Diagnosis and Treatment Suggestion for Parkinson Diseases Using ML Model

1st Dr.B.Rajesh Kumar
Professor and Head
Department of Computer Science and Engineering
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
hod.cse@dsce.ac.in

2nd Ramya R
Assistant Professor
Department of Computer Science and Engineering
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
ramyar@dsce.ac.in

3rd Yogesh S
UG Scholar
Department of Computer Science and Engineering
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
yogeshs.21@dsce.ac.in

3rd Vithyashri V R
UG Scholar
Department of Computer Science and Engineering
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
vithyashrivr.21@dsce.ac.in

4th Sabari Ganesh E
UG Scholar
Department of Computer Science and Engineering
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
sabariganeshe.21@dsce.ac.in

5th Ajay Krishna R
UG Scholar
Department of Computer Science and Engineering
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
ajaykrishnar.21@dsce.ac.in

***Abstract*—Machine learning algorithms have been advanced in various fields, including healthcare. This paper explores the Diagnosis and Treatment planning for Parkinson's Disease. By feeding the Existing Patient Data, ML models can accurately Diagnosis the disease and suggest the optimal treatment options. The paper discusses different ML approaches, like supervised learning, unsupervised learning, and reinforcement learning, and their potential benefits in improving diagnostic accuracy and treatment efficacy for Parkinson's Disease.**

Keywords— artificial neural networks, convolutional neural networks, deep learning, diabetes, disease diagnosis, heart disease, kidney disease, machine learning, causal ML

I. INTRODUCTION

In medical domains, machine learning (ML) primarily focuses on developing algorithms and techniques to determine whether a system's behavior is correct when diagnosing a disease. This diagnosis identifies the disease or conditions that explain a person's Conditions and signs [1]. Although there is no cure, treatment options vary and include medications, lifestyle adjustments, and surgery. While Parkinson's itself is not fatal, disease complications can be serious. Parkinson's Disease (PD) is a neuro-degenerative disorder characterized by the progressive loss of dopamine-producing neurons in the substantia nigra, a region of the brain. This loss leads to various symptoms, Like tremors, rigidity, bradykinesia, and postural instability. Early and accurate diagnosis of PD is crucial for initiating appropriate treatment Plan and improving the Patients life quality.

Traditional way of diagnosing PD relies on clinical assessments and neurological examinations. However, these methods can be prone to errors, especially in the early stages of the disease. In recent years, ML has emerged as one of the best tools for diagnosing the Diseases, offering the Efficient way to improve the accuracy in PD diagnosis.

This paper explores the Diagnosis and Treatment planning for Parkinson's Disease. We will discuss the various data that can be used for Training the ML models in PD. We will also review the different ML algorithms that can be used to Diagnosis and treatment planning for PD, such as decision trees, random forests, support vector machines, and deep learning models.

Finally, we will discuss the challenges and limitations of using machine learning for PD diagnosis and treatment planning. Despite the promising potential of machine learning, it is important to recognize that these methods are not perfect and require careful validation and clinical evaluation.

By understanding the Diagnosis and Treatment planning for Parkinson's Disease., we can gain insights into the future of personalized medicine for this debilitating disease.

II. RECENT WORKS

The field of ML, specifically in the context of diagnosing the Parkinson's Disease (PD) has seen significant advancements in recent years. Most of the studies have explored various ML algorithms to accurately classify PD patients from healthy controls.

1. Data Collection

The foundation of any ML project lies in the quality and quantity of the data used. Researchers have employed diverse data collection methods, including:

- Traditional medical records, patient histories, and symptom assessments provide valuable information for PD diagnosis.

- Techniques like MRI, PET, and SPECT scans offer insights into brain structure and function, aiding in the identification of PD-related abnormalities.

2. Data Preprocessing

Raw data often requires extensive preprocessing to ensure its suitability for ML models. Common Preprocessing steps include:

- Handling missing values, outliers, and inconsistencies to maintain data integrity.
- Deriving meaningful features from raw data, such as statistical measures or time-domain features.
- Identifying the most relevant features to improve the Performance of the Model and reduce computational costs.

3. Exploratory Data Analysis

- Feature Engineering: Researchers have explored various feature engineering techniques to extract meaningful information from the data Collected from the Hospitals. These techniques include time-domain features, frequency-domain features, and statistical features.
- Visualization: It can provide insights into the underlying patterns and relationships. Techniques such as scatter plots, histograms, and heat maps have been used to explore the characteristics of PD data.

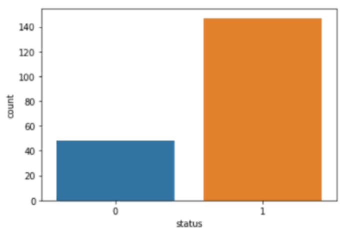

Fig.1

4. Dataset Balancing and Scaling:

- Class Imbalance: PD datasets often suffer from class imbalance, where the number of samples in one class (e.g., PD patients) is significantly smaller than the number of samples in the other class (e.g., healthy controls). Techniques such as oversampling, under-sampling, and SMOTE (Synthetic Minority Over-sampling Technique) have been employed to address this issue.
- Feature Scaling: Normalizing or standardizing features can improve the performance of ML algorithms. Techniques like min-max scaling and z-score normalization have been commonly used.

5. ML Models Training and Evaluation

- Model Selection: Various ML algorithms have been applied to detect PD, including Decision Trees, Random Forests, Support Vector Machines, and Deep Learning models.
- Model Evaluation: Metrics such as accuracy, precision, recall, F1-score, and ROC curves are used to evaluate the performance of the trained models.
- Ensemble Methods: Combining multiple models can often improve performance. Techniques like bagging and boosting have been explored in PD detection.

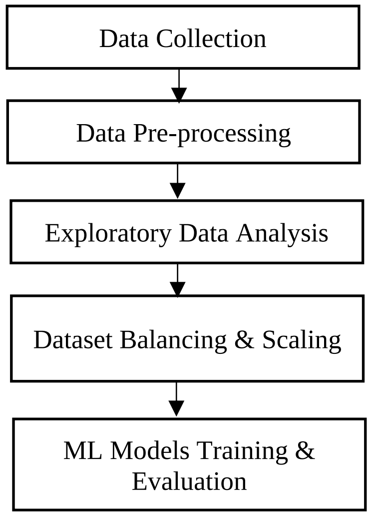

Fig.2 Project Performance

III. PROPOSED SYSTEM

This System uses various ML models for more accuracy, it includes:

- Decision Tree Classifier
- Random Forest Classifier
- Logistic Regression
- Support Vector Machine Classifier
- Naive Bayes Classifier
- K Nearest Neighbour Classifier
- XGBoost Classifier

1. Decision Tree Classifier

- **Decision Tree Classifiers** are a popular ML algorithm often used in Parkinson's Disease (PD) diagnosis. They create a tree-like model where each internal node represents a test on an attribute, and each leaf node represents a class label (in this case, PD or non-PD).
- Create a decision tree by recursively partitioning the data based on the most informative features. The Decisions and Outcomes are Represents in Each Nodes and Branch. Select a splitting criterion (e.g., Gini impurity, entropy) to determine the best feature to split the data at each node.
- Divide the data into training and testing sets to evaluate the model's performance and Calculate metrics like accuracy, precision, recall, F1-score, and AUC-ROC to assess the model's ability to correctly classify Parkinson's Disease cases.

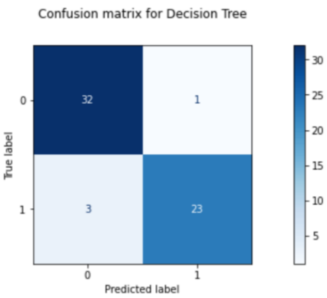

Fig.3

2. Random Forest Classifier

- **Random Forest Classifier** is a powerful machine learning algorithm that has been successfully applied to various medical domains, including Parkinson's Disease (PD) diagnosis and treatment planning. Its ability to handle large datasets, handle both numerical and categorical features, and provide feature importance measures makes it well-suited for this application.
- The algorithm creates multiple decision trees, each trained on a random subset of the data and features. The final prediction is made by combining the predictions from all the trees using a voting mechanism (e.g., majority vote).
- Random Forest provides a feature importance score for each feature, indicating its contribution to the model's accuracy. The trained Random Forest model is used to predict the likelihood of PD in new patients based on their input features. Based on the diagnosis and the patient's specific characteristics, the model can suggest appropriate treatment options, such as medication, therapy, or surgical interventions.

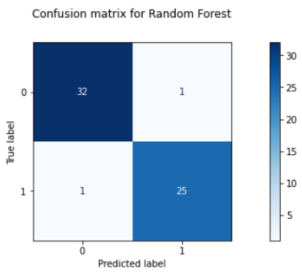

Fig.4

3. Logistic Regression

- **Logistic Regression** is a statistical model that predicts the probability of an event occurring. In the context of Parkinson's Disease (PD), this event could be the presence or absence of the disease.
- The logistic regression model is trained on a dataset of patients with known PD status. The model learns the relationship between the features and the probability of having PD. For a new patient, the model predicts the probability of having PD based on their feature values. If the probability exceeds a predefined threshold, the patient is classified as having PD.
- Logistic regression provides coefficients for each feature, indicating their relative importance in predicting PD. This can help clinicians understand the factors contributing to the diagnosis.

Fig.5

4. Support Vector Machine (SVM)

- **Support Vector Machine (SVM)** is a powerful ML algorithm widely used for classification and regression tasks in Parkinson's Disease (PD) diagnosis and treatment planning, SVM offers several advantages due to its ability in handling the high-dimensional data and find optimal decision boundaries.
- Relevant features from PD patients are extracted, such as voice characteristics (pitch, intensity, jitter, shimmer), gait parameters, or medical history.
- The extracted features are preprocessed to handle missing values, and outliers, and normalize data for consistent scaling. An SVM model is trained on a labeled dataset containing features of PD patients.
- SVM can be used to predict the effectiveness of different treatment options for individual PD patients based on their specific features and disease progression. SVM can help predict the rate of disease progression, allowing for more proactive treatment planning and monitoring.

```
Test Set Accuracy: 0.8135593220338984
Train Set Accuracy: 0.8638297872340426
```

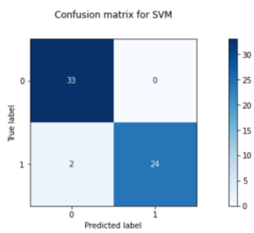

Fig.6

5. Navi Bayes Classifier

- **Naive Bayes** is a probabilistic classification algorithm that's often used in recommendation systems. It can be used in Parkinson's Disease diagnosis and treatment planning for Extracting features from patient data, such as tremors, rigidity, bradykinesia, and speech impairment. Consider factors like age, family history, and other medical conditions. It includes data points such as gender, occupation, and lifestyle.

- It Calculate the probability of a patient having Parkinson's Disease (PD) given a set of symptoms and other factors. Apply Bayes' theorem to compute this probability based on the prior probability of PD, the probability of symptoms given PD, and the probability of symptoms.
- Based on the calculated probability, assess the patient's risk level for disease progression Suggest appropriate treatments, such as medication, therapy, or surgery, tailored to the patient's risk level and symptoms.

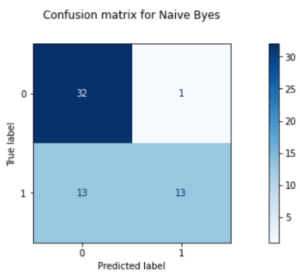

Fig.7

6. K-Nearest Neighbor (KNN)

- **K-Nearest Neighbour (KNN)** is a ML algorithm that classifies new data points based on the majority class of their k nearest neighbors in the training set. It's a non-parametric algorithm, meaning it doesn't make assumptions about the underlying data distribution.
- Based on the diagnosis of the patient's and the treatments that were effective for similar patients in the training set. It Estimates the likelihood of different treatment outcomes for a given patient and Tailor treatment plans to individual patient characteristics and preferences.
  - KNN can be a valuable tool for Parkinson's disease diagnosis and treatment planning, it's often combined with other ML algorithms or used as a baseline for comparison to explore more complex models.

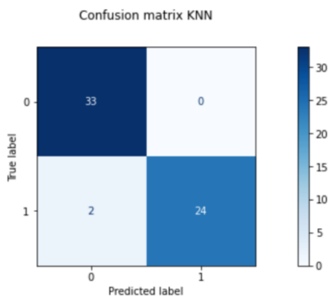

Fig.8

7. XGBoost Classifier

In this section, we have trained an XGBoost Classifier, for the classification of Instances to be Parkinson's or Not. The following parameters of the XGBoost Classifier have been optimized in this section:

- **Max Depth**: This value is used to determine the Maximum Depth of the Tree.
- **ETA**: This is also known as Learning Rate.
- **Reg_Lambda**: This is the L2 Regularization for the weights.
- **Random State**: This is used to evaluate and determine the performance of the model based on different random states.

The Parameter Optimization has been performed using GridSearchCV with the following parameters:

- **Scoring Parameter**: F1 Score
- **Cross Validation**: 3

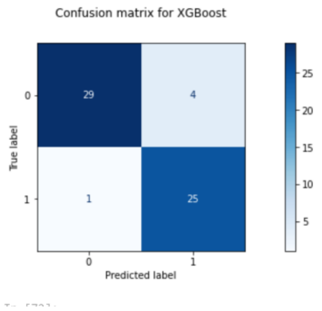

Fig.9

IV. System Approach and Result

1. Data Collection and Preprocessing

Gathering relevant data from various sources, such as health records, patient surveys, and medical literature. Cleaning and preparing the data for analysis by handling missing values, outliers, and inconsistencies.

2. Feature Extraction and Selection

Extracting relevant features from the collected data, such as age, gender, symptoms, medical history, and genetic information. Selecting the most informative features using techniques like correlation analysis, mutual information, or feature importance from ML models.

3. Model Training and Evaluation

Training various machine learning models on the preprocessed and selected features, such as decision trees, random forests, logistic regression, support vector machines, naive Bayes, K-nearest neighbors, and XGBoost. Evaluating the performance of each model using appropriate metrics like accuracy, precision, recall, F1-score, and AUC-ROC.

4. Model Selection and Deployment

Selecting the best-performing model based on the evaluation metrics. Integrating the selected model into a web application or mobile app for real-time diagnosis and treatment plan recommendations.

5. *Treatment Plan Recommendation*

Generating personalized treatment plans based on the diagnosis from the selected model, considering factors like medication, therapy, or lifestyle changes. Customizing the treatment plan to individual patient needs and preferences.

```
For Test Set
Accuracy : 0.9152542372881356
Precision : 0.8620689655172413
Recall : 0.9615384615384616
R2 Score : 0.6561771561771561
```

Fig.10 Test set Result

```
For Train Set
Accuracy : 1.0
Precision : 1.0
Recall : 1.0
R2 Score : 1.0
```

Fig.11 Train set Result

	Metric	DT	RF	LR	SVM	NB	KNN	XGB
0	Accuracy	0.932203	0.966102	0.830508	0.966102	0.762712	0.966102	0.915254
1	F1-Score	0.920000	0.961538	0.782609	0.960000	0.650000	0.960000	0.909091
2	Recall	0.884615	0.961538	0.692308	0.923077	0.500000	0.923077	0.961538
3	Precision	0.958333	0.961538	0.900000	1.000000	0.928571	1.000000	0.862069
4	R2-Score	0.724942	0.862471	0.312354	0.862471	0.037296	0.862471	0.656177

Fig.12 Result of Implementation

References

1. Grover, S.; Bhartia, S.; Yadav, A.; Seeja, K.R. Predicting the severity of Parkinson's disease using deep learning. Procedia Comput. Sci. 2018, 132, 1788–1794. [CrossRef]
2. Sriram, T.V.; Rao, M.V.; Narayana, G.S.; Kaladhar, D.; Vital, T.P.R. Intelligent Parkinson disease prediction using machine learning algorithms. Int. J. Eng. Innov. Technol. (IJEIT) 2013, 3, 1568–1572.
3. Esmaeilzadeh, S.; Yang, Y.; Adeli, E. End-to-end Parkinson disease diagnosis using brain mr-images by 3d-CNN. arXiv 2018, arXiv:1806.05233.
4. Warjurkar, S.; Ridhorkar, S. A Study on Brain Tumor and Parkinson's Disease Diagnosis and Detection using Deep Learning. In Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), Online, 27–28 August 2021; Atlantis Press: Amsterdam, The Netherlands, 2021; pp. 356–364.
5. SherlyPusphaAnnabel, L.; Sreenidhi, S.; Vishali, N. A Novel Diagnosis System for Parkinson's Disease Using K-means Clustering and Decision Tree. In Communication and Intelligent Systems; Springer: Berlin/Heidelberg, Germany, 2021; pp. 607–615.
6. Asmae,O.;Abdelhadi,R.; Bouchaib, C.; Sara, S.; Tajeddine, K. Parkinson's disease identification using KNN and ANN Algorithms based on Voice Disorder. In Proceedings of the 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 19–20 March 2020; IEEE: Manhattan, NY, USA, 2020; pp. 1–6.
7. Güriiler, H. A novel diagnosis system for Parkinson's disease using a complex-valued artificial neural network with k-means clustering feature weighting method. Neural Comput. Appl. 2017, 28, 1657–1666. [CrossRef]
8. Shetty, S.; Rao, Y. SVM-based machine learning approach to identify Parkinson's disease using gait analysis. In Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; IEEE: Manhattan, NY, USA, 2016; Volume 2, pp. 1–5.
9. Nurrohman, A.; Abdullah, S.; Murfi, H. Parkinson's disease subtype classification: Application of decision tree, logistic regression and logit leaf model. In AIP Conference Proceedings; AIP Publishing LLC: Melville, NY, USA, 2020; Volume 2242, p. 030015.