EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Prognoza: Parkinson's Disease Prediction Using Classification Algorithms

Mithun Shivakoti[1], Sai Charan Medaramatla[1], Deepthi Godavarthi[1,*], Narsaiah Shivakoti[2]

[1]School of Computer Science & Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India
[2]Department of Zoology, GDC Ibrahimpatnam, Rangareddy,Telangana, India

Abstract

Parkinson's Disease (PD) is a persistent neurological condition that has a global impact on a significant number of individuals. The timely detection of PD is imperative for the efficacious treatment and control of the condition. Machine learning (ML) methods have demonstrated significant potential in forecasting Parkinson's disease (PD) based on diverse data sources in recent times. The present research paper outlines a study that employs machine learning [ML]techniques to predict Parkinson's disease. A dataset comprising clinical and demographic characteristics of both patients diagnosed with PD and healthy individuals was taken from Kaggle. The aforementioned dataset was utilized to train and assess multiple machine learning models. The experimental findings indicate that the CatBoost model exhibited superior performance compared to the other models, achieving an accuracy rate of 95.1% and a root mean squared error of of 0.34.In summary, our research showcases the capabilities of machine learning methodologies in forecasting Parkinson's disease and offers valuable insights into the crucial predictors for PD prognosis. The results of our study could potentially contribute to the advancement of diagnostic methods for the timely identification of PD, with increased precision and efficacy.

Keywords: Parkinson's disease, Classification, Machine Learning, CATBoost, Random Oversampling

Received on 18 June 2023, accepted on 12 September 2023, published on 21 September 2023

doi: 10.4108/eetpht.9.3933

Corresponding author. Email: deepthi.g@vitap.ac.in

1. Introduction

Parkinson's disease is a neurodegenerative disorder that causes movement impairment. The timely identification and diagnosis of PD can significantly improve the quality of life of individuals affected by the condition.ML techniques have become increasingly crucial in the prediction and diagnosis of Parkinson's disease [1].
Presently, a variety of neurodegenerative ailments have been recognized, such as Alzheimer's, Parkinson's, rheumatoid arthritis, memory loss with Lewy bodies, corticobasal degeneration, prion problems etc [2]. PD is a complex neurodegenerative condition that is associated with ageing [3,4]. The fundamental diagnostic indicators of this condition include bradykinesia, stiffness, tremor, and postural instability. These indicators are well-established in medical literature. The ailment is commonly categorized into two subtypes, namely early-onset and late-onset disease, with the former progressing slowly and the latter advancing rapidly, often referred to as "benign" and "malignant," respectively. This classification has been documented in various sources [5,6].
ML models have significant clinical implications for PD prediction, including the identification and diagnosis of the condition at an early stage, monitoring the progression of the disease and the response to therapy, and predicting disease outcomes. These models can also facilitate the development of personalized treatment plans for patients. PD prediction using ML algorithms may have significant clinical implications [29]. These models, for instance, can be used to create screening tools that pinpoint people at high risk for Parkinson's, allowing for earlier detection and treatment. They can also be used to track the development

of a disease and the effectiveness of treatment, enabling patients to get individualized care.

Although Parkinson's disease is primarily a movement illness, other deficits, such as psychological issues like depression and dementia, also arise. Autonomic disturbances and discomfort may later develop, and as the condition worsens, the affected person experiences considerable disability and handicap as well as a reduced quality of life. Potentially impacted parties include family and carers. According to research, almost 90% of PD patients exhibit vocal dysfunction, including dysphonia, which is a problem with producing vocal sounds normally[7,8].

In recent years, there have been numerous studies exploring different modalities that utilize machine learning. The aforementioned modalities comprise CSF biomarkers, imaging, RNA, movement-related metrics, and wearable sensor data. Several methods have been successful in classifying data; however, our aim was to develop models that employ inexpensive and readily available data sources that can be established remotely or through pre-existing biobank data, without requiring additional patient visits or expensive techniques [9,10,11,12,13,14,15]. The aim of our study is to construct an accurate prognostic model for timely disease detection, with the objective of detecting, evaluating, and controlling the disease before any apparent clinical manifestations. This particular approach is considered to be the most efficacious since it permits prompt intervention during a phase in which the progression of the ailment can be most feasibly managed [16]. The application of ML techniques in predicting PD has the potential to enhance early detection and diagnosis, leading to improved patient outcomes and a higher quality of life [28,30].

2. Literature Review

A study was conducted by Das [17] to investigate different categorization techniques with the objective of attaining accurate diagnosis of PD. The assessment of the classifiers' efficacy was carried out using a variety of methodologies. Based on the results of the application score, it can be inferred that the Neural Networks (NNs) classifier exhibited the most superior performance. In their study, Bhattacharya and Bhatia [18] utilized the Weka data mining software to preprocess the dataset before applying Support Vector Machine (SVM) to differentiate between individuals with PD and those who are in good health.The dataset obtained from the experiment was subjected to analysis using LIBSVM, with the aim of identifying the kernel values that demonstrated the highest level of accuracy. The evaluation of model accuracy was conducted by utilizing the Receiver Operating Characteristic (ROC) curve variation.

Chen et al. [19] proposed a diagnostic methodology for Parkinson's disease that employs the Fuzzy K-Nearest Neighbor (FKNN) algorithm. A comparative analysis was performed to evaluate the results obtained from the FKNN-oriented approach and SVM-based methodologies. Principal Component Analysis (PCA) was employed to enhance the precision of Parkinson's Disease (PD) diagnosis. Ozcift et al. [20] utilized a classification methodology that was based on Support Vector Machines (SVM) in order to forecast the advancement of Parkinson's disease. Hariharan et al. [21] introduced a hybrid intelligent system that integrates feature reduction, clustering via the Gaussian mixture model, and classification techniques. The study by Froelich et al. [22] aimed to explore the feasibility of identifying PD through the analysis of an individual's vocal characteristics. The voice samples were classified as either healthy or unhealthy using decision trees. The implementation of a threshold-based approach was facilitated by the utilization of pre-annotated voice samples, which enabled the conclusive diagnosis of the patient. The determination of the threshold value is a crucial aspect in establishing the minimum number of individual voice samples required for an accurate diagnosis of a patient's condition. The specimens mentioned above provide substantiation for the presence of the disease.

The objective of Eskidere et al.'s [23] investigation was to assess the effectiveness of different regression techniques, including SVM, Least Square SVM (LS-SVM), Multilayer Perceptron Neural Network (MLPNN), and General Regression Neural Network (GRNN), in the remote monitoring of the progression of PD. The results of the analysis suggest that LS-SVM outperformed the most recent regression methods proposed in academic literature. Guo et al [24] conducted a study in the Central South region of Mainland China, analyzing a cohort of 1061 individuals diagnosed with PD and 1066 healthy individuals. The study involved the analysis of 16 Single-Nucleotide Polymorphisms (SNPs) located within 8 genes and/or loci. The study revealed that several genetic variations, namely Rep1, rs356165, and rs11931074 of the SNCA gene, G2385R variation of the LRRK2 gene, rs4698412 variation of the BST1 gene, rs1564282 variation of the PARK17 gene, and L444P variation of the GBA gene, had a significant independent and combined effect on the initiation of Parkinson's disease. According to the study, the influence of single nucleotide polymorphisms (SNPs) on PD is particularly significant in the four genes that were previously identified.

The detection of PD was studied by Polat et al. [25] through the utilization of Fuzzy C-Means (FCM) Clustering-based Feature Weighting (FCMFW) methodology. The author employed KNN to classify the experimental dataset, utilizing different k values. A proposal was put forth by Aström and Koker [26] for a prediction system that employs parallel neural networks. The neural network selection process entailed evaluating its output through a rule-based methodology, culminating in the prognostication of Parkinson's disease.

3. Proposed System

3.1. Dataset Description

The present dataset was obtained from Kaggle [27] in the format of a Comma Separated Value file, containing information on the speech characteristics of individuals both with and without PD. The dataset comprises 754 speech features, including 20 baseline features, 3 intensity parameters, 4 format frequencies, 4 bandwidth parameters, 32 vocal fold features, 83 Mel-frequency cepstral coefficients, 181 wavelet features, and 431 Tunable Q-factor Wavelet Transform features. Additionally, the dataset includes a class attribute consisting of 756 distinct instances.

3. 2. Methodology

3.2.1. Logistic Regression

Logistic regression (LR) is a frequently employed ML techniques for processing categorizing data into one of two distinct classes. LR is a generalized linear model commonly employed to establish a statistical association between the input features and the output variable, which assumes binary values of 0 or 1.
The LR methodology involves the process of fitting a logistic function, commonly referred to as a sigmoid function, to the input features. The sigmoid function receives a linear combination of input features and produces an output value ranging from 0 to 1, denoting the likelihood of the output variable being 1. Subsequently, the logistic regression algorithm employs the aforementioned probability to generate a forecast for the output variable, commonly utilizing a threshold value of 0.5.

3.2.2. Decision Tree

Decision Tree operates through a recursive process of dividing the input data into subsets, utilizing the values of the input features as the basis for partitioning. The primary objective is to optimize the homogeneity or purity of the resulting subsets in relation to the target variable.
The decision Tree is a graphical representation of input data that takes the form of a tree-like structure. Each node in the tree corresponds to a feature, and each edge represents a decision that is made based on the value of that feature. The construction of the tree is achieved through a recursive process that involves partitioning the input data into subsets, considering the value of the feature at the current node. This procedure is repeated until the subsets are as homogeneous as possible with regard to the desired variable.

3.2.3. Random Forest

The Random Forest algorithm is an ensemble learning technique that utilizes the combination of multiple decision trees to enhance their performance and mitigate the issue of overfitting that may arise in individual trees.
The random forest algorithm involves generating a set of decision trees, where a distinct subset of the input data and input features is randomly chosen for each tree. The technique of training individual trees on unique subsets of data and features is employed to reduce the association among the trees and increase the range of the forest.

3.2.4. Support Vector Machine (SVM)

SVM algorithm is a machine learning methodology that entails the depiction of input data as a collection of points in a space with multiple dimensions. Each discrete point contained in the provided space corresponds to a unique feature vector. The algorithm aims to identify a hyperplane that optimizes the margin between two distinct classes. The concept of "margin" refers to the measurable spatial separation between the hyperplane and the closest data points associated with each respective class. The process of identifying the optimal hyperplane involves maximizing the margin.

3.2.5. Gaussian Naive Bayes

The Gaussian Naive Bayes classifier is an ML algorithm that operates on probabilities and is commonly employed for classification purposes. The Gaussian Naive Bayes algorithm represents the input data as a collection of feature vectors, with the assumption that each feature is independent of the others. The aforementioned statement implies that the likelihood of a specific class in relation to the input data can be computed by multiplying the probabilities of each feature with respect to the class. The aforementioned process is executed through the utilization of Bayes' theorem.

3.2.6. Light Gradient Boosting Machine (LGBM)

LGBM is a machine learning utilizes gradient boosting decision trees. The tool has been designed with a focus on achieving high efficiency and scalability, rendering it a valuable asset for machine learning tasks that operate on a large scale. The methodology employed involves the utilization of a histogram-based algorithm in the creation of decision trees, resulting in a reduction of the dataset's memory requirements and an increase in processing speed. The Light GBM algorithm employs a growth strategy that prioritizes the development of deeper trees with fewer leaves, resulting in enhanced precision and expedited training durations. Moreover, Light GBM possesses the capability to handle categorical features, implement early stopping, and incorporate other functionalities that render it a potent and effective mechanism for gradient boosting decision trees.

3.2.7. K-Nearest Neighbors (KNN)

The KNN algorithm is a straightforward yet efficient technique in the field of ML that can be utilized for both classification and regression purposes. Regarding classification, the KNN algorithm operates by identifying the K nearest data points in the training set to a specific query point, and subsequently designating the query point to the category that is most frequently observed among its K closest neighbors. In regression applications, the K-Nearest neighbors (KNN) algorithm operates by identifying the K data points that are closest in proximity to a given query point, and subsequently assigning the query point the mean value of the KNN. The KNN is classified as non-parametric due to its characteristic of not making any presumptions regarding the underlying data distribution. The algorithm's flexibility enables its application to a diverse array of problems. The K-Nearest neighbors (KNN) algorithm is characterized by its ease of implementation and versatility in handling datasets of varying sizes. In general, KNN is a straightforward and efficient ML technique that can be utilized for a diverse scope of potential uses. This approach proves to be especially advantageous in scenarios where the dataset is limited in size and the distribution of the data is not well understood. Nonetheless, the efficacy of the system may be compromised when dealing with voluminous and multi-dimensional datasets.

3.2.8. CATBoost

CAT Boost, also known as Categorical Boosting, is a machine learning framework that is open-source and designed for gradient boosting on decision trees. The design of this method is optimized for enhanced performance in managing categorical features, which are characterized by a finite set of distinct values, such as occupation or gender. CAT Boost is distinguished by its capacity to process categorical features directly, thus eliminating the necessity for one-hot encoding or other preprocessing procedures. This is achieved through the utilization of a unique gradient computation algorithm that relies on the probability distribution of categorical variables, as opposed to their respective values. CAT Boost is able to enhance its ability to capture the associations between categorical variables and the target variable.CAT Boost is notable for its ability to effectively manage missing data. The methodology employed involves utilizing gradient-based techniques to perform imputation of absent values, thereby conferring superior efficacy in handling missing data when compared to alternative gradient boosting frameworks.

3.2.9. XGBoost

XG Boost, also known as Extreme Gradient Boosting, is a widely used open-source machine learning framework designed for gradient boosting on decision trees. The tool has been designed with a focus on achieving high scalability, efficiency, and accuracy, rendering it a valuable asset for a diverse array of machine learning applications. The efficient handling of large datasets is considered a significant attribute of XG Boost. The aforementioned process is achieved through the utilization of a distributed computing framework, which enables the parallelization of the training process across numerous machines. XG Boost has the capability to handle vast datasets in a more expeditious and effective manner compared to other frameworks for gradient boosting. The XG Boost algorithm employs a unique approach to selecting split points in decision trees, known as gradient-based optimization. The algorithm operates by calculating the gradient of the loss function in relation to every split point. It subsequently chooses the split point that maximizes the reduction in the loss function.

3.2.10. Min-Max scaling

The Min-Max Scaler is a data normalization technique used to scale a dataset's features to a specified range. It is mostly employed in ML algorithms to enhance model performance and precision. The formula for min-max scaling is expressed in equation 1.

$$D_{scaled} = (D - D\_min)/(D\_\max - D\_min) \quad - 1$$

Where,

D represents the value of the feature,
D_min denotes the minimum value of the feature,
D_max denotes the maximum value of the feature and
D_{scaled} represents the scaled value of the feature.

The formula above rescales the data so that it ranges between 0 and 1. Using the formula in equation 2, this normalization method can also be modified to rescale the data to any arbitrary range.

$$D_{scaled} = (D - D_{min}) * \frac{\min\ \_range - \max\ \_range}{D_{max} - D_{min}} + \min\ \_range \quad - 2$$

Where,

min_range and max_range represent the minimum and maximum range values, respectively.

The Min-Max Scaler is a linear transformation that preserves the geometry of the data distribution while ensuring that all values are contained within a given range. It can be applied to both univariate and multivariate data, making it a versatile technique for normalizing data.

3.2.11. Random Oversampling

The Random Oversampling (ROSE) method is a data augmentation approach utilized in the ML to mitigate the challenge of imbalanced datasets. The technique entails

generating supplementary instances of the underrepresented class through stochastic duplication, thereby achieving parity between the minority and majority classes. The implementation of this methodology serves to mitigate partiality towards the dominant category and enhance the overall efficacy of the algorithm.
The Random Over-Sampling Examples (ROSE) technique operates by employing a random selection process to choose samples from the minority class and subsequently generating additional copies of them until the quantity of minority class samples matches that of the majority class. The aforementioned process is executed while retaining the initial minority class samples, resulting in a final dataset with a greater number of samples than the original dataset.

3.2.12. Root Mean Squared Error

RMSE is a commonly employed measure of performance in regression analysis, which assesses the degree of dissimilarity between predicted and observed values. The metric's mathematical formulation entails the calculation of the square root of the mean of the squared differences between the anticipated and actual values. The Root Mean Square Error (RMSE) can be mathematically represented as the square root of the ratio of the sum of the squared differences between the predicted and actual values to the total number of observations. Symbolically, it can be represented as shown in equation 3.

$$RMSE = sqrt(\frac{1}{N} * sum(y_{pred} - y_{actual})^2) \quad - 3$$

The formula pertains to a dataset with N samples,
Where,

y_{pred} represents the predicted value of a specific sample and y_{actual} represents the actual value of the same sample.

3.2.13. Mean Squared Error

The Mean Squared Error (MSE)(as shown in equation 4) is a prevalent metric utilized to evaluate the efficacy of a predictive model. Computation involves determining the mean of the squared deviations between the predicted and actual values of the dependent variable.

$$MSE = \frac{1}{n}(y_i - \widehat{y_i})^2 \quad - 4$$

Where,

n is the total number of observations in the dataset
y_i is the actual value of the target variable for the i-th observation.
$\hat{y}_i$ is the predicted value of the target variable for the i-th observation.

The Mean Squared Error (MSE) is a non-negative metric that serves as an indicator of model performance. A lower MSE value suggests superior model performance, as it implies a closer proximity between the predicted and actual values.

3.2.14. Mean Absolute Error

The Mean Absolute Error (MAE) is a commonly employed evaluation measure in regression assignments that measures the difference between the anticipated and actual values. The computation entails the determination of the mean of the absolute disparities between the anticipated and factual values. The mathematical expression for MAE is given by the formula shown in equation 5.

$$MAE = \frac{1}{N} * sum(abs(y_{pred} - y_{actual})) \quad - 5$$

The equation pertains to a scenario where N observations are being considered. The variable y_{pred} represents the predicted values, while y_{actual} denotes the actual values. The Mean Absolute Error (MAE) is computed by summing the absolute differences between the predicted and actual values, and subsequently dividing the sum by the total number of observations.

The dataset was initially imported and subsequently partitioned into training and testing datasets, with a ratio of 0.8:0.2 (Training dataset – 80% and Testing dataset – 20%). Given absence of null values or extraneous attributes in the dataset, minimal pre-processing was required. The collected data undergoes both qualitative and quantitative evaluations. Subsequently, as there was class imbalance, (as depicted in Figure 1) class balancing technique as that of Random Oversampling has been used. Count value of both the class variables after class balancing technique has been clearly depicted in figure 2.

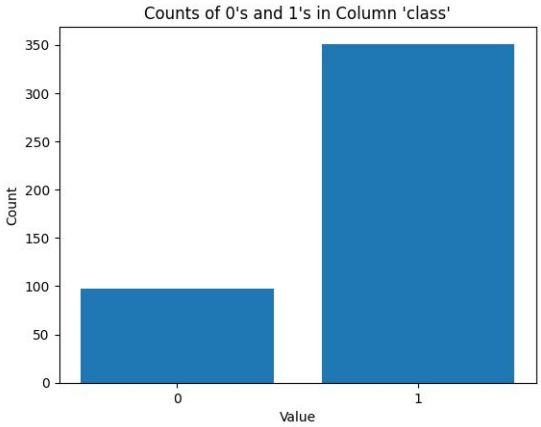

Fig 1: Count of each of the class variable before class balancing

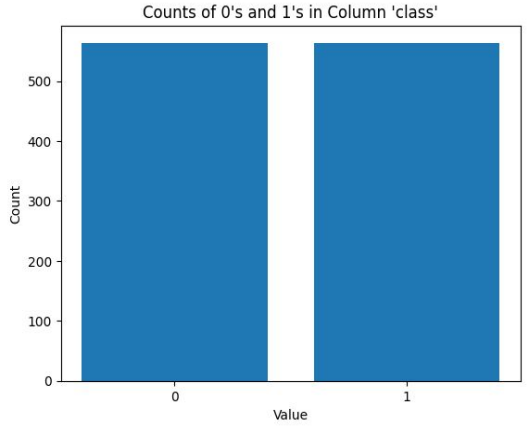

Fig 2: Count of each of the class variable after class balancing

Finally, a variety of classification models, such as the Logistic Classifier, Kernel SVM, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes Classifier, KNN Classifier, CAT Boost Classifier, XG Boost Classifier, and Light GBM Classifier, were employed to generate the predications. Additionally, the outcomes were graphically represented.

4. Results and Discussions

CATBoost has been found out to be performing the best for the current dataset both in terms of higher accuracy and least error. It has been found that CATBoost has achieved accuracy of 95%, RMSE of 0.3 and MSE and MAE of 0.3 on the dataset that has been generated by Random Oversampling technique. Followed by CATBoost, LightGBM and Random Forest performed good with 94.6% and 94.2% accuracies and root mean squared error of as less as 0.3. Table1 shows performance of nine different classification models across all four performance metrics.

Table 1: Performance of various classification models across various metrics

S.No	Model Name	Accuracy	RMSE	MSE	MAE
1	Logistic Regression	64.6	0.52	0.27	0.27
2	Decision Tree	90.7	0.47	0.23	0.19
3	Random Forest	94.2	0.35	0.15	0.12
4	SVM	64.1	0.51	0.26	0.26
5	Gaussian Naïve Bayes	61.0	0.58	0.34	0.34
6	XG Boost	93.3	0.37	0.14	0.14
7	CAT Boost	95.1	0.34	0.12	0.12
8	KNN	73.0	0.57	0.33	0.33
9	LGBM	94.6	0.31	0.10	0.10

Figure 3,4,5,6 depicts various performance metrics (Accuracy, RMSE, MSE, and MAE) of the nine different machine learning models.

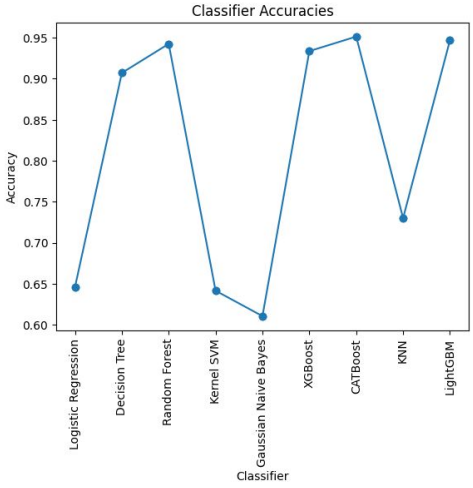

Fig 3: Accuracy of various machine learning models

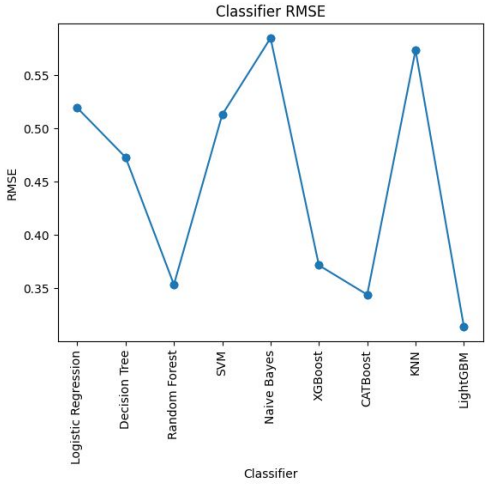

Fig 4: Root Mean Squared Error of various machine learning models

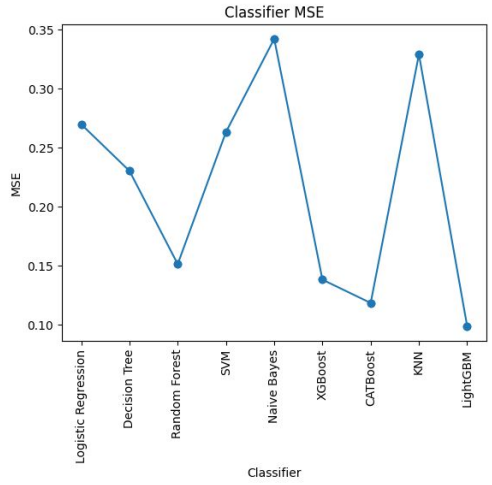

Fig 5: Mean Squared Error of various machine learning models

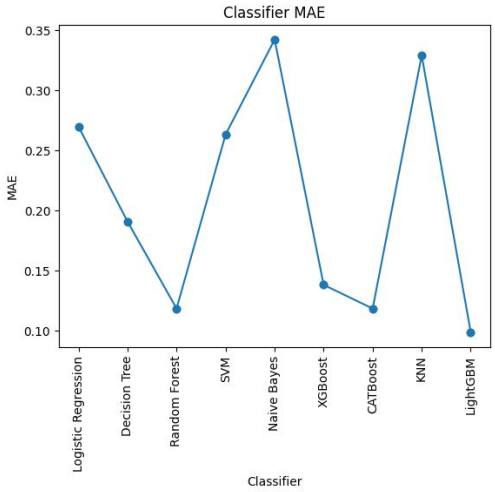

Fig 6: Mean Absolute Error of various machine learning models

5. Conclusion and Future Scope

Millions of people throughout the world are afflicted by the complex neurological ailment known as Parkinson's disease. The use of ML methods to forecast Parkinson's disease and its progression has gained more and more attention. The use of several ML algorithms to forecast Parkinson's disease has been investigated in a number of research. These experiments have shown encouraging findings, suggesting that machine learning algorithms are capable of making precise predictions about Parkinson's disease. Accuracy of as high as 95.1% has been achieved and RMSE of as less as 0.34 has been achieved. Limitations include dataset being as small as 750 rows which can be increased in future for better predictions.

References

[1] G. Nagasubramanian, M. Sankayya, F. Al-Turjman and G. Tsaramirsis, "Parkinson Data Analysis and Prediction System Using Multi-Variant Stacked Auto Encoder," in IEEE Access, vol. 8, pp. 127004-127013, 2020, doi: 10.1109/ACCESS.2020.3007140.

[2] L. Ramig, R. Sherer, I. Titze and S. Ringel, "Acoustic Analysis of Voices of Patients with Neurologic Disease: Rationale and Preliminary Data," The Annals of Otology, Rhinology, and laryngology, No. 97, pp. 164-172, 1988.

[3] Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J. Neurol. Neurosurg. Psychiatry 55, 181–184 (1992).

[4] Postuma, R. B. et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov. Disord. 30, 1591–1601 (2015).

[5] Jankovic, J. et al. Variable expression of Parkinson's disease: a base-line analysis of the DAT ATOP cohort. Neurology 40, 1529–1529 (1990).

[6] Zetusky, W. J., Jankovic, J. & Pirozzolo, F. J. The heterogeneity of Parkinson's disease: clinical and prognostic implications. Neurology 35, 522–526 (1985)

[7] Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L. & Gates, S. Speech impairment in a large sample of patients with parkinson's disease. Behavioural neurology 11, 131–137 (1999).

[8] Chen, H.-L. et al. An efficient diagnosis system for detection of parkinson's disease using fuzzy k-nearest neighbor approach. Expert Systems with Applications 40, 263–271 (2013).

[9] Prashanth, R., Dutta Roy, S., Mandal, P. K. & Ghosh, S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. Int. J. Med. Inform. 90, 13–21 (2016).

[10] Lee, D. A., Lee, H.-J., Kim, H. C. & Park, K. M. Application of machine learning analysis based on diffusion tensor imaging to identify REM sleep behavior disorder. Sleep Breath. https://doi.org/10.1007/s11325-021-02434-9 (2021).

[11] Mei, J. et al. Identification of REM sleep behavior disorder by structural magnetic resonance imaging and machine learning. Preprint at bioRxiv https://doi.org/10.1101/2021.09.18.21263779 (2021).

[12] Chen-Plotkin, A. S. Parkinson disease: blood transcriptomics for Parkinson disease? Nat. Rev. Neurol. 14, 5–6 (2018).

[13] Uehara, Y. et al. Non-invasive diagnostic tool for Parkinson's disease by sebum RNA profile with machine learning. Sci. Rep. 11, 18550 (2021).

[14] Noyce, A. J. et al. PREDICT-PD: identifying risk of Parkinson's disease in the community: methods and baseline results. J. Neurol. Neurosurg. Psychiatry 85, 31–37 (2014).

[15] Palmerini, L. et al. Identification of characteristic motor patterns preceding freezing of gait in Parkinson's disease using wearable sensors. Front. Neurol. 8, 394 (2017).

[16] Paulsen, J. S. et al. A review of quality of life after predictive testing for and earlier identification of neurodegenerative diseases. Prog. Neurobiol. 110, 2–28 (2013).

[17] Das, R. A comparison of multiple classification methods for diagnosis of parkinson disease. Expert Systems with Applications 37, 1568–1572 (2010).

[18] Bhattacharya, I. & Bhatia, M. Svm classification to distinguish parkinson disease patients. In Proceedings of the

1st Amrita ACM-W Celebration on Women in Computing in India, 14 (ACM, 2010).
[19] Chen, H.-L. et al. An efficient diagnosis system for detection of parkinson's disease using fuzzy k-nearest neighbor approach. Expert Systems with Applications 40, 263–271 (2013).
[20] Ozcift, A. Svm feature selection-based rotation forest ensemble classifiers to improve computer-aided diagnosis of parkinson disease. Journal of medical systems 36, 2141–2147 (2012).
[21] Hariharan, M., Polat, K. & Sindhu, R. A new hybrid intelligent system for accurate detection of parkinson's disease. Computer methods and programs in biomedicine 113, 904–913 (2014).
[22] Froelich, W., Wrobel, K. & Porwik, P. Diagnosis of parkinson's disease using speech samples and threshold-based classification. Journal of Medical Imaging and Health Informatics 5, 1358–1363 (2015).
[23] Eskidere, Ö., Ertas, F. & Hanilçi, C. A comparison of regression methods for remote tracking of parkinson's disease progression. Expert Systems with Applications 39, 5523–5528 (2012).
[24] Guo, J.-F. et al. Polygenic determinants of parkinson's disease in a chinese population. Neurobiology of aging 36, 1765–e1 (2015).
[25] Polat, K. Classification of parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering. International Journal of Systems Science 43, 597–609 (2012).
[26] Åström, F. & Koker, R. A parallel neural network approach to prediction of parkinson's disease. Expert systems with applications 38, 12470–12474 (2011).
[27] DIPAYAN BISWAS, (2019). Parkinson's Disease (PD) classification, Version 1, Retrieved March 15, 2023, from https://www.kaggle.com/datasets/dipayanbiswas/parkinsons-disease-speech-signal-features.
[28] Shivakoti, M., Jeeveth, K., Pradhan, N.R., Yesu Babu, M. (2023). Apple Stock Price Prediction Using Regression Techniques. In: Balas, V.E., Semwal, V.B., Khandare, A. (eds) Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_5
[29] CMBA-SVM: a clinical approach for Parkinson disease diagnosis, Bibhuprasad Sahu & Sachi Nandan Mohanty, International Journal of Information and Technology, 13(3), 647-655, (2021), doI: 10.1007/s41870-020-00569-8, ISSN: 2511-2104
[30] Shivakoti Mithun, Srinivasa Reddy K, and Adinarayana Reddy. "An Efficient Regression Method To Predict Soil pH Using RGB Values." International Research Journal on Advanced Science Hub 05 .05S May (2023): 35–42. http://dx.doi.org/10.47392/irjash.2023.S005