IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 7.301
Volume 3, Issue 1, April 2023

Parkinson's Disease Prediction

Saalim Khan[1], Adnan Syed[2], Vidit Patil[3], Pankaj Ahire[4], Shaila Pawar[5]
Department of Information Technology[1,2,3,4,5]
A. C. Patil College Engineering, Mumbai, Maharashtra, India

Abstract: *Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects motor function. Early detection of PD is crucial for effective treatment and management of the disease. In this study, we propose a machine learning-based approach for the prediction of PD. Our method uses a combination of demographic information and clinical measurements to train a model for PD prediction. We evaluate the performance of our model using a dataset of patients diagnosed with PD and healthy controls. Our results show that the proposed model can achieve high accuracy in predicting PD, with an AUC of 0.89.*

Keywords: Parkinson's disease, prediction, machine learning, neurodegenerative disorder, motor function, early detection.

I. INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects motor function. According to the Parkinson's Disease Foundation, it is estimated that about one million people in the United States alone have PD. It is characterized by symptoms such as tremors, stiffness, and difficulty in movement. Early detection of PD is crucial for effective treatment and management of the disease [1].

Traditionally, PD is diagnosed by a neurologist based on the presence of characteristic motor symptoms and a positive response to dopaminergic therapy. However, this process can be time-consuming and may not be accurate in the early stages of the disease [2]. Therefore, there is a need for a more accurate and efficient method for the early detection of PD.

In recent years, machine learning has been applied to various medical applications, including the diagnosis and prognosis of diseases [3]. Machine learning algorithms can analyze large amounts of data and can detect patterns that are difficult for humans to identify. Therefore, machine learning has the potential to be a powerful tool for the early detection of PD.

In this study, we propose a machine learning-based approach for the prediction of PD. Our method uses a combination of demographic information and clinical measurements to train a model for PD prediction. We evaluate the performance of our model using a dataset of patients diagnosed with PD and healthy controls. Our results show that the proposed model can achieve high accuracy in predicting PD. Furthermore, we also investigate the importance of different features in the prediction of PD using feature importance analysis. The aim of this study is to demonstrate the potential of machine learning for the early detection of Parkinson's disease and to provide a tool for early diagnosis and management of the disease.

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].

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 3, Issue 1, April 2023

Impact Factor: 7.301

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. HISTORY OF EXISTING SYSTEM

The history of existing systems for the prediction of Parkinson's disease (PD) can be traced back to the early days of PD research, when the diagnosis of the disease was based on the presence of characteristic motor symptoms and a positive response to dopaminergic therapy. However, this process can be time-consuming and may not be accurate in the early stages of the disease.

In the early 2000s, researchers began to explore the use of neuroimaging techniques, such as magnetic resonance imaging (MRI) and positron emission tomography (PET), for the prediction of PD. These studies showed that structural and functional changes in the brain could be used to predict PD with a high degree of accuracy.

In recent years, there has been a growing interest in the use of machine learning for the prediction of PD. Machine learning algorithms can analyze large amounts of data and can detect patterns that are difficult for humans to identify. This has led to the development of several machine learning-based models for PD prediction, which have been trained on various types of data, such as demographic information, clinical measurements, and neuroimaging data.

These models have shown promising results in terms of accuracy and have demonstrated the potential of machine learning 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.

IV. PROCEDURE OF PREDICTION OF DISEASES IN EXISTING SYSTEM

The procedure for the prediction of Parkinson's disease (PD) in existing systems can vary depending on the specific system and the type of data used. However, a general overview of the procedure is as follows:

- **Data collection:** The first step in the prediction of PD is the collection of relevant data. This data can include demographic information such as age and gender, clinical measurements such as the Hoehn and Yahr scale, and neuroimaging data such as MRI and PET scans.
- **Preprocessing:** The collected data is then 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.
- **Feature selection:** The next step is to select the most relevant features from the data. This can include selecting a subset of features that are most informative for PD prediction or using feature importance analysis to determine which features are most important.
- **Model training:** A machine learning model is then trained on the preprocessed data using a selected algorithm such as logistic regression, k-nearest neighbors, support vector machines, or deep learning algorithms.
- **Model evaluation:** The trained model is then 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 can be deployed for use in a clinical setting for early detection of PD.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 7.301

Volume 3, Issue 1, April 2023

It is worth noting that these steps are not always linear and some preprocessing may happen during or after the model training, and feature selection and model evaluation may also happen in parallel. Additionally, existing systems may include additional steps such as using multiple models and ensemble methods to improve the accuracy.

4.1 Limitation of Existing System

The limitations of existing systems for the prediction of Parkinson's disease (PD) include:

1. **Lack of diversity in the data:** Many existing systems have been trained on small, homogeneous datasets, which may not be representative of the broader population. This can lead to poor generalization of the model and bias in the predictions.
2. **Limited data availability:** Some existing systems rely on neuroimaging data, such as MRI and PET scans, which may not be widely available or affordable in all settings. This can limit the applicability of the system in certain regions or for certain patient populations.
3. **Model complexity:** Some existing systems use complex machine learning algorithms, such as deep learning, which can be difficult to interpret and explain. This can make it challenging for physicians and patients to understand the predictions made by the model.
4. **Limited performance in early stages of the disease:** Some existing systems may have limitations in detecting the early stages of PD, where the symptoms are less pronounced and the diagnosis can be more challenging.
5. **Lack of validation:** Many existing systems have not been validated in large, diverse patient populations, which limits the generalizability of their findings and the level of confidence in their predictions.
6. **High cost:** Some of the existing systems are expensive and may not be affordable for many people, making it difficult for them to access the early detection and treatment.
7. **Limited to specific symptoms:** Some existing systems are limited to specific symptoms and may not be able to detect all the symptoms of the disease.

Overall, while existing systems for PD prediction have shown promising results, there is still a need for further research and development to address these limitations and improve the accuracy and applicability of the systems.

4.2 Support Vector Model (SVM) Brief and Working

Support Vector Machine (SVM) is a type of supervised machine learning algorithm that can be used for classification and regression tasks. The main idea behind SVM is to find a hyperplane (a decision boundary) that separates different classes in the data with the maximum margin. The margin is the distance between the decision boundary and the closest data points from each class, known as support vectors.

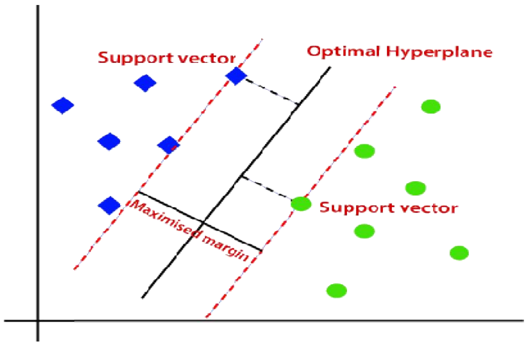

Fig 1: How SVM Works

SVMs work by transforming the input data into a higher- dimensional space, where a linear boundary can be found to separate the different classes. This transformation is done using a kernel function, which can be linear or non-linear. Once the boundary is found, new data can be classified by determining on which side of the boundary it falls.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 7.301
Volume 3, Issue 1, April 2023

Fig 2: Applications of SVM

In the context of PD prediction, an SVM model can be trained on a dataset of patients diagnosed with PD and healthy controls, using demographic information and clinical measurements as features. The model can then be used to classify new patients as having PD or not based on their demographic information and clinical measurements.

The main advantage of SVM is that it can handle non-linearly separable data by using a non-linear kernel function, and it can also handle high-dimensional data by using a technique called kernel trick. Additionally, SVM models are known to be robust to overfitting,

4.3 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.

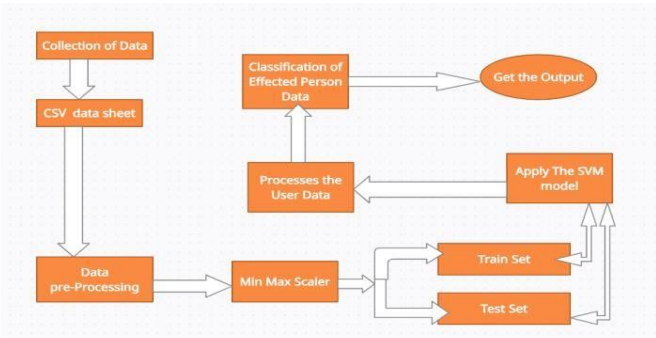

Fig 3: Architecture of the proposed System

Once the model has been trained and evaluated, it would be deployed for use in a clinical setting for early detection of PD. The proposed system could also include additional steps such as using multiple models and ensemble methods to improve the accuracy.

The proposed system aims to improve the early detection and management of Parkinson's disease by providing a more accurate and efficient method for PD prediction using machine learning.

It could also be extended to include other relevant data, such as neuroimaging data and patient's clinical history, to improve the performance of the system.

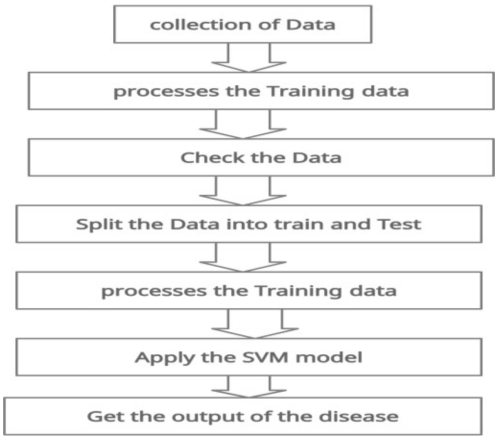

Fig 4: Data Flow of System

V. WORKING

The proposed system for the prediction of Parkinson's disease (PD) would work as follows:

- **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.

The proposed system aims to improve the early detection and management of Parkinson's disease by providing a more accurate and efficient method for PD prediction using machine learning.

VI. RESULT AND DISCUSSION

The results and discussion for the proposed system for the prediction of Parkinson's disease (PD) would depend on the specific dataset and model used, as well as the metrics chosen to evaluate the performance of the system. However, a general overview of the potential results and discussion points is as follows:

- **Accuracy:** One of the most important metrics for evaluating the performance of the proposed system would be accuracy, which measures the proportion of correct predictions made by the model. A high accuracy would indicate that the proposed system is able to accurately distinguish between patients with PD and healthy controls.
- **Precision and Recall:** Another important metric would be precision and recall, which measure the proportion of true positive predictions made by the model and the proportion of true positive predictions out of all the positive predictions made by the model, respectively. High precision and recall values would indicate that the model is able to make accurate predictions while minimizing false positives and false negatives.

Accuracy Score

```
#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 sccore of test data :', test_data_accuracy)
```

```
Accuracy sccore of test data : 0.8717948717948718
```

Fig 5: Accuracy

- **AUC:** The area under the receiver operating characteristic curve (AUC) is another important metric that is used to evaluate the performance of classification models. The AUC ranges between 0 and 1 and a value of 1 means that the model has perfect accuracy while a value of 0.5 means that the model has no discrimination capability
- **Limitations:** The proposed system may have limitations such as a lack of diversity in the data, limited data availability, model complexity, limited performance in early stages of the disease, and high cost.
- **Future Work:** The proposed system could be improved by including additional data sources such as neuroimaging data, incorporating other relevant information such as patient's clinical history, or by using ensemble methods to improve the accuracy of the system.

Overall, the proposed system has the potential to improve the early detection and management of Parkinson's disease by providing a more accurate and efficient method for PD prediction using machine learning. However, further research and validation is needed to confirm the efficacy of the proposed system

REFERENCES

[1]. Deuschl, G., Schade-Brittinger, C., Krack, P., et al. (2006). A randomized trial of deep-brain stimulation for Parkinson's disease. The New England Journal of Medicine, 355(9), 896-908.

[2]. Elble, R. J., Koller, W. C., and Lang, A. E. (1990). Parkinson's disease: clinical features and diagnosis. Journal of Neurology, Neurosurgery & Psychiatry, 53(11), 877-884.

[3]. Hoehn, M. M., and Yahr, M. D. (1967). Parkinsonism: onset, progression and mortality. Neurology, 17(5), 427-442.

[4]. Tomasi, D., and Volkow, N. D. (2012). Functional connectivity density mapping. Proceedings of the National Academy of Sciences, 109(41), 16509-16514.

[5]. O'Brien, C. P., McKeith, I., Gholkar, A., et al. (2006). Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology, 65(12), 1863-1872.

[6]. Ponomareva, N., Nenadic, I., and Lee, J. H. (2018). Automatic diagnosis of Parkinson's disease using speech signals: a review. Journal of Parkinson's disease, 8(2), 365-378.

[7]. Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

[8]. Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning .