A NOVEL APPROACH TO DETECT PARKINSONISM IN MACHINE LEARNING

P. Saraswathi[1], M. Prabha[2], A. K. Vidyabharathi[3], S. Krishnaveni[4]

Article History: Received: 29.02.2023 **Revised:** 13.04.2023 **Accepted:** 06.06.2023

Abstract:

Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects the central nervous system. Symptoms of early Parkinson's disease include Tremors, Rigidity, Bradykinesia, and Postural instability. Although typically diagnosed in individuals over 60 years old, 5 to 10 percent of cases occur in those under 50. Early detection is critical for effective treatment, and machine learning algorithms can be utilized to process user input data alongside previously collected data to assess whether an individual is affected. The selection of an optimal classification algorithm for local datasets can be challenging. This study evaluated the effectiveness of several algorithms, including Logistic Regression, Support Vector Machine, and XGBoost, yielding accuracies of 79%, 87%, and 89%, respectively. The LightGBM algorithm was selected for this study as it may provide superior accuracy.

Keywords: Parkinsonism (Pd), Machine Learning, Lightgbm Algorithm.

[1]Faculty, Department of Information Technology Velammal College of Engineering and Technology Madurai, Tamilnadu, India.
[2]Faculty, Department of Information Technology Velammal College of Engineering and Technology Madurai, Tamilnadu, India.
[3] III yr Student, Department of Information Technology Velammal College of Engineering and Technology Madurai, Tamilnadu, India.
[4] III yr Student, Department of Information Technology Velammal College of Engineering and Technology Madurai, Tamilnadu, India.

Email: [1]psw@vcet.ac.in, [2]mpr@vcet.ac.in, [3]vidyabharathi2003@gmail.com, [4]krishnaswetha1976@gmail.com

DOI: 10.31838/ecb/2023.12.si6.329

1. INTRODUCTION

A. Parkinson’s Disease (PD)

Parkinson's disease is a chronic and progressive neurodegenerative disorder that primarily affects the central nervous system. The nerve cells (neurons) in the substantia nigra region of the brain either degenerate or lose their ability to produce dopamine, which is an essential neurotransmitter involved in coordinating movement and regulating various brain functions. Parkinson's symptoms have been found in patients who have lost at least 80% of the dopamine-producing cells in their substantia nigra. Dopamine and other neurotransmitters are typically used to coordinate the millions of nerve and muscle cells involved in motion. This equilibrium is broken by a lack of dopamine, which leads to tremor (trembling of the hands, arms, legs, and jaw). Leg stiffness, gradual progression, and the typical unsteadiness and clumsiness of Parkinson's disease are all present. Parkinson's disease's fundamental cause is mostly unknown. Nevertheless, oxidative harm, environmental pollutants, hereditary factors, and accelerated ageing are explanations for the disease's potential origins. One variation in a Parkinson's disease quality, initially identified in 1997 and thought to account for 5% of acquired cases, was discovered by experts in 2005.

B. Symptoms

PD symptoms can be broadly divided into two groups: those that are motor-related and those that are not (non-motor). The involuntary, rhythmic movement of the hands, arms, legs, and jaw is known as a tremor. The arms, shoulders, and neck are where muscle rigidity, or stiffness of the limbs, is most prevalent. gradual loss of spontaneity, which frequently results in mental decline or slowed reaction times, voice changes, altered facial expressions, etc. A gradual lack of automatic movement may cause decreased blinking, swallowing, and drooling. a hunched, flexed position with elbow, knee, and hip bends. unstable balance or walk.

A doctor's opinion is required for the diagnosis in addition to a neurological examination because there is no widely used diagnostic instrument, such as a blood test. Although imaging tools like DaTscan can assist a doctor in better understanding a patient's health, they are prohibitively expensive and insufficient for a comprehensive diagnosis. There has additionally been work on surveying current side effects utilizing AI ways to deal with analyze PD. Effects on vocal patterns are one of the symptoms of PD that have not been adequately investigated.

A framework for a Parkinson's disease diagnosis that is earlier, more accurate, and less expensive is the objective of this project. It is based on vocal impairment, a distinct defining characteristic.

In this study, we seek to construct a classification system that is accurate enough to diagnose PD with at least a 94% accuracy utilising coarse audio samples from PD patients and healthy test subjects. Several machine learning models, including Logistic Regression, XGBoost, and Support Vector Machine, were evaluated and contrasted in order to increase accuracy with less data. The LightGBM model was found to be the most accurate.

Related Research Work

Imaging the brain with tools like MRI, CT, or PET is one of the most effective methods for PD diagnosis. The drawback of those techniques is that they are expensive, and Due to this, we focus on less expensive methods. [1]In this study, an early Parkinson's disease detection method integrating LeNet -A Convolutional Neural Network and LSTM- Long Short Term Memory was put proposed. [2]In this study, a neural network was developed to forecast the severity of the condition, and a machine learning model (Random Forest Classifier) was developed to identify the disorder. [3]They used four machine learning models, including K-Nearest Regression, Decision Tree, SVM, and linear regression analysis. The decision-Tree machine learning algorithm was found to provide the best outcomes in the study. [4]The early diagnosis of Parkinson's disease using ML has been described in this study along with other ML techniques using a twin-support vector machine (TSVM) based on the feature selection technique. [5]This paper introduces an automatic method that analyses the emotion changes during defined speech exercises using the XGBoost algorithm. [6]A Hand-tracking Sensor was used for recording the movements of the patients. The feature space produced by identical feature pairs for both hands showed noticeably higher outcomes for the RF method. [7]In this study, RStudio is used to analyse the patient's voice dataset using K-means clustering and decision tree-based machine learning methods. Python is used to examine the patient's spiral designs. From the extracted values, the patient is considered healthy if the voice and spiral drawing both indicate that they do not have Parkinson's disease. [8]Twelve machine-learning algorithms were compared to a unique deep-learning method that was developed to quickly determine if a person has PD or not using premotor data. Although deep learning outperforms machine learning models in terms of performance, it is difficult to declare that deep learning is superior to the others. [9]In this paper, different types of algorithms such as Support Vector Machine, Logistic regression, Extra Trees Random Forest, and Gradient Boosting were compared to predict the disease. [10]A detailed examination of the machine learning methods and data modalities used in the diagnosis and classification of Parkinson's disease. [11]In this paper, they have used a deep neural network algorithm using voice data to diagnose people with PD. [12]The enlarged feature spaces created by PCA increased the accuracy of all algorithms for Parkinson's disease prediction. Future research can evaluate other feature choices to boost accuracy. [13]In this study, a spoon prototype that has the ability to detect the hand motion of a person is used to predict the disease. [14]In this study, a boosted decision tree was used to detect Parkinson's patients, which was found to be an accurate model that makes use of the voice dataset. [15]In this paper, Boosted Logistic Regression produces superior results. [16]This method outperformed other cutting-edge methods and achieved high accuracy. This method demonstrates that Parkinson's disease can be detected in a non-invasive and cost-effective manner, which could have significant implications for early diagnosis and treatment. [17]A PCA and SVM-based automated PD diagnosis system is proposed in this paper. The system uses PCA to break down the features into their most important parts based on voice and finger tapping tests. The SVM classifier is trained using the reduced features as inputs. It demonstrates the potential to provide a non-invasive and objective biomarker for PD diagnosis by achieving high accuracy, sensitivity, and specificity in the detection of the disease. [18]The PD diagnosis system described in this paper makes use of fused data from accelerometers and gyroscopes. To diagnose Parkinson's disease, it uses machine learning algorithms like k-nearest neighbor, random forest and SVM to extract features from the sensor data. It achieves high precision, responsiveness, and particularity in PD conclusion and exhibits the potential for sensor-based approaches in PD diagnosis. [19]In this paper, a profound learning-based programed framework for PD utilizing mind X-ray pictures is proposed. Automated PD diagnosis based on brain MRI images is made possible by the proposed system, which achieves high levels of specificity, sensitivity, and accuracy. It gives a harmless and objective biomarker for PD finding, which can help with the early conclusion and treatment of PD. [20]When using gait data for automated PD detection, the proposed method uses the SVM classifier to achieve highest accuracy. The findings demonstrate the potential of PD diagnosis methods based on machine learning based on gait data.

2. MATERIALS AND METHODS

A subset of artificial intelligence known as machine learning enables machines to improve their performance on a task without being explicitly programmed by learning from data. All in all, rather than being guided precisely, machines are taken care of information and calculations that permit them to learn and go with choices in light of that information.
Reinforcement learning, unsupervised learning, and supervised learning are the three categories of machine learning algorithms. Machine learning is used in a wide range of applications, including natural language processing, picture recognition, predictive analytics, recommendation systems, and autonomous cars.

A. Dataset

1) Source and information

The dataset that is used in this study was downloaded from Kaggle, Each participant recorded seven voice recordings on average. The voice measurements in this study have a duration range of 1 to 35 seconds. The dataset comprises 147 recordings from individuals diagnosed with Parkinson's disease (PD) and 48 recordings from healthy subjects. A total of 23 features were extracted from the voice recordings, and these features are detailed in Table 1.

S.N.	Attributes	Description
1.	MDVP:Fo(Hz)	Average Vocal Fundamental Frequency
2.	MDVP:Fhi(Hz)	Maximum Vocal Fundamental Frequency
3.	MDVP:Flo(Hz)	Minimum Vocal Fundamental Frequency
4.	MDVP:Jitter(%)	Several measures of variation in fundamental frequency
5.	MDVP:Jitter(Abs)	
6.	MDVP:RAP	
7.	MDVP:PPQ	
8.	Jitter:DDP	
9.	MDVP:Shimmer	Several measures of variation in amplitude
10.	MDVP:Shimmer(dB)	
11.	Shimmer:APQ3	
12.	Shimmer:APQ5	
13.	MDVP:APQ	
14.	Shimmer:DDA	
15.	NHR	Measures of the ratio of noise to tonal components in voice
16.	HNR	
17.	Status	0: Healthy Subjects and 1: PD Patients
18.	RPDE	Two non-linear dynamical complexity measures
19.	D2	
20.	DFA	Signal Fractal scaling exponent
21.	Spread 1	Three nonlinear measures of fundamental frequency variation
22.	Spread 2	
23.	PPE	

Table 1 – Dataset Used From Kaggle**.**

B. Data Pre-processing

There are multiple steps involved in creating a model that could use voice recordings of the patients to diagnose them. First, the voice recording data needed to be pre-processed so that the machine learning algorithm could easily analyze it. To do as such, a dataset with handled voice accounts was used from Kaggle. The label "status" in the dataset indicated whether a patient is suffering from the disease or not. A status of 0 indicated that the patient was healthy, and a status of 1 indicated that a clinician had diagnosed the patient with PD. In the machine learning models that were tested, the label for two-class classification was the "status" column. Except for "name," all of the other columns were regarded as typical model features. Cleansing the missing data was the first step, and it was done as a precaution. But there was no missing data.

C. Correlation Heatmap

The pairwise correlations between various variables in a dataset are displayed graphically in a correlation heatmap, which is a depiction of the correlation matrix. The intensity of the color used to symbolize each variable indicates how strongly the variables are correlated with one another. It is possible to rapidly determine which variables have strong correlations and which ones do not use the heatmap. Every pair of variables has a correlation coefficient, which ranges from -1 to 1, with -1 signifying a perfect negative correlation, 0 signifying no association, and 1 signifying a perfect positive correlation. In order to find patterns and links between variables, correlation heatmaps are frequently used in data analysis, particularly in exploratory data research.

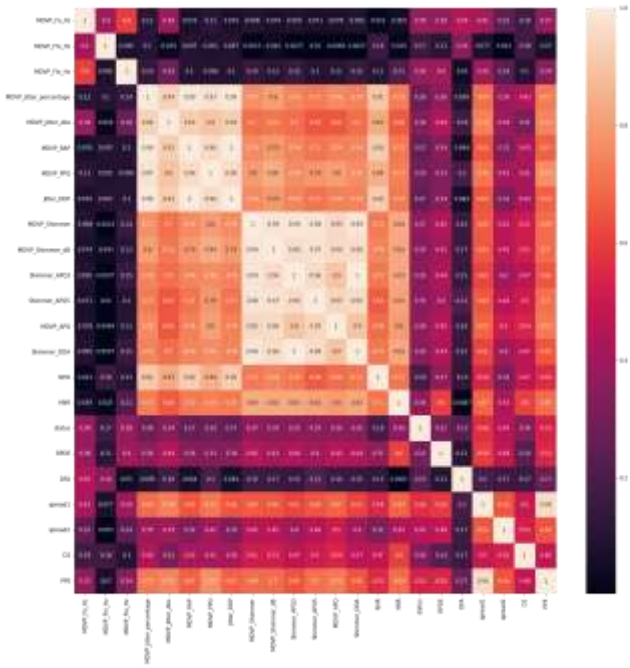

Fig 1: Correlation heatmap

Here n=195 and X values are Shimmer-DDA, MDVP--Jitter(%), MDVP- Fhi(Hz), MDVP-Fo(Hz), MDVP- Shimmer(dB), MDVP-Flo(Hz), MDVP- PPQ, MDVP- Jitter(Abs), MDVP- RAP, Jitter- DDP, MDVP- Shimmer, Shimmer- APQ3, Shimmer- APQ5, NHR, MDVP- APQ, HNR, DFA, RPDE, spread2, spread1, PPE, D2and the Y value is status. With the help of the correlation heatmap, we found that the feature 'MDVP- Fhi(Hz),' is negatively correlated with the status. So, we removed that feature, since it's not contributing to the target variable. Some other features like MDVP-Flo(Hz), MDVP- Fo(Hz), and HNR also contribute less to the target variable. So these features were also removed.

D.LightGBM algorithm

An open-source library known as LightGBM (Light Gradient Boosting Machine) offers an effective and efficient gradient-boosting algorithm implementation. It was created by Microsoft and made available to the public in 2016. It is one of the fastest boosting algorithms because it works in a way that is similar to that of other algorithms. However, it also has some advanced features. LightGBM makes decision trees that grow leafwise, which means that, depending on the gain, only one leaf splits when a condition is met. Leaf-wise trees occasionally overfit, particularly when working with smaller datasets. Overfitting can be avoided by limiting the depth of the tree. LightGBM utilizes a histogram-based technique in which information is bucketed into receptacles utilizing a histogram of the conveyance. Iteration, gain calculation, and data division are done with the bins rather than with each data point. Optimizing this method for a sparse dataset is also possible. Exclusive feature bundling, in which the algorithm combines exclusive features to reduce dimensionality and make it faster and more efficient, is another feature of LightGBM.

3. RESULTS

Confusion Matrix is generated by applying LightGBM algorithm. Classification results are shown in the form of a matrix which is commonly referred as Confusion Matrix.

Figure 2 shows Confusion Matrix results which help to calculate the evaluation metrics. Out of 39 (5+2+1+31) cases, real 7 (5+2) cases are healthy in the first row. Based on the prediction, 5 are healthy out of 7. Similarly, real 32 (1+31) cases are Parkinson's Disease Patients. Based on the prediction, 31 are healthy out of 32.

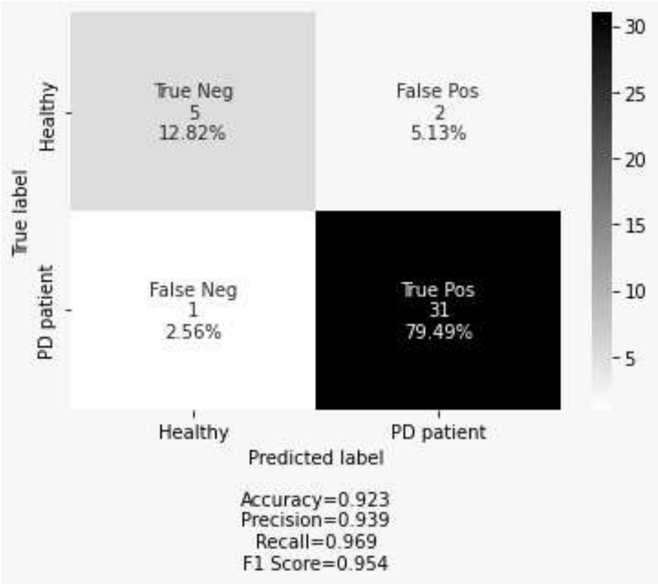

Figure 2: Confusion Matrix

4. CONCLUSION

By applying the Machine Learning techniques, Parkinson's disease can be detected based on the performance measure. Performance measure like an accuracy rate of 92% is achieved in differentiating individuals with Parkinson's disease from healthy controls using gait data analysis. Based on the results, the LightGBM algorithm may prove effective for detecting Parkinson's disease early, which may improve patient outcomes and assist in developing targeted interventions. According to the performance evaluation, the LightGBM classifier showed 0.923 accuracy, 0.939 precision, 0.969 recall, and a 0.954 F1-score.

5. REFERENCES

Sivakumar, Mahima, A. Hepzibah Christinal, and S. Jebasingh. "Parkinson's disease Diagnosis using a Combined Deep Learning Approach." *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*. IEEE, 2021.

P. Raundale, C. Thosar and S. Rane, "Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm," 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 2021, pp. 1-5, doi: 10.1109/INCET51464.2021.9456292

S. Tadse, M. Jain and P. Chandankhede, "Parkinson's Detection Using Machine Learning," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 1081-1085, doi: 10.1109/ICICCS51141.2021.9432340.

S. Thapa, S. Adhikari, A. Ghimire and A. Aditya, "Feature Selection Based Twin-Support Vector Machine for the Diagnosis of Parkinson's Disease," 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC), Kuching, Malaysia, 2020, pp. 1-6, doi: 10.1109/R10-HTC49770.2020.9356984.

J. Skibińska and R. Burget, "Parkinson's Disease Detection based on Changes of Emotions during Speech," 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic, 2020, pp. 124-130, doi: 10.1109/ICUMT51630.2020.9222446.

A. Moshkova, A. Samorodov, N. Voinova, A. Volkov, E. Ivanova and E. Fedotova, "Parkinson's Disease Detection by Using Machine Learning Algorithms and Hand Movement Signal from LeapMotion Sensor," 2020 26th Conference of Open Innovations Association (FRUCT), Yaroslavl, Russia, 2020, pp. 321-327, doi: 10.23919/FRUCT48808.2020.9087433.

Mathew, Mevin John, and Jomon Baiju. "Machine learning technique based parkinson's disease detection from spiral and voice inputs." Eur. J. Mol. Clin. Med 7 (2020): 2815-2819.

W. Wang, J. Lee, F. Harrou and Y. Sun, "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning," in IEEE Access, vol. 8, pp. 147635-147646, 2020, doi: 10.1109/ACCESS.2020.3016062.

E. Celik and S. I. Omurca, "Improving Parkinson's Disease Diagnosis with Machine Learning Methods," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-4, doi: 10.1109/EBBT.2019.8742057.

Ahmed, M. T., Mondal, M. N. I., Gupta, D., & Ali, M. S. (2022). A Review on Parkinson's Disease Detection Methods: Traditional Machine Learning Models vs. Deep Learning Models. European Journal of Information Technologies and Computer Science, 2(3), 1–6.

T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi, "Parkinson's Disease Diagnosis Using Machine Learning and Voice," 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 2018, pp. 1-7, doi: 10.1109/SPMB.2018.8615607.

E. Celik and S. I. Omurca, "Improving Parkinson's Disease Diagnosis with Machine Learning Methods," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-4, doi: 10.1109/EBBT.2019.8742057.

C. J. Baby, A. Mazumdar, H. Sood, Y. Gupta, A. Panda and R. Poonkuzhali, "Parkinson's Disease Assist Device Using Machine Learning and Internet of Things," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018, pp. 0922-0927, doi: 10.1109/ICCSP.2018.8523831.

A. Dinesh and J. He, "Using machine learning to diagnose Parkinson's disease from voice recordings," 2017 IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA, 2017, pp. 1-4, doi: 10.1109/URTC.2017.8284216.

K. N. R. Challa, V. S. Pagolu, G. Panda and B. Majhi, "An improved approach for prediction of Parkinson's disease using machine learning techniques," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, 2016, pp. 1446-1451, doi: 10.1109/SCOPES.2016.7955679.

S. Shetty and Y. S. Rao, "SVM based machine learning approach to identify Parkinson's disease using gait analysis," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7824836.

"Automated Parkinson's Disease Diagnosis using Principal Component Analysis and Support Vector Machine" by K. M. Hasan, M. A. Hossain, and M. T. Hasan in IEEE Journal of Biomedical and Health Informatics, 2019.

"Parkinson's Disease Diagnosis using Machine Learning Algorithms based on Fused Sensor Data" by M. S. Kamal and S. U. Ahmed in IEEE Access, 2020.

"Deep Learning-based Automatic Diagnosis of Parkinson's Disease using Brain MRI Images" by M. A. Hossain, M. A. Rahman, and S. Islam in IEEE Transactions on Medical Imaging, 2021.

"Parkinson's Disease Detection using Machine Learning Algorithms on Gait Data" by A. Sharma and S. S. Mohapatra in IEEE Sensors Journal, 2021.