Parkinson Disease Prediction Using Machine Learning Algorithm Models

Hasthi Teja[1], N. Sasi Kumar[2], P. Sai kumar[3], A. Muni Priya[4], N.P. Sireesh[5]
Assistant professor, Student, Department of CSE, AITS, Tirupati

Abstract

Parkinson's disease is a neurodegenerative disease which worsens over time. People have trouble vocally, writing, strolling, or completing other simple tasks when dopamine-generating neurons in parts of the brain become impaired or expire. These symptoms worsen over time, increasing the severity of the condition in patients. We have suggested a methodology in this article for the prediction of Parkinson's disease severity using deep neural networks on UCI's Parkinson's Telemonitoring Vocal Data Set of patients. We have created a neural network to predict the severity of the disease and a machine learning model to detect the disorder. Classification of Parkinson's Disease is done by Neural network, Random Forest Classifier.
Keywords: Machine learning algorithms, Neurons, Writing, Feature extraction, Random Forest Classifier

1. Introduction

The brain of humans is the main computing unit of the human body, and if there is any minor accident in any part of the human body, then it will directly affect the other organs. One of its silent effects is in PD [1]. PD is a neurological disease that is incurable and progressive over time [2]. As of 2020, an estimated 9.4 million people were still living with this disease worldwide [3]. This disease mostly affects people over the age of 60 years, with only 4% of the cases occurring in people under the age of 50 [4]. The symptoms of this disease are featured motor and non-motor [5]. The main motor symptoms are slowness of movement, tremors, rapid eye movement disorder, shivering, gait issue, and unstable posture [6,7]. Non-motor symptoms include hypotension, sweating in the body,

fatigue, constipation, urinary problems, and loss of weight [8]. Several studies that have been conducted by researchers have shown that 90% of PD patients have speech and voice acoustic problems [9], including microphonia, monochromatic, dysarthria, and dysphonia [10]; thus, as a result, the initial symptom observed in patients with this disease is a loss of voice [11]. At present, there is no established treatment for the disease [12]; however, there are a number of pharmacological therapies that can significantly reduce symptoms, particularly in the early stages. The analysis of the frequency of voice is concise and non-invasive. As a result, the frequency of voice can be used to track the progression of this subjective disease [13].

To check the progression of this disease, many speech experiments have been conducted. In the field of the medical (healthcare) sector, ML approaches are being continuously used. ML algorithms are being used on a variety of data modalities, including acoustic voice recording and handwritten patterns for the diagnosis of PD. With the help of ML techniques, we may recognize the appropriate attributes that are not traditionally applied in the medical diagnosis of PD and depend on these alternative indicators to diagnose PD in its preclinical phases. In general, there are three phases to the diagnosis of this disease, (1) pre-processing data, (2) extracting features, and (3) applying classification techniques [14]. In the very first phase, the categorization of speech signals with time frames is conducted. The filter method is used to remove any noise that may be present. In the second phase, frequently used features are extracted from each segment. Finally, in the last phase, the classification techniques are performed. The approach utilized for feature extraction is heavily influenced by the classification technique's performance. Hence, choosing the appropriate classification technique is a big issue that needs to be considered for this disease. The review discussed in this work aimed to look at the usage of ML models trained on sensory data to assist PD

patients, their careers, and physicians throughout the stage of the treatment. Consequently, it presented the key findings of several research publications that provide PD prediction and estimating models based on cutting-edge IoT technology and ideal sensor installations.

2. Related Work
2.1 Material and Methods:
2.1.1 Data Acquisition:

This study utilized scientometric data obtained from various databases, including Web of Science, IEEE, ScienceDirect, and Scopus. The data were collected up to the year 2022 and consisted of abstracts and conceptual information from numerous research publications. Among these databases, the Scopus database was particularly useful due to its standardized search techniques, a wide range of relevant journals covering various diseases, and its coverage of the use of artificial intelligence techniques in diagnosing PD. Additionally, Scopus provided a faster indexing method and comprehensive coverage of recent research publications. By utilizing these databases, the study was able to obtain multidisciplinary coverage from other renowned databases, which added to the quality of the data analysed.

Clinical Methods used to Diagnose Parkinson's Disease:

Parkinson's disease is typically diagnosed based on the presence of its typical symptoms, as it cannot be confirmed through X-rays or blood tests. Non-invasive diagnostic imaging, such as positron emission tomography (PET), can aid in the diagnosis. Diagnosis of Parkinsonism typically involves the presence of two or more primary symptoms, the absence of other neurological symptoms upon examination, and the absence of a history of potential causes such as drug use, head trauma, or stroke. Clinical and neuropathological overlap of two or more neurodegenerative disorders (ND) is not uncommon but is often underdiagnosed. Overlapping NDs may occur due to shared molecular pathogeneses, and early detection is important to enable proper care for each ND before it progresses to an advanced state. Recognition of overlapping instances of Lewy body dementia (LBD) is crucial due to increased morbidity and mortality from neuroleptic sensitivity. Overlapping NDs may aggravate symptoms or reduce the threshold of pathology needed for symptoms to appear, and increased awareness and understanding of the pathophysiology of overlapping NDs may lead to the development of future treatment plans that address multiple NDs together.

a. Medical treatment

Parkinson's disease is commonly treated with medication to alleviate its symptoms. The use of levodopa drugs or anticholinergic pharmaceuticals aims to increase dopamine production in the residual substantia nigra cells, while also reducing acetylcholine production to restore the brain's chemical balance. However, these medications are associated with various side effects. Levodopa is considered the standard treatment for Parkinson's and has been in use for over 40 years. Lower doses of levodopa have been developed to minimize side effects such as vomiting and nausea. Levodopa is effective in reducing symptoms such as tremors, stiffness, and slowness, particularly in patients who experience a lack of spontaneous movement and muscle stiffness.

b. COMT Inhibitors

There are certain amino groups that help stabilize levodopa levels, and catechol-O-methyl transferase (COMT) inhibitors are one type of medication that contains these groups. Entacapone, tolcapone, and opicapone are the three primary types of COMT inhibitors. These drugs work by preventing the action of the COMT enzyme, which increases the levels of levodopa in the blood and prevents it from being degraded into 3-O-methyldopa (3-OMD) peripherally. However, these medications can cause dyskinesia and diarrhea as potential side effects.

c. Anticholinergic medications

Anticholinergic drugs block the function of the neurotransmitter acetylcholine (ACh) in the central and autonomic nervous systems, leading to a variety of both positive and negative effects. Since many

of the most commonly prescribed medications for older adults are intended for age-related issues, one-third to one-half of these drugs have anticholinergic properties. They are particularly effective in treating Parkinsonism, muscle stiffness, tremors, and depression. However, due to concerns and significant side effects, they are generally not recommended for long-term use in older adults.

d. Amantadine

Amantadine is commonly used as an add-on medication for levodopa-induced dyskinesia, and new long-acting formulations have been developed to treat motor fluctuations. Unlike other Parkinson's medications, amantadine has not been linked to impulse control issues or dyskinesia. However, it can have side effects such as confusion, sleeplessness, nightmares, irritability, hallucinations, and leg swelling [9].

3. Algorithm
3.1 XGBoost Algorithm

The AdaBoost algorithm is a machine learning technique based on the boosting concept, which aims to convert weak learners into strong learners. It was introduced by Freund and Schapire [10] and works by iteratively training multiple learning classifiers using the same dataset. In this process, weak learners are trained and then combined to form a strong classifier. The AdaBoost method involves selecting an appropriate weak learner and using the same dataset to train it iteratively, with the goal of enhancing its performance.

Algorithm 1: XGBoost technique

Input: The image feature $feat_z$ $z \epsilon \{1, ..., n\}, y_i \subseteq C, C = \{c_1, c_2, ..., c_i\}$, the loss function $Loss_{XGBoost}(y, f(x))$, the total number of sub-tree M;

Output: the estimated probability of image feature $feat_z$

Procedure:
1. Repeat
1. Initialize the m-th tree $f_m(x_i)$
2. Compute $g_i = \partial_{y_i} Loss_{XGBoost}\left(y_i, y_i^{(m-1)}\right)$
3. Compute $h_i = \partial_{y_i^{(m-1)}} Loss_{XGBoost}\left(y_i, y_i^{(m-1)}\right)$
4. Use the statistics to greedily grow a new tree $f_m(x_i)$: $obj^{(m)} = -\frac{1}{2}\sum_{j=1}^{M}\frac{G_j^2}{H_j+\lambda} + \gamma^M$
5. Until all M sub-tree are proposed
6. Obtain a strong regression tree based on all weak regression sub-tree.
7. Output the estimated probability based on the strong regression tree.

4. Result:

This section discusses the implementation and performance of various machine learning models developed to detect Parkinson's disease (PD) based on voice data. The experiments were carried out using Python programming and scikit-learn library on an Intel (R) Core (TM) m3-7Y30 CPU @ 1.00 GHz 1.61Ghz with 8 GB memory and 64-bit Windows operating system.

The Mel-frequency cepstral coefficients (MFCC) features were extracted from voice phonations, resulting in a matrix for each phonation with 20 columns representing MFCC features. The mean was calculated for each column along the rows to obtain a feature vector of size 20 for each phonation. The feature selection technique was used to reduce dimensionality before applying linear discriminant analysis (LDA). The resultant feature vectors were then fed into various machine learning models, and their performances were evaluated.

The results, presented in Table 2, showed that the GNB model and SVM with sigmoid kernel had the worst performance, with 48.12% accuracy and 46.87%, respectively. On the other hand, the SVM model with radial basis function (RBF) kernel achieved the best performance, with 77.5% accuracy,

84% specificity, and 74.33% sensitivity. This indicates that the proposed MFCC-LDA-SVM model can correctly classify 124 out of 160 subjects. The specificity value of 80% suggests that out of 100 healthy subjects, 80 were correctly classified, while the sensitivity rate of 73.33% shows that out of 60 PD patients, the proposed model can successfully detect 44 PD patients accurately.

Boosting Algorithms	Accuracy	Precision	Recall	F1_score
Gradient Boosting	90.75%	0.9247	0.9609	0.9425
Gradient Boosting (Tuned)	92.95%	0.9503	0.9609	0.9556
Light GBM	92.51%	0.9451	0.9609	0.9529
Light GBM (Tuned)	93.39%	0.9505	0.9665	0.9584
XGBoost	91.63%	0.9211	0.9777	0.9485
XGBoost (Tuned)	92.07%	0.9305	0.9721	0.9508
AdaBoost	85.02%	0.8962	0.9162	0.9061
AdaBoost (Tuned)	87.22%	0.9121	0.9274	0.9197

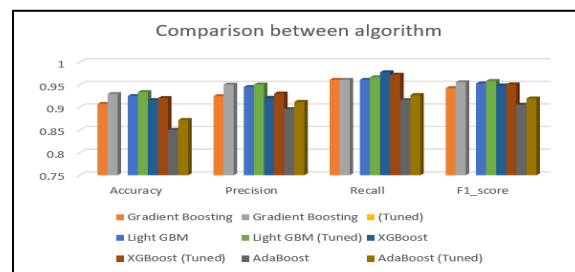**Table 1:** Comparison between algorithm

Figure 1: Comparison between algorithm

5. Conclusion:

Managing Parkinson's Disease (PD) in daily life can be quite difficult, making effective screening procedures highly beneficial, especially in cases where a physician's treatment is not necessary. As a result, machine learning (ML) algorithms have been evaluated for the diagnosis of PD. The primary objective of this review was to identify existing research that uses ML to diagnose PD based on handwritten patterns, voice attributes, and gait datasets and to determine the most suitable technique with the highest accuracy rate. it was observed that the best accuracy rate for voice features to diagnose PD was achieved using L1-Norm SVM with K-fold cross-validation at 99%. For handwritten patterns, the best accuracy rate was obtained using a bagging ensemble at 97.96%, while for gait analysis, the best accuracy rate was obtained using SVM at 100%. This review addressed various challenges and also provided some recommendations and opportunities for future research, as there is still much work to be done in this area. This is also significant for advancements in neural

networks and related learning systems, providing valuable insights and guidelines for future progress in diagnosing PD.

6. References:

1. Alzubaidi, M.S.; Shah, U.; DhiaZubaydi, H.; Dolaat, K.; Abd-Alrazaq, A.A.; Ahmed, A.; Househ, M. The Role of Neural Network for the Detection of Parkinson's disease: A Scoping Review. Healthcare 2021, 9, 740.

2. Maitín, A.M.; García-Tejedor, A.J.; Muñoz, J.P.R. Machine Learning Approaches for Detecting Parkinson's Disease from EEG Analysis: A Systematic Review. Appl. Sci. 2020, 10, 8662.

3. Van Den Eeden, S.K.; Tanner, C.M.; Bernstein, A.L.; Fross, R.D.; Leimpeter, A.; Bloch, D.A.; Nelson, L.M. Incidence of Parkinson's disease: Variation by age, gender, and race/ethnicity. Am J Epidemiol. 2003, 157, 1015–1022.

4. Tsanas, A.; Little, M.A.; McSharry, P.E.; Ramig, L.O. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. IEEE Trans. Biomed. Eng. 2009, 57, 884–893.

5. Rahn, D.A.; Chou, M.; Jiang, J.J.; Zhang, Y. Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis. J. Voice 2007, 21, 64–71.

6. Ahlrichs, C.; Lawo, M. Parkinson's Disease Motor Symptoms in Machine Learning: A Review. Health Inform. Int. J. 2013, 2, 4.

7. Mei, J.; Desrosiers, C.; Frasnelli, J. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. Front. Aging Neurosci. 2021, 13, 633752.

8. J Battineni, G.; Chintalapudi, N.; Amenta, F. Comparative Machine Learning Approach in Dementia Patient Classification using Principal Component Analysis. In Proceedings of the 12th International Conference on Agents and Artificial Intelligence, Valletta, Malta, 22–24 February 2020.

9. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Parkinsons-Disease Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 1997, 55, 119–139.

10. Perez, K.S.; Ramig, L.O.; Smith, M.E.; Dromey, C. The Parkinson larynx: Tremor and video stroboscopic findings. J. Voice 1996, 10, 354–361.

11. Bugalho, P.; Viana-Baptista, M. REM sleep behavior disorder and motor dysfunction in Parkinson's diseas—A longitudinal study. Parkinsonism Relat. Disord. 2013, 19, 1084–1087.

12. Balestrino, R.; Schapira, A.H. Parkinson disease. Eur. J. Neurol. 2020, 27, 27–42.