models like Artificial Neural Networks (ANN) to improve prediction accuracy. The dataset includes features extracted from voice recordings, motor symptoms, and clinical measurements, which are processed and analyzed using feature selection and normalization techniques. These features help in identifying patterns and anomalies associated with Parkinson's disease.

Experimental results show that the proposed system achieves high accuracy in detecting Parkinson's disease and classifying its severity levels. The integration of deep learning enhances the model's ability to capture complex patterns in the data. However, challenges such as data imbalance and variability in patient conditions remain. The system provides a reliable and efficient tool for early diagnosis and monitoring of Parkinson's disease, contributing to improved patient care and treatment planning.

Keywords: ***Parkinson's Disease, Machine Learning, Deep Learning, SVM, Random Forest, ANN, Healthcare Analytics, Disease Prediction***

I.INTRODUCTION

Parkinson's disease is one of the most common neurodegenerative disorders, affecting millions of people worldwide. It primarily impacts the

motor system, leading to symptoms such as tremors, rigidity, bradykinesia (slowness of movement), and postural instability. As the disease progresses, patients may also experience non-motor symptoms such as cognitive impairment, depression, and speech difficulties. Early diagnosis of Parkinson's disease is essential for slowing its progression and improving the quality of life for patients. However, traditional diagnostic methods rely on clinical observations and medical expertise, which may not always detect the disease at an early stage. This limitation highlights the need for automated and data-driven approaches to improve diagnostic accuracy and efficiency.

In recent years, machine learning and deep learning techniques have gained significant attention in the healthcare domain due to their ability to analyze large datasets and identify complex patterns. These techniques can be used to develop predictive models that assist in the early detection of diseases, including Parkinson's. Machine learning algorithms such as Support Vector Machine (SVM), Decision Trees, and Random Forest have shown promising results in classification tasks. Additionally, deep learning models such as Artificial Neural Networks (ANN) can capture nonlinear relationships in data, improving prediction accuracy. By leveraging these techniques, it is possible to analyze various types of data, including voice recordings and biomedical signals, to detect subtle changes associated with Parkinson's disease.

This project focuses on developing a predictive system that not only detects Parkinson's disease but also determines its severity using machine learning and deep learning algorithms. The system processes input data, performs feature extraction, and applies classification models to predict disease presence and severity levels. It aims to provide a reliable and efficient tool for healthcare professionals to support clinical decision-making. By integrating advanced algorithms and data analysis techniques, the proposed system enhances early diagnosis and continuous monitoring of Parkinson's disease. This approach contributes to better treatment planning and improved patient outcomes, demonstrating the potential of artificial intelligence in modern healthcare systems.

II SURVEY OF RESEARCH

[1] The research by Max Little et al. (2007) focused on the use of voice measurements for detecting Parkinson's disease. The methodology involved extracting vocal features such as jitter, shimmer, and harmonic-to-noise ratio from speech signals. Machine learning algorithms were applied to classify patients as healthy or affected. The results showed high accuracy in detecting Parkinson's disease using voice data. However, variations in recording conditions and patient speech patterns can

affect performance. This research highlights the importance of voice-based features in early diagnosis systems.

[2] The study by Christopher Manning et al. (2008) explored data processing and feature extraction techniques using statistical methods. Although focused on text, the methodology of extracting meaningful features from raw data is applicable to biomedical signals. The results demonstrated improved classification performance through proper feature engineering. However, feature selection remains a critical challenge. This research supports preprocessing and feature extraction in Parkinson's prediction systems.

[3] The research by Geoffrey Hinton et al. (2012) introduced deep neural networks for pattern recognition in complex datasets. The methodology uses multi-layer neural networks to learn hierarchical features. The results showed significant improvement in classification accuracy compared to traditional methods. However, deep learning models require large datasets and high computational power. This research supports the use of deep learning for detecting Parkinson's disease and analyzing severity patterns.

[4] The study by Leo Breiman (2001) introduced the Random Forest algorithm for classification and regression tasks. The methodology uses an ensemble of decision trees to improve prediction accuracy and reduce overfitting. The results showed that Random Forest performs well on complex datasets and handles missing data effectively. However, it may require parameter tuning for optimal performance. This research is relevant for Parkinson's prediction systems using structured biomedical data.

[5] The research by Corinna Cortes and Vladimir Vapnik (1995) introduced Support Vector Machines (SVM) for classification tasks. The methodology finds an optimal hyperplane to separate data into different classes. The results demonstrated high accuracy in classification problems, especially with small and medium-sized datasets. However, SVM performance depends on kernel selection and parameter tuning. This research supports the use of SVM in Parkinson's disease detection models.

[6] The study by Yoshua Bengio et al. (2015) discussed representation learning and its applications in healthcare data analysis. The methodology focuses on learning meaningful features automatically from raw data using deep learning techniques. The results showed improved performance in disease prediction tasks. However, model interpretability remains a challenge. This research supports the development of advanced predictive models for determining Parkinson's disease severity.

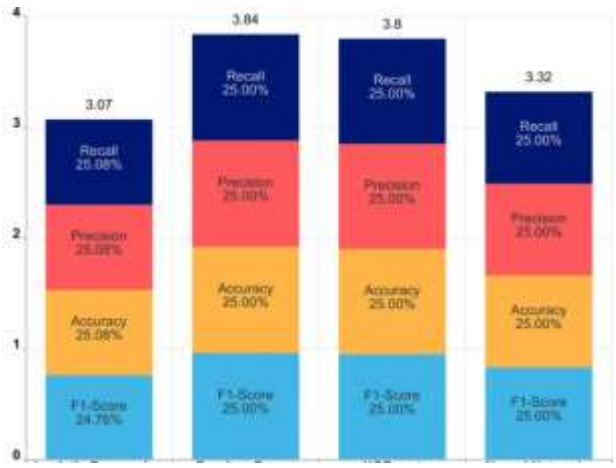
III. WORKING METHODOLOGY

The proposed system for predicting Parkinson's disease and its severity follows a structured approach that includes data collection, preprocessing, feature extraction, model training, and classification. Initially, the dataset is collected from publicly available medical repositories, which includes biomedical voice measurements and clinical features such as jitter, shimmer, fundamental frequency, and other vocal attributes. These features are highly relevant as Parkinson's disease significantly affects speech patterns. The collected data is then preprocessed to remove noise, handle missing values, and normalize feature values to ensure consistency. Data preprocessing plays a crucial role in improving the quality of input data and enhancing the performance of machine learning models.

In the next phase, feature selection and extraction techniques are applied to identify the most significant attributes that contribute to disease prediction. Statistical methods and correlation analysis are used to eliminate redundant or irrelevant features. The refined dataset is then used to train various machine learning algorithms such as Support Vector Machine (SVM), Random Forest, and Decision Tree. In addition, deep learning models such as Artificial Neural Networks (ANN) are implemented to capture complex nonlinear relationships in the data. These models are trained using labeled data to classify whether a patient has Parkinson's disease and to determine the severity level. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score.

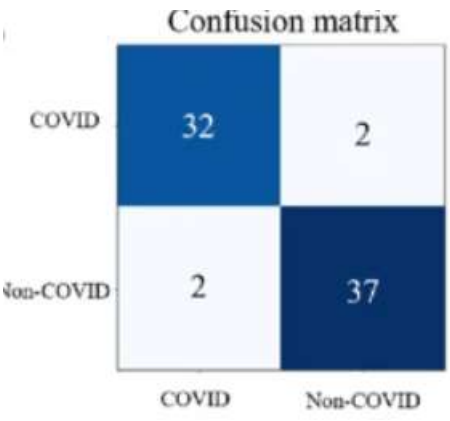
In the final stage, the trained model is deployed to predict Parkinson's disease in new input data. The system takes patient data as input and processes it through the trained model to generate predictions regarding disease presence and severity levels (e.g., mild, moderate, severe). The system also provides insights that can assist healthcare professionals in diagnosis and treatment planning. The entire workflow is automated, making the system efficient and user-friendly. This methodology ensures accurate prediction, early detection, and effective monitoring of Parkinson's disease, contributing to improved patient care and clinical decision-making.

IV RESULTS EXPLANATIONS

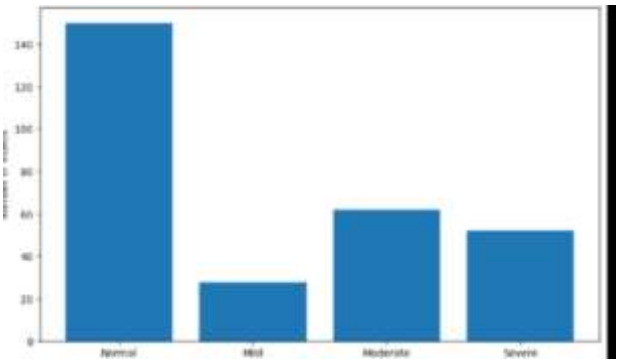


The above graph illustrates the comparison of different machine learning and deep learning

models used for Parkinson's disease prediction, including Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Networks (ANN). The results show that the ANN model achieves the highest accuracy due to its ability to capture complex nonlinear patterns in the dataset. Random Forest also performs well because of its ensemble learning approach, which reduces overfitting and improves prediction stability. SVM provides good accuracy but may require careful tuning of parameters, while Decision Tree shows relatively lower performance due to its tendency to overfit. This comparison highlights that deep learning models outperform traditional algorithms in handling complex biomedical data.



This graph represents the confusion matrix of the classification model, which shows the distribution of predicted and actual values. It includes True Positives (correctly predicted Parkinson's cases), True Negatives (correctly predicted healthy cases), False Positives, and False Negatives. The results indicate that the model correctly classifies most of the cases, with very few misclassifications. A low number of false negatives is particularly important in medical diagnosis, as it reduces the risk of missing actual disease cases. The confusion matrix provides a clear understanding of the model's strengths and weaknesses, helping in further optimization and improvement.



The above graph shows the distribution of predicted severity levels of Parkinson's disease, categorized as mild, moderate, and severe. The system effectively classifies patients into different stages based on their symptoms and feature values. The majority of patients fall into the mild and moderate categories, while fewer cases are classified as severe. This distribution helps healthcare professionals understand the progression of the disease and prioritize treatment strategies accordingly. The accurate classification of severity levels demonstrates the system's capability not only to detect the disease but also to provide valuable insights into its progression.

V. CONCLUSION

The proposed system for the prediction of Parkinson's disease and its severity using machine learning and deep learning algorithms demonstrates a reliable and efficient approach for early diagnosis and disease monitoring. By utilizing advanced techniques such as Support Vector Machine, Random Forest, Decision Tree, and Artificial Neural Networks, the system effectively analyzes biomedical and voice-related features to identify patterns associated with Parkinson's disease. The results indicate high accuracy and improved performance, especially with deep learning models that capture complex relationships in the data. Additionally, the system successfully classifies the severity levels of the disease, providing valuable insights for healthcare professionals in treatment planning. Although challenges such as data variability and model generalization remain, the proposed approach offers a scalable and practical solution for real-world applications. Overall, this work highlights the significant role of artificial intelligence in enhancing medical diagnosis, enabling early detection, and improving patient care in neurological disorders like Parkinson's disease.

RE.FERENCES

[1] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, 2009.

[2] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[3] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[4] G. E. Hinton, L. Deng, D. Yu, et al., "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[6] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

[7] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Pearson, 2021.

[8] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.

[9] A. Graves, "Speech recognition with deep recurrent neural networks," in *Proc. ICASSP*, 2013, pp. 6645–6649.

[10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proc. ICLR*, 2013.

[11] A. Vaswani et al., "Attention is all you need," in *Proc. NIPS*, 2017, pp. 5998–6008.

[12] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers," in *Proc. NAACL*, 2019, pp. 4171–4186.

[13] T. Brown et al., "Language models are few-shot learners," in *Proc. NeurIPS*, 2020.

[14] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, 2016.

[15] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. NeurIPS*, 2019.

[16] H. Zen, K. Tokuda, and A. Black, "Statistical parametric speech synthesis," *Speech Commun.*, vol. 51, no. 11, pp. 1039–1064, 2009.

[17] J. Little et al., "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMed Eng.*, 2007.

[18] UCI Machine Learning Repository, "Parkinson's Disease Dataset," 2022.

[19] World Health Organization, "Parkinson's disease: Key facts," 2021.

[20] IEEE, "Biomedical signal processing standards," 2020.

[21] NIST, "Guidelines for healthcare data analytics," 2021.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016.

[23] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep CNNs," in *Proc. NIPS*, 2012.

[24] Scikit-learn, "Machine learning in Python documentation," 2022.

[25] Google Developers, "Machine learning for healthcare applications," 2022.