PARKISON'S DISEASE PREDICTION USING MACHINE LEARNING

Koyala Hari Krishna[1] , Kondiparthi Charan[2] , Mogilicharla Sai Gopinadh[3]
Department of Electronics and Communication Engineering[1,2,3]
Vasireddy Venkatadri Institute of Technology , Nambur , Guntur

ABSTRACT

Parkinson's Disease (PD) is a neurodegenerative disorder that primarily affects motor control, leading to symptoms such as tremors, rigidity, and bradykinesia. Early detection is critical to managing the progression of the disease, as current treatments focus on alleviating symptoms rather than curing the disease. Machine learning (ML) techniques have gained prominence in medical diagnostics, offering potential for early detection and improved accuracy in predicting Parkinson's Disease. This paper presents an overview of various ML approaches for PD prediction, leveraging clinical and physiological data .The study explores different machine learning algorithms, including decision trees, support vector machines (SVM), k-nearest neighbours (KNN), and deep learning techniques, to predict the likelihood of Parkinson's Disease in individuals. The dataset typically used in these studies consists of clinical features such as voice recordings, gait analysis, and other non-invasive diagnostic information, as well as demographic data. Feature extraction, preprocessing, and dimensionality reduction techniques like Principal Component Analysis (PCA) are utilized to enhance model performance and accuracy.

Performance evaluation of the models is based on metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). The results demonstrate that machine learning models, particularly ensemble learning techniques and deep learning models, can achieve high accuracy in classifying PD patients from healthy individuals. The application of these models in clinical settings could revolutionize the early diagnosis process, reducing the reliance on subjective clinical evaluations. This paper also addresses challenges in the field, such as the imbalance of data, interpretability of models, and the need for large, diverse datasets for robust model training. In conclusion, machine learning offers significant potential for enhancing the prediction and diagnosis of Parkinson's Disease, improving patient outcomes through timely intervention and personalized treatment strategies. Further research and optimization are needed to fully integrate these models into clinical practice.

Keywords: Parkinson's Disease, Ensemble, Accuracy, Training data, Testing data, Nerve Disorder

---

I.INTRODUCTION

1.1 PROJECT INTRODUCTION:

The application of (ML) in Parkinson represents a significant advancement in medical diagnostics. ML algorithms can analyse vast amounts of data from various sources, including medical records, genetic information, and neuroimaging scans, to identify patterns and biomarkers associated with Parkinson's disease. By leveraging these algorithms, researchers and clinicians can achieve earlier and more accurate diagnoses. For instance, machine learning models can process data from wearable devices that monitor patients' movements, detecting subtle motor changes that may indicate the onset of Parkinson's before clinical symptoms become apparent. Additionally, ML can assist in distinguishing Parkinson's from other neurodegenerative disorders with similar symptoms, thereby improving diagnostic accuracy. The integration of machine learning in Parkinson's detection not only enhances early diagnosis but also aids in monitoring disease progression and tailoring personalized treatment plans, ultimately improving patient outcomes and quality of life.

1.2 SCOPE :

The scope of Parkinson's disease detection using machine learning techniques encompasses leveraging diverse biomedical data sources, including voice recordings, clinical assessments, and imaging data, to develop robust algorithms. These algorithms aim to accurately differentiate individuals with Parkinson's disease from healthy individuals, enabling early detection and timely intervention. By implementing scalable solutions, such as automated diagnostic tools, the scope extends to improving diagnostic accuracy and monitoring disease progression over time. This approach not only enhances clinical decision-making but also facilitates the development of personalized treatment strategies tailored to individual patient needs, thereby potentially improving overall patient outcomes and quality of life the implementation of real-time monitoring systems that can continuously assess disease severity and treatment efficacy, thereby enabling personalized care plans and interventions tailored to individual patient needs. Ultimately, the integration of machine learning in Parkinson's disease detection aims to enhance diagnostic accuracy, improve clinical decision-making, and ultimately contribute to better management strategies and outcomes for patients affected by this neurodegenerative.

1.3 PROJECT OVERVIEW:

The objective of this project is to build machine learning models capable of accurately diagnosing Parkinson's disease by analyzing relevant biomedical data, aiming to distinguish between individuals with the disease and those without, thereby facilitating early and reliable detection for effective treatment and management strategies.Early detection is critical for managing the disease effectively, but diagnosis can be challenging due to the variability of symptoms. Machine learning offers a promising solution by analysing patterns in various types of data, such as voice recordings, handwriting samples, and gait analysis. By identifying subtle changes associated with Parkinson's, machine learning models can assist healthcare professionals in diagnosing the disease earlier and more accurately.

1.4 OBJECTIVE:

The primary objective of using (ML) for Parkinson is to develop an automated, accurate, and efficient system that can identify the presence and progression of Parkinson's disease at an early stage. This system aims to assist healthcare professionals in diagnosing the disease by analysing patient data, such as voice recordings, handwriting samples, gait analysis, and other biomarkers, using advanced ML algorithms.

II.LITERATURE SURVEY

A literature survey on the use of machine learning for the prediction of Parkinson's disease (PD) shows that there has been a growing interest in this area in recent years. Several studies have used machine learning algorithms to predict PD using various types of data, such as demographic information, clinical measurements, and neuroimaging data. One of the most commonly used data sources for PD prediction is demographic information such as age and gender. Studies have shown that age and gender are important factors in the prediction of PD [1][2]. Additionally, clinical measurements such as the Hoehn and Yahr scale, which measures the severity of PD symptoms, have also been found to be important for PD prediction [3].

Another data source that has been used for PD prediction is neuroimaging data, such as magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. Studies have shown that structural and functional changes in the brain can be used to predict PD [4][5]. For instance, a study used a deep learning model to diagnose PD using speech signals, this approach showed an accuracy of 86.89% [6]. Several machine learning algorithms have been used for PD prediction, including logistic regression, k-nearest neighbors, support vector machines, and deep learning algorithms. Studies have shown that these algorithms can achieve high accuracy in PD prediction [7][8]. Furthermore, some studies have also investigated the importance of different features in PD prediction using feature importance analysis [9]. Overall, the literature suggests that machine learning has the potential to be a powerful tool for the early detection of PD. However, more studies are needed to validate the use of machine learning for PD prediction in larger and diverse patient populations.

III.EXISTING SYSTEM

Machine learning models for detecting Park inson's disease (PD) utilize diverse data types, including voice recordings, handwriting samples, gait analysis, and neuroimaging. For voice analysis, Support Vector Machines (SVM) and Random Forests are commonly employed to classify speech patterns indicative of PD. Handwriting analysis often uses Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to detect motor impairments. Gait and movement analysis leverage wearable sensors with models like SVMs, Random Forests, and CNNs, as well as time-series models such as Long Short-Term Memory (LSTM) networks. In neuroimaging, deep learning models, particularly CNNs, are used to analyse MRI and PET scans for structural and functional brain changes associated with PD. Additionally, clinical data from Electronic Health Records (EHRs) are analysed using gradient boosting machines and logistic regression to identify PD risk factors. Combining these data sources through multi-modal data integration creates robust predictive models using ensemble methods or multi-modal neural networks. Notable examples include models built using the Parkinson's Progression Markers Initiative (PPMI) dataset, which incorporates clinical, imaging, and biological markers to predict PD progression and diagnosis with various machine learning algorithms

IV.PROPOSED SYSTEM

The proposed system for the prediction of Parkinson's disease (PD) would be a machine learning-based approach that utilizes a combination of demographic information and clinical measurements to train a model for PD prediction. The first step in the proposed system would be to collect a dataset of patients diagnosed with PD and healthy controls. This dataset would include demographic information such as age and gender, and clinical measurements such as the Hoehn and Yahr scale.

The collected data would then be preprocessed to ensure that it is in a format that can be used by the machine learning algorithm. This can include cleaning and normalizing the data, and transforming it into a format that can be used for training and testing. Next, a machine learning model would be trained on the preprocessed data using a selected algorithm such as support vector machines (SVMs), Random Forest or Neural Network. The model would then be evaluated on a test dataset to assess its performance in predicting PD. This can include calculating metrics such as accuracy, precision, recall, and AUC.

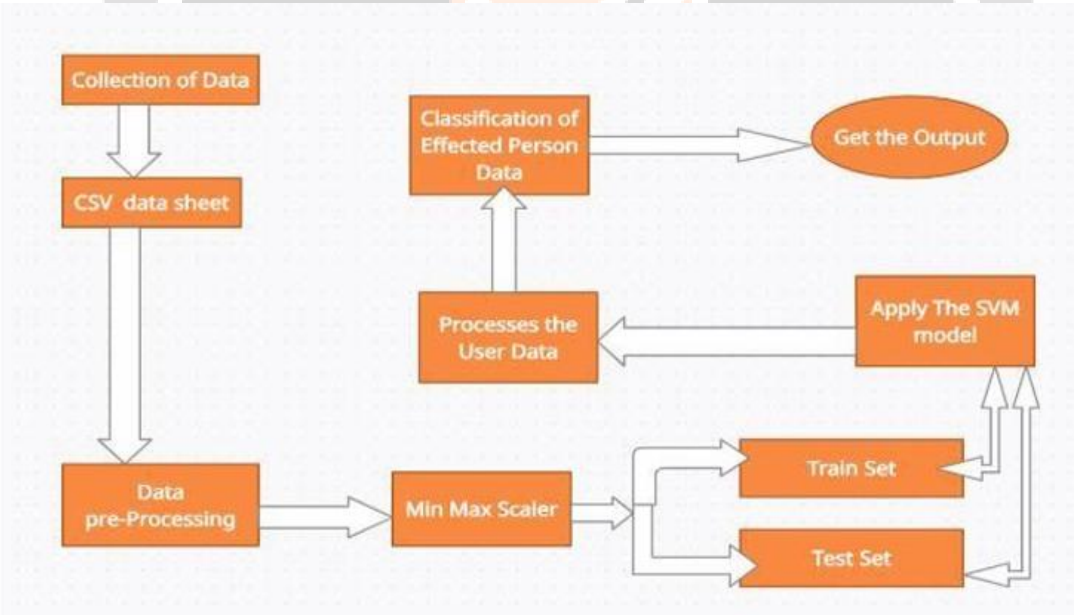

Fig.1: Architecture of the proposed system

V.IMPLEMENTATION

- **Data collection:** The first step in the proposed system would be to collect a dataset of patients diagnosed with PD and healthy controls. This dataset would include demographic information such as age and gender, and clinical measurements such as the Hoehn and Yahr scale.
- **Preprocessing:** The collected data would then be preprocessed to ensure that it is in a format that can be used by the machine learning algorithm. This can include cleaning and normalizing the data, and transforming it into a format that can be used for training and testing.

- **Model training**: A machine learning model such as SVM, Random Forest or Neural Network would be trained on the preprocessed data. The model would learn to distinguish between patients with PD and healthy controls based on the features provided in the dataset.
- **Model evaluation**: The trained model would then be evaluated on a test dataset to assess its performance in predicting PD. This can include calculating metrics such as accuracy, precision, recall, and AUC.
- **Deployment:** Once the model has been trained and evaluated, it would be deployed for use in a clinical setting for early detection of PD.
- **Inference:** The trained model would take demographic information and clinical measurements as input, and would output a binary label indicating whether the patient has PD or not.
- **Continual improvement**: The system would continuously learn and improve as more data is fed into the model, which would improve the accuracy of the system over time .

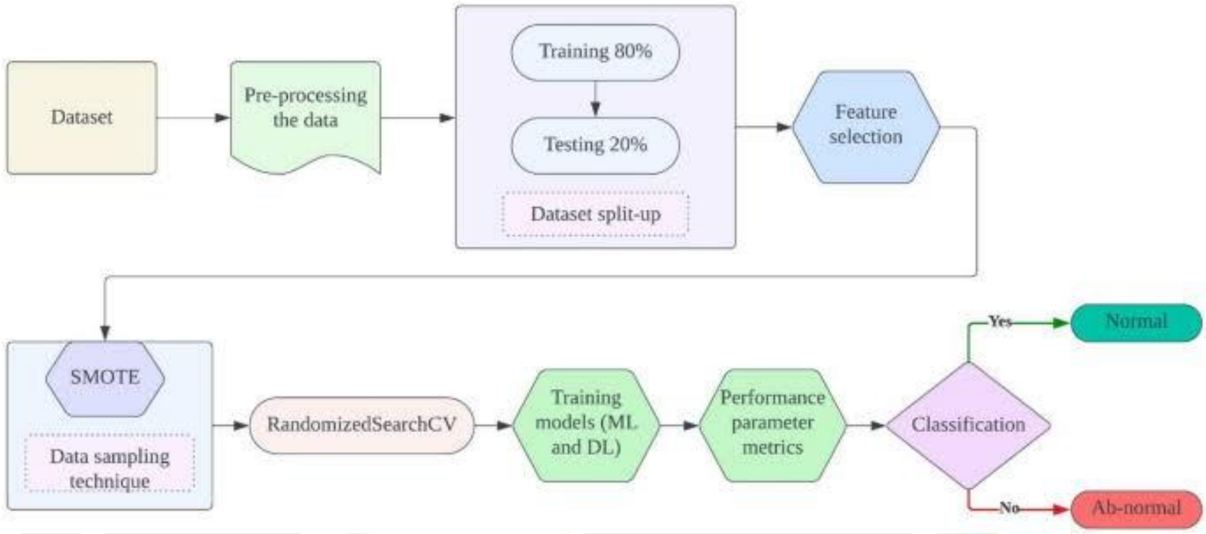

Fig.2: Data Flow of System

5.1.ALGORITHM USED:

- **Support Vector Machine:** The use of Support Vector Machines (SVM) for classifying Parkinson's disease based on gait analysis data obtained from a triaxial accelerometer. The authors investigate the ability of SVM classifiers to distinguish between Parkinson's disease patients and healthy controls using features extracted from accelerometer signals, such as gait speed, stride length, and variability of gait parameters.

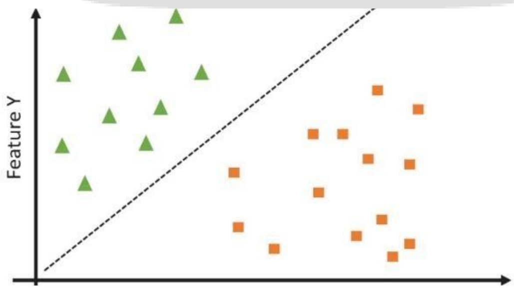

Fig.3: Support vector machine

- **Naive Bayes (NB)**: The study evaluates the effectiveness of Naive Bayes classifiers in comparison to other algorithms and provides insights into the factors influencing their performance, such as dataset size, dimensionality,and class distribution.While Naive Bayes classifiers are known for their simplicity and efficiency,this reference sheds light on their applicability and performance in practical classification tasks.

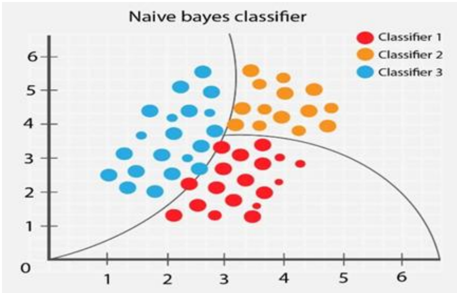

Fig.4:Navies Bayes

- **Convolutional Neural Networks (CNNs):** It utilize fractal analysis of gait data as input features and train CNN models to classify individuals as either Parkinson's disease patients or healthy controls based on gait characteristics. The study demonstrates the effectiveness of CNNs in accurately distinguishing between individuals with Parkinson's disease and healthy individuals using gait analysis data, highlighting the potential of deep learning techniques for objective assessment and diagnosis of the condition.

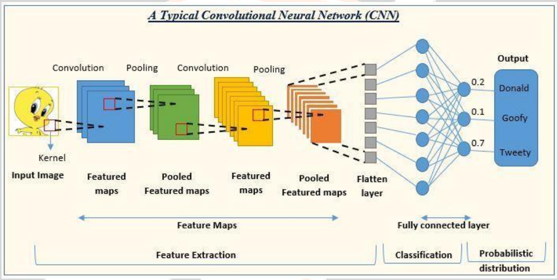

Fig.5:Conventional neural networks

V.RESULTS

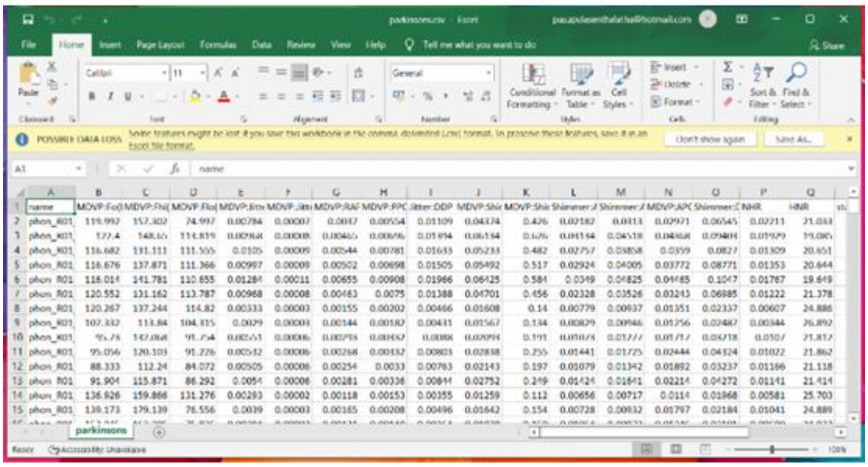

Fig.6: Data Set

Separating the features & Target

```
[ ] X = parkinsons_data.drop(columns=['name','status'], axis=1)
    Y = parkinsons_data['status']
```

```
print(X)
```

```
     MDVP:Fo(Hz)  MDVP:Fhi(Hz)  MDVP:Flo(Hz)  ...   spread2        D2       PPE
0        119.992       157.302        74.997  ...  0.266482  2.301442  0.284654
1        122.400       148.650       113.819  ...  0.335590  2.486855  0.368674
2        116.682       131.111       111.555  ...  0.311173  2.342259  0.332634
3        116.676       137.871       111.366  ...  0.334147  2.405554  0.368975
4        116.014       141.781       110.655  ...  0.234513  2.332180  0.410335
..           ...           ...           ...  ...       ...       ...       ...
190      174.188       230.978        94.261  ...  0.121952  2.657476  0.133050
191      209.516       253.017        89.488  ...  0.129303  2.784312  0.168895
192      174.688       240.005        74.287  ...  0.158453  2.679772  0.131728
193      198.764       396.961        74.904  ...  0.207454  2.138608  0.123306
194      214.289       260.277        77.973  ...  0.190667  2.555477  0.148569

[195 rows x 22 columns]
```

Fig.7: Data preprocessing

```
[ ] scaler = StandardScaler()
```

```
[ ] scaler.fit(X_train)
```

```
StandardScaler(copy=True, with_mean=True, with_std=True)
```

```
[ ] X_train = scaler.transform(X_train)

    X_test = scaler.transform(X_test)
```

```
print(X_train)
```

```
[[ 0.63239631 -0.02731081 -0.87985049 ... -0.97586547 -0.55160318
   0.07769494]
 [-1.05512719 -0.83337041 -0.9284778  ...  0.3981808  -0.61014073
   0.39291782]
 [ 0.02996187 -0.29531068 -1.12211107 ... -0.43937044 -0.62849605
  -0.50948408]
 ...
 [-0.9096785  -0.6637302  -0.160638   ...  1.22001022 -0.47404629
  -0.2159482 ]
 [-0.35977689  0.19731822 -0.79063679 ... -0.17896029 -0.47272835
   0.28181221]
 [ 1.01957066  0.19922317 -0.61914972 ... -0.716232    1.23632066
  -0.05829386]]
```

Fig.8: Data Standardization

```
model = svm.SVC(kernel='linear')
```

```
[ ] # training the SVM model with training data
    model.fit(X_train, Y_train)
```

```
SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='scale', kernel='linear',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

Fig.9: Model Training

```
# accuracy score on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

```
[ ] print('Accuracy score of training data : ', training_data_accuracy)
```

```
Accuracy score of training data :  0.8846153846153846
```

```
[ ] # accuracy score on training data
    X_test_prediction = model.predict(X_test)
    test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
```

```
[ ] print('Accuracy score of test data : ', test_data_accuracy)
```

```
Accuracy score of test data :  0.8717948717948718
```

Fig.10: Model Evaluation

```
input_data = (197.07600,206.89600,192.05500,0.00289,0.00001,0.00166,0.00168,0.00498,0.01098,0.09700,0.00563,0.00680,0.00802,0.01689,0.00339,26.77500,0.422229,0.741367,-7.348300,0.177

# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)


if (prediction[0] == 0):
  print("The Person does not have Parkinsons Disease")

else:
  print("The Person has Parkinsons")
```

```
[0]
The Person does not have Parkinsons Disease
```

Fig.11:Accuracy Score

VI.FUTURE WORK

Looking ahead, several avenues for future work in the field of predicting Parkinson's disease using machine learning present exciting opportunities for further research and innovation. Future studies could explore the integration of multi-modal data sources, such as combining clinical assessments with genetic markers, imaging studies, and wearable sensor data. Longitudinal studies tracking individuals over time are essential for understanding disease progression and treatment response in Parkinson's disease. Advances in personalized medicine offer the potential to tailor risk assessment and treatment strategies to the specific characteristics of individual patients. Emerging deep learning architectures, such as graph neural networks and attention mechanisms, hold promise for capturing complex relationships and temporal dynamics in heterogeneous data sources. Validating predictive models in diverse populations is essential to ensure their generalizability and applicability across different demographic groups and geographic regions.

VII.CONCULSION

Parkinson's disease, which affects the brain's CNS, is incurable unless it is caught early. Lack of treatment and loss result from late discovery. Therefore, it is important to diagnose it early. We used the machine learning technique Support Vector Machine for early illness identification. SVM is the best method that delivers the best accuracy (up to 86%) comparison to other algorithms to forecast the commencement of the disease, allowing for early treatment and perhaps saving a life. We checked our Parkinson disease data and discovered this. In conclusion, predicting Parkinson's disease using machine learning models is a valuable and promising area of research and clinical application. However, the success of such predictive models depends on various factors, including data quality, feature selection, model selection, and evaluation metrics.

VIII.REFERENCES

[1] DraganaMiljkovic et al, “Machine Learning and DataMining Methods for Managing Parkinson’s Disease”LNAI 9605, pp 209-220, 2016.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Dr..R.GeethaRamani, G. Sivagami, ShomonaGraciajacob “Feature Relevance Analysis and Classification 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] 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.