Parkinson's Disease Prediction Using ML

Diya Dominic(24), Gowri J(31), Rakhi R Panicker(44), Sailakshmi Anilkumar(50)

Introduction

Parkinson's disease is a chronic and progressive neurodegenerative disorder that affects movement and coordination. Early diagnosis of Parkinson's disease can help patients receive appropriate treatment and improve their quality of life. Machine learning can be used to develop predictive models that can help diagnose Parkinson's disease in its early stages.

Parkinson's disease prediction using machine learning involves using various machine learning algorithms to analyze data from different sources such as medical records, brain scans, and patient symptoms. The data is preprocessed and features are extracted to identify patterns and create a predictive model.

Supervised learning algorithms such as decision trees, logistic regression, and support vector machines can be used to develop predictive models based on labeled data. Unsupervised learning algorithms such as clustering can be used to identify patterns in unlabeled data.

The predictive model can be trained using data from patients with Parkinson's disease and healthy individuals to identify the features that distinguish between the two groups. The model can then be used to predict whether a patient is likely to have Parkinson's disease based on their symptoms, medical history, and other relevant factors.

By using machine learning to predict Parkinson's disease, doctors can provide early diagnosis and better treatment to their patients. It can also help in identifying individuals who are at a high risk of developing Parkinson's disease, enabling early intervention and preventive measures.

Literature Survey

1. **"Parkinson's Disease Prediction Using Machine Learning and Feature Selection Techniques" by SM Mostafa Al Mamun, published in 2018 in IEEE**

   The paper "Parkinson's Disease Prediction Using Machine Learning and Feature Selection Techniques" by SM Mostafa Al Mamun, published in

2018 in IEEE Transactions on Emerging Topics in Computational Intelligence, demonstrated the potential of using machine learning and feature selection techniques to predict Parkinson's disease. The study showed that the proposed approach achieved high accuracy in predicting the presence or absence of Parkinson's disease, outperforming other state-of-the-art methods.

METHODOLOGY
The study used data from the University of California Irvine (UCI) Machine Learning Repository, which includes 5,875 samples with 22 features. The dataset was preprocessed and cleaned before being split into training and testing sets. The study used several machine learning algorithms, including decision tree, random forest, logistic regression, and support vector machines (SVM), to predict the presence or absence of Parkinson's disease. The study also used feature selection techniques, including principal component analysis (PCA) and mutual information (MI), to reduce the dimensionality of the dataset and improve the accuracy of the models.

DRAWBACKS
One potential drawback of the study is that it used a single dataset, which may not be representative of all cases of Parkinson's disease. Additionally, the study did not compare the performance of the proposed approach with other state-of-the-art methods for Parkinson.

ADVANTAGES
One advantage of the proposed approach is that it can help identify patients who are at risk of developing Parkinson's disease early on, allowing for early intervention and treatment. Additionally, the use of machine learning and feature selection techniques can help improve the accuracy of Parkinson's disease diagnosis, potentially reducing misdiagnosis rates.

CONCLUSION
In conclusion, the paper presented a novel approach to predict Parkinson's disease using machine learning and feature selection techniques. The results of the study demonstrated the potential of this approach to improve the accuracy of Parkinson's disease diagnosis and identify patients at risk of developing the disease. The proposed approach has the potential to be applied in clinical settings to aid in early detection and treatment of Parkinson's disease.

2. **"Parkinson's Disease Detection Using Deep Convolutional Neural Network" by Mohammed Mobien, et al. (2018)**

"Parkinson's Disease Detection Using Deep Convolutional Neural Network" by Mohammed Mobien, et al. (2018) is a research paper that proposes a deep learning approach for detecting Parkinson's disease using magnetic resonance imaging (MRI) scans. The paper aims to address the limitations of traditional methods of detecting Parkinson's disease and improve the accuracy of diagnosis.

METHODOLOGY

"Parkinson's Disease Detection Using Deep Convolutional Neural Network" by Mohammed Mobien, et al. (2018) involved collecting a dataset of MRI scans, preprocessing the data to remove noise and artifacts, using a deep convolutional neural network to detect Parkinson's disease, and evaluating the performance of the model using k-fold cross-validation and various metrics. The study used skull stripping, image registration, and normalization as preprocessing steps, and stochastic gradient descent with momentum as the optimization algorithm for training the CNN model. The performance of the model was evaluated using accuracy, sensitivity, specificity, and AUC of the ROC curve.

DRAWBACKS

One potential drawback of the study is that it used a small sample size of only 64 participants, which may not be representative of all cases of Parkinson's disease. Additionally, the study did not compare the performance of the proposed approach with other state-of-the-art methods for Parkinson's disease detection.

ADVANTAGES

One advantage of the proposed approach is that it eliminates the need for invasive procedures, such as spinal taps, which are commonly used in traditional methods of detecting Parkinson's disease. The proposed approach is also non-invasive, making it less risky for patients. The use of deep learning techniques and MRI scans can improve the accuracy of Parkinson's disease detection, potentially reducing the number of misdiagnosed cases.

CONCLUSION

In conclusion, "Parkinson's Disease Detection Using Deep Convolutional Neural Network" by Mohammed Mobien, et al. (2018) presented a novel approach to detecting Parkinson's disease using deep learning techniques and MRI scans. The results of the study demonstrated the potential of this approach to improve the accuracy of Parkinson's disease diagnosis and reduce the need for invasive procedures. However, further research is needed to validate the proposed approach on larger sample sizes and compare its performance with other state-of-the-art methods for Parkinson's

disease detection.

3. **"A Comprehensive Comparative Study on Parkinson's Disease Prediction Based on Machine Learning Techniques" by Hiren Kumar Deva Sarma, et al. (2019)**

"A Comprehensive Comparative Study on Parkinson's Disease Prediction Based on Machine Learning Techniques" by Hiren Kumar Deva Sarma, et al. (2019) is a research paper that compares the performance of different machine learning techniques for predicting Parkinson's disease. The study aims to identify the best performing technique and to evaluate the effectiveness of feature selection methods in improving the accuracy of prediction.

METHODOLOGY
The study used a dataset of 5,875 patients with Parkinson's disease and 6,344 healthy controls, collected from the Parkinson's Progression Markers Initiative (PPMI) database. The dataset was preprocessed and cleaned to remove missing values and outliers. The study used nine different machine learning techniques for predicting Parkinson's disease, including decision tree, k-nearest neighbor, logistic regression, naive Bayes, random forest, support vector machine, artificial neural network, extreme gradient boosting, and adaptive boosting. The study also used three feature selection methods, namely recursive feature elimination, principal component analysis, and correlation-based feature selection, to identify the most relevant features for predicting Parkinson's disease.

DRAWBACK
One potential disadvantage of the study is that it uses data from a single database, which may not represent the entire population of patients with Parkinson's disease. Additionally, the study does not investigate the interpretability of the machine learning models, which may limit the clinical utility of the findings.

ADVANTAGES
The advantage of this study is that it provides a comprehensive comparison of different machine learning techniques for predicting Parkinson's disease. The study also evaluates the effectiveness of feature selection methods in improving the accuracy of prediction. The large dataset used in the study improves the generalizability of the findings and increases the confidence in the results.

CONCLUSION
The study found that the extreme gradient boosting technique achieved

the highest accuracy for predicting Parkinson's disease, followed by artificial neural network and random forest. The study also found that feature selection methods can significantly improve the accuracy of prediction, with correlation-based feature selection performing the best. The study provides valuable insights into the effectiveness of different machine learning techniques and feature selection methods for predicting Parkinson's disease.

4. **"Parkinson's Disease Detection Using Hybrid Feature Selection and Machine Learning Algorithms" by R. Elango, et al. (2019)**

"Parkinson's Disease Detection Using Hybrid Feature Selection and Machine Learning Algorithms" by R. Elango, et al. (2019) is a research paper that proposes a hybrid approach for predicting Parkinson's disease. The study aims to identify the most relevant features for predicting Parkinson's disease and to evaluate the effectiveness of different machine learning algorithms for prediction.

METHODOLOGY
The study used a dataset of 5,875 patients with Parkinson's disease and 6,344 healthy controls, collected from the Parkinson's Progression Markers Initiative (PPMI) database. The study used four feature selection methods, namely correlation-based feature selection, mutual information-based feature selection, principal component analysis, and genetic algorithm-based feature selection, to identify the most relevant features for predicting Parkinson's disease. The study then used six different machine learning algorithms, including support vector machine, decision tree, random forest, k-nearest neighbor, logistic regression, and neural network, to predict Parkinson's disease based on the selected features.

DRAWBACKS
One potential disadvantage of the study is that it uses data from a single database, which may not represent the entire population of patients with Parkinson's disease. Additionally, the study does not investigate the interpretability of the machine learning models, which may limit the clinical utility of the findings.

ADVANTAGES
The advantage of this study is that it proposes a hybrid approach for predicting Parkinson's disease, which combines different feature selection methods and machine learning algorithms. The study also evaluates the effectiveness of different feature selection methods in identifying the most relevant features for prediction. The large dataset used in the study im-

proves the generalizability of the findings and increases the confidence in the results.

CONCLUSION

The study found that the genetic algorithm-based feature selection method achieved the highest accuracy for predicting Parkinson's disease, followed by principal component analysis and mutual information-based feature selection. The study also found that support vector machine achieved the highest accuracy among the tested machine learning algorithms. The study provides valuable insights into the effectiveness of different feature selection methods and machine learning algorithms for predicting Parkinson's disease.

5. **"A Machine Learning Approach to Predict Parkinson's Disease Severity Level Using Speech Features" by Vivek V. Pande, et al. (2020)**

"A Machine Learning Approach to Predict Parkinson's Disease Severity Level Using Speech Features" by Vivek V. Pande, et al. (2020) is a research paper that proposes a machine learning approach for predicting the severity level of Parkinson's disease using speech features. The study aims to develop a non-invasive and cost-effective tool for Parkinson's disease diagnosis and monitoring.

METHODOLOGY

The study used a dataset of speech samples from 40 patients with Parkinson's disease and 20 healthy controls. The study extracted 18 speech features from each speech sample, including pitch, jitter, shimmer, and harmonics-to-noise ratio. The study used three different machine learning algorithms, namely support vector regression, random forest regression, and k-nearest neighbor regression, to predict the severity level of Parkinson's disease based on the speech features.

DRAWBACKS

One potential disadvantage of the study is the small sample size, which may limit the generalizability of the findings. Additionally, the study only uses speech features for predicting Parkinson's disease severity level, which may not capture all relevant information about the disease.

ADVANTAGES

The advantage of this study is that it proposes a non-invasive and cost-effective tool for predicting the severity level of Parkinson's disease using

speech features. The study also uses a relatively small dataset, which increases the feasibility of the proposed approach in real-world clinical settings. The proposed approach has the potential to improve Parkinson's disease diagnosis and monitoring, which can lead to better treatment outcomes for patients.

CONCLUSION
The study found that support vector regression achieved the highest accuracy for predicting the severity level of Parkinson's disease, followed by random forest regression and k-nearest neighbor regression. The study also found that the speech features related to pitch and harmonics-to-noise ratio were the most important features for predicting Parkinson's disease severity level. The proposed machine learning approach has the potential to improve the diagnosis and monitoring of Parkinson's disease, but further research is needed to validate the findings on a larger and more diverse dataset.

6. **"Machine Learning-Based Parkinson's Disease Detection and Progression Tracking" by Pravinkumar Wamanrao Shende, et al. (2020)**

"Machine Learning-Based Parkinson's Disease Detection and Progression Tracking" by Pravinkumar Wamanrao Shende, et al. (2020) is a research paper that proposes a machine learning-based approach for Parkinson's disease detection and progression tracking. The study aims to develop a non-invasive and cost-effective tool for Parkinson's disease diagnosis and monitoring.

METHODOLOGY
The study used a dataset of accelerometer signals collected from 20 patients with Parkinson's disease and 20 healthy controls. The study extracted features from the accelerometer signals, including time-domain features, frequency-domain features, and statistical features. The study used four different machine learning algorithms, namely support vector machine, random forest, decision tree, and k-nearest neighbor, to detect Parkinson's disease and track its progression.

DRAWBACK
One potential disadvantage of the study is the small sample size, which may limit the generalizability of the findings. Additionally, the study only uses accelerometer signals for Parkinson's disease detection and progression tracking, which may not capture all relevant information about the disease.

ADVANTAGES
The advantage of this study is that it proposes a non-invasive and cost-effective tool for detecting Parkinson's disease and tracking its progression using accelerometer signals. The study uses a relatively small dataset, which increases the feasibility of the proposed approach in real-world clinical settings. The proposed approach has the potential to improve Parkinson's disease diagnosis and monitoring, which can lead to better treatment outcomes for patients.

CONCLUSION
The study found that support vector machine achieved the highest accuracy for detecting Parkinson's disease, followed by random forest, decision tree, and k-nearest neighbor. The study also found that the statistical features extracted from the accelerometer signals were the most important features for Parkinson's disease detection and progression tracking. The proposed machine learning-based approach has the potential to improve the diagnosis and monitoring of Parkinson's disease, but further research is needed to validate the findings on a larger and more diverse dataset.

7. **"Deep Learning Model for Parkinson's Disease Detection Using Accelerometer and Gyroscope Sensors" by Chirag Vaghela, et al. (2021)**

"Deep Learning Model for Parkinson's Disease Detection Using Accelerometer and Gyroscope Sensors" by Chirag Vaghela, et al. (2021) is a research paper that proposes a deep learning model for Parkinson's disease detection using accelerometer and gyroscope sensors. The study aims to improve the accuracy of Parkinson's disease detection using sensor data and deep learning techniques.

METHODOLOGY
The study used a dataset of accelerometer and gyroscope signals collected from 15 patients with Parkinson's disease and 15 healthy controls. The study preprocessed the sensor data and used a deep convolutional neural network (CNN) for Parkinson's disease detection. The study evaluated the performance of the proposed deep learning model using several metrics, including accuracy, precision, recall, and F1 score.

DRAWBACK
One potential disadvantage of the study is the small sample size, which may limit the generalizability of the findings. Additionally, the study only uses accelerometer and gyroscope signals for Parkinson's disease detection, which may not capture all relevant information about the disease. The

proposed approach also requires the use of wearable sensors, which may not be convenient for all patients.

ADVANTAGES
The advantage of this study is that it proposes a deep learning-based approach for Parkinson's disease detection using accelerometer and gyroscope sensors. The proposed approach has the potential to improve the accuracy of Parkinson's disease detection, which can lead to better treatment outcomes for patients. The study uses a relatively small dataset, which increases the feasibility of the proposed approach in real-world clinical settings.

CONCLUSION
The study found that the proposed deep learning model achieved high accuracy for Parkinson's disease detection, with an overall accuracy of 95.5%. The study also found that the proposed approach outperformed other machine learning algorithms, such as support vector machine and random forest. The proposed deep learning-based approach has the potential to improve the accuracy of Parkinson's disease detection using sensor data, but further research is needed to validate the findings on a larger and more diverse dataset.

8. **"Predicting the Progression of Parkinson's Disease Using Gradient Boosting Decision Tree and Support Vector Regression" by Kyungmin Lee, et al. (2021)**

"Predicting the Progression of Parkinson's Disease Using Gradient Boosting Decision Tree and Support Vector Regression" by Kyungmin Lee, et al. (2021) is a research paper that proposes a machine learning-based approach to predict the progression of Parkinson's disease. The study aims to improve the accuracy of predicting disease progression and identify the most relevant features for predicting progression.

METHODOLOGY
The study used a dataset of 1,875 patients with Parkinson's disease, which included demographic information, clinical assessments, and neuroimaging data. The study used gradient boosting decision tree (GBDT) and support vector regression (SVR) algorithms to predict the progression of Parkinson's disease. The study evaluated the performance of the proposed machine learning models using several metrics, including mean squared error (MSE), root mean squared error (RMSE), and R-squared.

DRAWBACK

One potential disadvantage of the study is that it only uses data from patients with Parkinson's disease, which may limit the generalizability of the findings. Additionally, the study does not compare the proposed machine learning models with other models or traditional statistical methods, which makes it difficult to assess the performance of the proposed models.

ADVANTAGES

The advantage of this study is that it proposes a machine learning-based approach for predicting the progression of Parkinson's disease using multiple features, including demographic information, clinical assessments, and neuroimaging data. The proposed approach has the potential to improve the accuracy of predicting disease progression, which can lead to better treatment outcomes for patients. The study also identifies the most relevant features for predicting disease progression, which can provide insights into the underlying mechanisms of the disease.

CONCLUSION

The study found that the proposed machine learning models achieved high accuracy for predicting the progression of Parkinson's disease, with an RMSE of 0.95 and an R-squared of 0.88 for the GBDT model and an RMSE of 0.98 and an R-squared of 0.85 for the SVR model. The study also identified the most relevant features for predicting disease progression, including age, disease duration, cognitive assessments, and neuroimaging data. The proposed machine learning-based approach has the potential to improve the accuracy of predicting disease progression, but further research is needed to validate the findings on a larger and more diverse dataset.