ISSN 2320 - 2602

Volume 13 No. 7, July 2024

International Journal of Advances in Computer Science and Technology

Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst011372024.pdf

https://doi.org/10.30534/ijacst/2024/011372024

Parkinson’s Disease Prediction Using Machine Learning Models

Srinivas Kanakala [1]

[1]Department of CSE, VNR Vignana Jyothi Institute of Engineering and Technology, India,
srinivaskanakala@gmail.com

Received Date : May 25 , 2024 Accepted Date : June 27, 2024 Published Date : July 07, 2024

ABSTRACT

Parkinson's disease is a neurodegenerative condition that affects billions of persons worldwide. This abstract aims to shed light on the causes and consequences of this debilitating condition. The primary cause of Parkinson's disease is the progressive degeneration of dopaminergic neurons in the substantia nigra region of brain. This neuronal loss results in a depletion of dopamine, a crucial neurotransmitter responsible for regulating movement and coordination. Therefore, individuals with Parkinson's disease have symptoms like tremors, rigidity, bradykinesia, and postural instability. These signs profoundly impact the quality of life, causing difficulties with daily activities and reducing independence. In addition to motor symptoms, non-motor symptoms such as depression, cognitive impairment, and autonomic dysfunction often accompany the disease, further complicating the clinical picture. Research into the causes and consequences of Parkinson's disease is ongoing, with a focus on using efficient medications and refining the quality of life for those affected by this condition. Now by Using machine learning algorithms, we can predict whether a person has a specific disease based on input values like gender and age. These algorithms analyze patterns and relationships in data to get predictions about an individual's health status. This technology can assist in early disease detection and improve healthcare outcomes..

Key words: Parkinson’s Disease , Machine Learning**,** prediction, models.

1. INTRODUCTION

Parkinson's Disease is a disease that causes the muscles to This system uses fancy technology like computer learning and data study. It looks at lots of things like family history, certain signs in the body, and how people live. By looking at all these things together, it tries to tell if someone might get Parkinson's and how it might change over time. If it works well, this system could help doctors find Parkinson's sooner and create special plans to help people better. This document talks about how this system was made, what it might mean for doctors and people and how it could change how we deal with Parkinson's disease. become weak, the arms and legs shake and that gets worse over a period.

In recent years, the integration of machine learning techniques in healthcare has shown promising results, offering innovative solutions for early detection and prediction of various medical conditions, including Parkinson's disease. In the world of healthcare technology, a new system has been developed to predict Parkinson's disease before it shows up. Parkinson's is a tough disease that affects people's movements and other parts of their health. This document is all about this system, explaining how it works and how it could help find and treat Parkinson's disease early. To ensure accurate predictions, the collected data is meticulously preprocessed to address any inconsistencies, such as missing values or outliers. Key features relevant to the disease, including motor symptoms, cognitive functions, and genetic markers, are carefully selected from the dataset. Using machine learning algorithms, such as support vector machines, random forests, or deep learning models, predictive models are developed based on the preprocessed data. These models are trained on historical data to make accurate predictions regarding the likelihood of Parkinson's disease occurrence or progression in new cases. The performance of these models is rigorously evaluated using appropriate metrics to ensure their reliability and precision in predicting Parkinson's disease. Subsequently, the models are deployed in real- world scenarios and validated using new patient data. Continuous updates and refinements are made based on insights and information gathered from ongoing research and patient outcomes. Ultimately, the implementation of machine learning in Parkinson's disease prediction holds the potential to revolutionize early detection and intervention strategies. If successful, this research could lead to more personalized and timely interventions, ultimately improving the lives of those affected by Parkinson's disease and can potentially improve early detection, leading to improved patient outcomes and a better quality of life.

2. LITERATURE SURVEY

Landolfi [1] discussed various techniques used for the diagnosis and prognosis of Parkinson's disease. It covers different algorithms, feature extraction methods, and the effectiveness of these models in clinical settings. The survey emphasizes the importance of accurate data and advanced computational methods in improving prediction outcomes.

Althum [2] focused on the application of deep learning models in the detection and classification of Parkinson's disease. It discusses convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, highlighting their performance in analyzing medical imaging and time-series data from patients. [3] author discussed that CNN can be also used for detection of covid spread analysis.

Dixit [4] use of speech analysis as a non-invasive method for early detection of Parkinson's disease. It covers various acoustic features, signal processing techniques, and machine learning models that have been used to identify vocal biomarkers associated with the disease. [5] examined the integration of wearable devices and the Internet of Things (IoT) in monitoring and predicting Parkinson's disease progression. It discusses the types of sensors used, data collection methodologies, and the potential of continuous monitoring systems in providing real-time insights for clinicians. [6]. In [7] author discussed identification and utilization of genetic and genomic biomarkers for predicting and diagnosing Parkinson's disease. It discusses the latest findings in genome-wide association studies (GWAS), the role of specific genes, and the potential for personalized medicine approaches. In [8] comprehensive review covers the broad spectrum of artificial intelligence applications in Parkinson's disease. It includes discussions on diagnosis, symptom monitoring, treatment optimization, and patient management. The review highlights the strengths and limitations of current AI models and suggests future research directions.

3. SYSTEM ARCHITECTURE

The system architecture for the Parkinson's disease prediction system is shown in figure 1 which revolves around a cohesive framework that seamlessly integrates data processing, machine learning, and user interaction. Starting with data collection and integration, diverse datasets, including patient demographics and clinical assessments, undergo preprocessing to ensure uniformity. Feature extraction identifies crucial patterns, and the machine learning model, trained on historical data, forms the predictive core. The training pipeline maintains the model's accuracy by periodic updates. The prediction engine applies the model to new patient data, generating likelihood scores for Parkinson's disease. The user interface facilitates interaction for healthcare professionals to input data and view results, while consent management ensures ethical data usage. A feedback mechanism enhances system performance based on user input, and a security layer safeguards sensitive information. The database stores and manages patient data, while external integrations enhance data diversity. This architecture creates an effective, user-friendly, and secure Parkinson's disease prediction system.

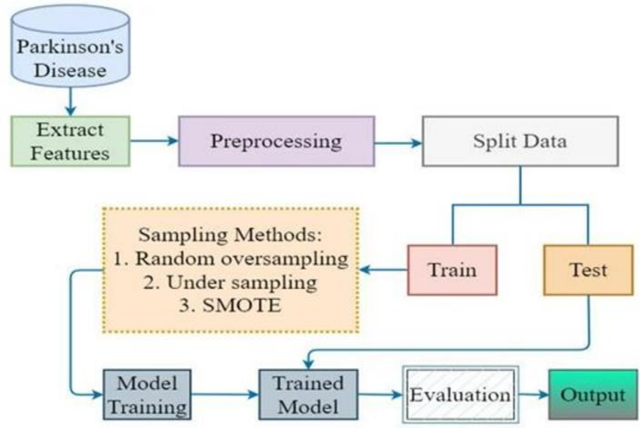

Figure 1: System Architecture.

4. PROPOSED SYSTEM

Gather relevant data, including clinical features, patient demographics, and possibly biomarkers associated with Parkinson's disease. Next Clean and preprocess the data, handling missing values, outliers, and ensuring compatibility with machine learning algorithms. Further identify important features using techniques like correlation analysis, recursive feature elimination, or domain knowledge. Create new features if necessary. Next Choose appropriate machine learning models such as Logistic Regression, XGBoost, or Support Vector Machine (SVM) based on their performance characteristics and suitability for the task. Train the selected models on the training dataset, optimizing hyperparameters to achieve the best performance. Evaluate model performance using metrics like ROC AUC, accuracy, precision, recall, and F1-score. Validate the models on a separate validation dataset. Enhance model interpretability using techniques like SHAP values or LIME to understand the contributions of features to predictions. Firstly, the modules required to run the application are imported. Using Kaggle API data is retrieved through the keys stored in CSV file. The contents of the dataset copy are extracted and stored in the form of a table using pandas. Cleaning the unwanted data from the dataset to analyze the required data easily, The column id is grouped here and the noise is reduced by decreasing the row count. By measuring the correlation factor we have reduced the number of columns (The columns which are having high correlation between them are removed). By using Minmax Scalar function we normalized our data in certain interval (i.e. in particular range) And by using chi-square test we will select the best thirty columns from the existing data (It selects the columns which have a significant association between them). For training We are splitting the data into training and validation set, inorder to compare the accuracy between the 2 sets as the data was highly imbalanced, we will balance it by adding repetitive rows of minority class. We fit different machine learning models and will check the accuracy for each model and select the best one. Logistic regression model is used to predict whether a person is affected with disease or not. As it has a minute difference between the accuracies of tracing set and validation set.

5. RESULTS

Logistic Regression is a statistical method used for binary classification, expecting the chances of an event occurence or not. It is used for classification tasks, not regression. The procedure models the relationship between independent variables and the probability of a particular outcome, applying the logistic function constrain predictions between 0 and 1. It's widely used in various fields for its simplicity, interpretability, and effectiveness in handling linearly separable data. ROC and AUC of logistic regression for Parkinson's disease prediction is shown in figure 2.

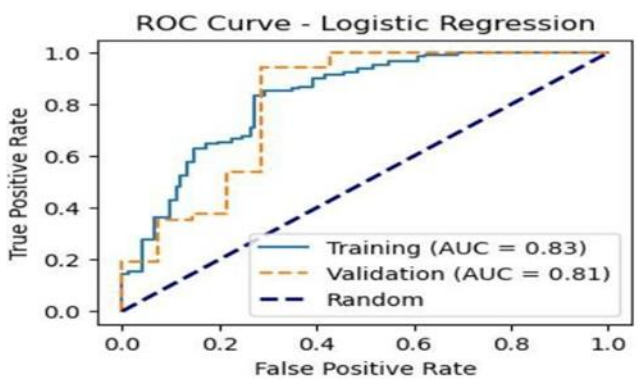

Figure 2: Logistic regression

XGBoost, short for Extreme Gradient Boosting, is a compelling machine learning algorithm renowned for its effectiveness in classification tasks. It comes under the family of boosting algorithm, combining the strength of multiple weak learners to build a robust predictive model. XGBoost optimizes both computational efficiency and model performance, incorporating regularization techniques to prevent overfitting. Its flexibility, speed, and capability to handle complex datasets make XGBoost a popular choice in diverse domains. ROC and AUC of XGBoost for Parkinson's disease prediction is shown in figure 3.

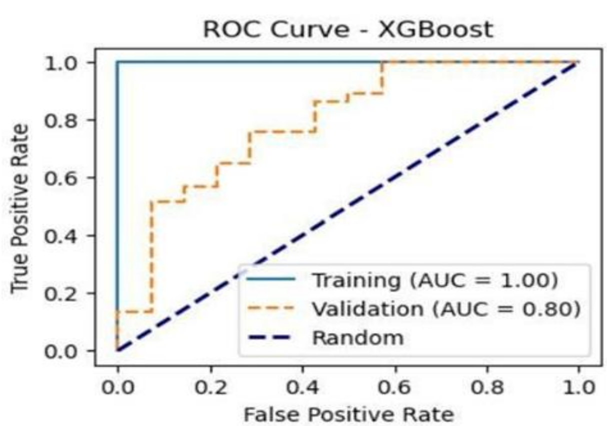

Figure 3: XGB Classification.

SVM, short for Support Vector Machine, is a powerful machine learning algorithm renowned for its effectiveness in classification tasks. It is a supervised machine learning algorithm, combining the strength of multiple weak learners to create a robust predictive model. SVM optimizes both computational efficiency and model performance, incorporating regularization techniques to prevent overfitting. Its flexibility, speed, and ability to handle complex datasets make SVM a popular choice in diverse domains such as finance, healthcare, and competitive machine learning. It classifies new data points by determining which side of the hyperplane they fall on. SVM is effective in high-dimensional spaces and is versatile, is able of handling linear and non-linear classification with the use of different kernel functions. ROC and AUC of SVM for Parkinson's disease prediction is shown in figure 4.

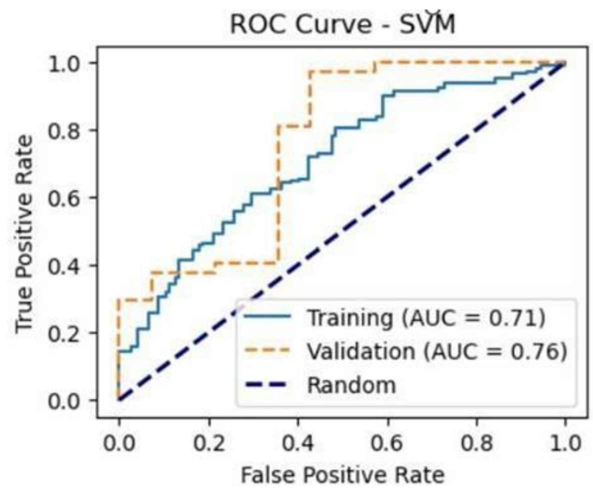

Figure 4: SVM Classification.

6. CONCLUSION

In this study on Parkinson's disease prediction, three different machine learning models, namely Logistic Regression, XGBoost, and Support Vector Machine (SVM), were evaluated. The models were trained and assessed using both training and validation datasets, and their performance was visualized through Receiver Operating Characteristic (ROC) curves. The Area Under the Curve (AUC) was utilized as a summary metric. Results indicated that all three models demonstrated promising performance in distinguishing between Parkinson's disease and non-Parkinson's instances, as evidenced by high AUC values. The ROC curves illustrated the balance among sensitivity and specificity at various probability threshold. The findings suggest that machine learning models hold potential for accurate and reliable prediction of Parkinson's disease depending on the provided characteristics. This may involve domain-specific insights or incorporating new biomarkers. Hyperparameter Tuning: Fine-tune the hyperparameters of the models to optimize their performance. Randomized search can be utilized for hyperparameter space Ensemble Methods. Investigate the use of ensemble methods, such as stacking or boosting, to combine the strengths of more than one models and raise overall predictive accurateness. External Validation: Validate the models on external datasets to assess their generalizability and robustness. This is crucial for confirming the models' effectiveness across diverse populations. Explain ability: Enhance the interpretability of systems by utilizing SHAP (SHapley Additive explanation's) values or LIME (Local Interpretable Model-agnostic Explanations) to understand the contributions of individual features. Longitudinal Data: If available, consider incorporating longitudinal data to observe the advancement of Parkinson's disease over time. This could give insights into the dynamics of the predictive features.

REFERENCES

1. Landolfi, Annamaria, et al. "Machine learning approaches in Parkinson’s disease." Current medicinal chemistry 28.32 (2021): 6548-6568.
2. Altham, Callum, Huaizhong Zhang, and Ella Pereira. "Machine learning for the detection and diagnosis of cognitive impairment in Parkinson’s Disease: A systematic review." Plos one 19.5 (2024): e0303644.
3. Kanakala, S., Prashanthi, V. (2021). Covid-19 Spread Analysis. In: Satapathy, S.C., Bhateja, V., Favorskaya, M.N., Adilakshmi, T. (eds) Smart Computing Techniques and Applications. Smart Innovation, Systems and Technologies, vol 224. Springer, Singapore
4. Dixit, Shriniket, et al. "A Comprehensive review on AI-enabled models for Parkinson’s disease diagnosis." Electronics 12.4 (2023): 783.
5. Pasluosta, Cristian F., et al. "An emerging era in the management of Parkinson's disease: wearable technologies and the internet of things." IEEE journal of biomedical and health informatics 19.6 (2015): 1873-1881.
6. Sharada, K.V., Prashanthi, V., Kanakala, S. (2022). Detection and Classification of Intracranial Brain Hemorrhage. In: Nayak, P., Pal, S., Peng, SL. (eds) IoT and Analytics for Sensor Networks. Lecture Notes in Networks and Systems, vol 244. Springer, Singapore
7. Gasser, Thomas. "Genomic and proteomic biomarkers for Parkinson disease." Neurology 72.7_supplement_2 (2009): S27-S31.
8. Patel, Urvish K., et al. "Artificial intelligence as an emerging technology in the current care of neurological disorders." Journal of neurology 268 (2021): 1623-1642.