Salud, Ciencia y Tecnología. 2025; 5:1523
doi: 10.56294/saludcyt20251523

ORIGINAL

Enhancing Parkinson's Disease Detection using AI Techniques

Mejora de la detección de la enfermedad de Parkinson mediante técnicas de IA

Varshitha D N[1], Ranjan Kumar H S[2], Dharini K R[3], Kavyashree N[3], Shreya R[4], Soundarya Govinda Rao[5]

[1]Department of CSE(AI&ML), Vidyavardhaka College of Engineering. Mysuru, Karnataka, India.
[2]Department of Artificial Intelligence and Data Science, Shri Madhwa Vadiraja Institute of Technology and Management. Udupi. Bantakal, India.
[3]Department of CSE, BGS Institute of Technology, Adichunchanagiri Univeristy. India.
[4]Tekzen Systems. Mysore, India.
[5]Accenture Bengaluru. India.

Cite as: D N V, H S RK, K R D, N K, R S, Rao SG. Enhancing Parkinson's Disease Detection using AI Techniques. Salud, Ciencia y Tecnología. 2025; 5:1523. https://doi.org/10.56294/saludcyt20251523

Submitted: 18-08-2024 **Revised:** 03-12-2024 **Accepted:** 27-02-2025 **Published:** 28-02-2025

Editor: Prof. Dr. William Castillo-González

Corresponding Author: Varshitha D N

ABSTRACT

One of the severe illnesses that causes uncontrollable and unexpected outcomes is Parkinson's disease (PD). People over 50 years of age are typically the ones who contract this illness. The patients' symptoms progressively become worse leading to a variety of abnormalities such as body part rigidity and abnormalities in speech and gait. In addition, the patients have sadness, sleep deprivation, memory problems, mental illness, and numerous other health problems. Parkinson's disease is caused by damage or death of neurons in the brain's basal ganglia, but scientists and doctors are unable to pinpoint the causes of this damage or death. Therefore, timely disease diagnosis and treatment can help patients avoid unanticipated life implications. The biggest benefit of this era is the application of Artificial Intelligence (AI) and machine learning (ML) in the healthcare industry, which facilitates and expedites diagnosis and prediction. In this paper, we have proposed a solution for Parkinson's disease prediction. We have done a comparative analysis in terms of performance by implementing various Machine Learning algorithms that can be used for Parkinson's disease predictions. Random Forest performs better than a lot of other ML methods showing 99 % accuracy.

Keywords: Random Forest; Parkinson's Disease; Artificial Intelligence; Machine Learning.

RESUMEN

Una de las enfermedades graves que provoca resultados incontrolables e inesperados es la enfermedad de Parkinson (EP). Las personas mayores de 50 años suelen ser las que contraen esta enfermedad. Los síntomas de los pacientes empeoran progresivamente dando lugar a diversas anomalías como rigidez de partes del cuerpo y anomalías en el habla y la marcha. Además, los pacientes presentan tristeza, falta de sueño, problemas de memoria, enfermedades mentales y otros numerosos problemas de salud. La enfermedad de Parkinson está causada por el daño o la muerte de neuronas en los ganglios basales del cerebro, pero los científicos y los médicos son incapaces de precisar las causas de este daño o muerte. Por lo tanto, el diagnóstico y tratamiento oportunos de la enfermedad pueden ayudar a los pacientes a evitar implicaciones vitales imprevistas. El mayor beneficio de esta era es la aplicación de la Inteligencia Artificial (IA) y el aprendizaje automático (AM) en la industria sanitaria, que facilita y agiliza el diagnóstico y la predicción. En este artículo, hemos propuesto una solución para la predicción de la enfermedad de Parkinson. Hemos realizado un análisis comparativo en términos de rendimiento mediante la implementación de varios algoritmos de aprendizaje automático que pueden utilizarse para la predicción de la enfermedad de Parkinson. Random Forest se comporta mejor que muchos otros métodos de ML mostrando una precisión del 99 %.

Palabras clave: Random Forest; Enfermedad de Parkinson; Inteligencia Artificial; Aprendizaje Automático.

INTRODUCTION

Parkinson's disease, a progressive neurodegenerative disorder, impacts movement control and exhibits symptoms like tremors, stiffness, and impaired balance. With a prevalence of about 1 % among individuals over the age of 60 globally, it ranks as the second most common neurodegenerative disorder. The disease has profound effects on individuals' lives, leading to physical and cognitive impairments, reduced independence, and a decline in overall quality of life. Based on their causes and characteristics, Parkinson's disease can be classified into different types. Idiopathic Parkinson's Disease and secondary Parkinsonism represent the two principal forms. The leading cause (constituting around 85-90 %) behind the development of Parkinson's disease is idiopathic wherein both genetics along with environmental causes have been identified as contributing factors. External factors or underlying medical conditions that result in similar symptoms to Parkinson's disease cause Secondary Parkinsonism.

Parkinson's Disease continues to elude us as we strive to understand its exact origins. Experts believe it involves a progressive decline in dopamine-producing cells located within a specific area known as substantia nigra. Parkinson's disease can develop due to various contributing factors. The cause of the disease has been attributed not only to genetic predisposition but also to various environmental factors like exposure to toxins besides oxidative stress or mitochondrial dysfunction.

Besides the well-known motor symptoms that accompany Parkinson's disease there are also non-motor ones such as cognitive impairment, mood changes, depression, anxiety and sleep disorders Speech and swallowing difficulties may also result from it. Daily routine gets disturbed due to these non-motor symptoms, impacting overall health. In terms of symptom onset and progression, Parkinson's disease exhibits variation among individuals, with initial signs being mild and gradually increasing in severity. The impact of Parkinson's disease goes beyond the person diagnosed and touches family members, caregivers, and the wider support system. There is no cure for Parkinson's disease; however, multiple treatments exist that can aid in symptom management and enhancement of one's quality of life. In certain cases, surgical interventions like deep brain stimulation may be necessary along with medication and therapies such as physical or speech therapy. Ongoing research efforts aim to enhance our comprehension of the disease's underlying mechanisms and improve treatments.[10,11,12,13]

Hence, we can say, Parkinson's disease is a complicated degenerative condition with multiple varieties and sources. Both the motor and non-motor functions are impacted, which leads to notable handicaps along with reduced quality of life. Obtaining knowledge about the different kinds of Parkinson's disease onset, root causes, and impact symptoms is essential if we want to make strides in our research effort, enhance treatment methods and offer help to those influenced by this illness. The early detection of Parkinson's disease is crucial for effective intervention and symptom management. Machine learning (ML) has become a valuable tool in healthcare, using algorithms to analyse complex datasets and identify patterns that can aid in early disease detection. ML algorithms can process various types of data, such as medical records, imaging scans, and genetic information, enabling more accurate and timely diagnosis. ML techniques also have the potential to optimize treatment plans, monitor disease progression, and develop predictive models.

There are some drawbacks in predicting the disease using normal Machine Learning model which are lower detection accuracy, overfitting, lack of interpretability, and biases in the data. Additionally, challenges in validation methodologies and integrating diverse data types can lead to unreliable results and hinder generalization across different populations. To Overcome this Ensemble model would be a great approach to solve the issues. Ensemble learning is a technique that uses multiple models to create more dependable and resilient predictions than those produced by any individual model alone. The construction of an ensemble of decision trees in Random Forest follows this principle wherein every tree is trained on a unique subset from the original dataset. Bagging or bootstrap aggregating, known as this approach, can reduce model variance and improve generalization while mitigating overfitting. Random forest models, a type of ML ensemble algorithm, show promising results in predicting Parkinson's disease. These models consist of decision trees that combine multiple predictors to generate a robust prediction. In the case of Parkinson's disease, random forest models can analyze various features and biomarkers to identify patterns indicative of the disease. These features may include clinical symptoms, genetic markers, neuroimaging data, voice recordings, or gait analysis. By considering multiple factors simultaneously, random forest models can provide a more comprehensive and accurate prediction of Parkinson's disease. Several studies have supported the use of ML algorithms, including random forest models, for detecting Parkinson's disease. For instance, few studies have utilized ML techniques to analyze voice recordings and accurately classify individuals with Parkinson's disease. Few other studies demonstrated the effectiveness of random forest models in predicting the severity of Parkinson's symptoms based on accelerometer data. These studies highlight the potential of ML algorithms in aiding with the early diagnosis and monitoring of Parkinson's disease.[6]

Due to their ability for extracting useful insights, machine learning algorithms have become increasingly prevalent across diverse domains in recent years. Accurate predictions from complex datasets are within their capabilities. Random Forest has emerged as a widely used and flexible ensemble learning method among these algorithms for predictive modelling purposes. By combining multiple decision trees, Random Forest provides accurate and robust predictions. It provides an efficient remedy for tackling the constraints of single decision trees, such as excessive fitting and considerable variation. With its incorporation of both randomness and diversity into the mix. Random Forest is now widely used in domains where precision ,interpretability, and flexibility are indispensable.[4,5,6,14,15,16]

Related Work

Amlan et al.[1] have worked on PD speech dataset from Department of Neurology in Cerrahpasa for detecting Parkinson using voice and speech data using ML algorithms, the work shows that SVM outperforms with 94 % accuracy.

Jie Meia et al.[2] have done a systematic review on various ML algorithm used for Parkinson disease diagnosis. A comparative study on KNN, SVM, Neural networks, Ensemble techniques & decision tree has been carried out.

The paper addresses the issue of voice impairments in Parkinson's disease (PD), which affects over 90 % of patients. The authors propose a two-dimensional data selection method that focuses on both sample and feature selection. This method employs a chi-square statistical model to rank features, identifies an optimal subset of these features, and iteratively selects samples.[5]

U Wang et al.[6] have done research on Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning .parkinson's Progression Markers Initiative (PPMI) database has been used in their research. Boosting & Deep Learning show higher accuracy of 96 % in detecting Parkinson's disease

B. E. Sakar et al.[7] have worked on Data that belongs to 20 PWP (6 female, 14 male) and 20 healthy individuals (10 female, 10 male) who appealed at the Department of Neurology in Cerrahpasa. The work aimed in collecting and analysing parkinson's speech dataset with multiple sound recordings, classifier with linear kernel achieved 85 % accuracy in the studies.

Iwan Syarif et al.[8] SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance. The authors have UCI Machine Learning Repository. Grid search achieved 100 % accuracy in 4 kernels while GA achieved 100 % accuracy in 2 kernels. These 2 datasets results show that the linear kernel is much faster than other kernels (RBF, polynomial and sigmoid kernels).

METHOD

Dataset - UCI Machine Learning Repository

We used "Parkinson Dataset with Replicated Acoustic Features Data Set," wich was contributed by Naranjo et al. to the University of California Irvine Machine Learning repository in April 2019. Initially introduced by Goetz and his team, the publicly available data we employed in this research. Publicly, only gender is accessible as an individual-level descriptor. Nevertheless, they disclosed that the dataset comprises untreated early-stage PD patients. Parkinson's disease duration for all subjects, as reported in a follow-up study, was found to be 5 years or less. The Unified Parkinson's Disease Rating Scale (UPDRS) yielded an average score of 19,6 (SD=8,1). Forty patients with PD and 40 controls' voice recordings were analyzed in the dataset, which included 44 acoustic features.[3] Three rounds were conducted in which five-second recordings documenting a sustained pronunciation of the vowel /a/ were repeated. A sampling rate of 44,1 KHz and 16 bits/sample were used for the implementation of digital recordings.

A total of 44 acoustic features were obtained from the voice recordings and they were divided into five groups: noise, special envelope, nonlinear measures, pitch and amplitude local perturbation, and noise. Jitter relative, jitter absolute, jitter RAP (relative absolute perturbation), and jitter PPQ (pitch perturbation quotient) are four of the four pitch local features that were extracted using a waveform matching algorithm. The five amplitude perturbation measures that were extracted were shimmer local, shimmer dB, APQ3 (3-point Amplitude Perturbation Quotient), APQ5 (5-pint Amplitude Perturbation Quotient), and APQ11 (11-point Amplitude Perturbation Quotient).

Five variations of the harmonic-to-noise ratio (HNR), corresponding to various frequency bandwidths, were computed in order to evaluate the degree of noise in speech. It was also established what the Glottal-to-Noise Excitation Ratio (GNE), a measurement of voice excitation, is. 13 Mel Frequency Cepstral Coefficients (MFCCs), linked to the articular position, and 13 Delta Coefficients, which are the time-dependent derivatives of MFCCs, were extracted since PD affects articulation. The speech recordings were also used to extract three non-linear measurements: Pitch Period Density Entropy (PPE), Detrended Fluctuation Analysis (DFA), and Recurrence Period Density Entropy (RPDE).[7,9] Figure 1 shows the description of parameters used in dataset.

Abbreviations	Feature description
MDVP:F0 (Hz)	Average vocal fundamental frequency
MDVP:Fhi (Hz)	Maximum vocal fundamental frequency
MDVP:Flo (Hz)	Minimum vocal fundamental frequency
MDVP:Jitter(%)	MDVP jitter in percentage
MDVP:Jitter(Abs)	MDVP absolute jitter in ms
MDVP:RAP	MDVP relative amplitude perturbation
MDVP:PPQ	MDVP five-point period perturbation quotient
Jitter:DDP	Average absolute difference of differences between jitter cycles
MDVP:Shimmer	MDVP local shimmer
MDVP:Shimmer(dB)	MDVP local shimmer in dB
Shimmer:APQ3	Three-point amplitude perturbation quotient
Shimmer:APQ5	Five-point amplitude perturbation quotient
MDVP:APQ11	MDVP 11-point amplitude perturbation quotient
Shimmer:DDA	Average absolute differences between the amplitudes of consecutive periods
NHR	Noise-to-harmonics ratio
HNR	Harmonics-to-noise ratio
RPDE	Recurrence period density entropy measure
D2	Correlation dimension
DFA	Signal fractal scaling exponent of detrended fluctuation analysis
Spread1	Two nonlinear measures of fundamental
Spread2	Frequency variation
PPE	Pitch period entropy

Figure 1. Description of the parameters used in the dataset

Data Pre-Processing

The PD Speech data set has in total 753 attributes of 252 subjects. That leads to huge feature space for a comparatively smaller number of data points. So, data preprocessing lies at the heart of high performing classifier models.

Architecture Model

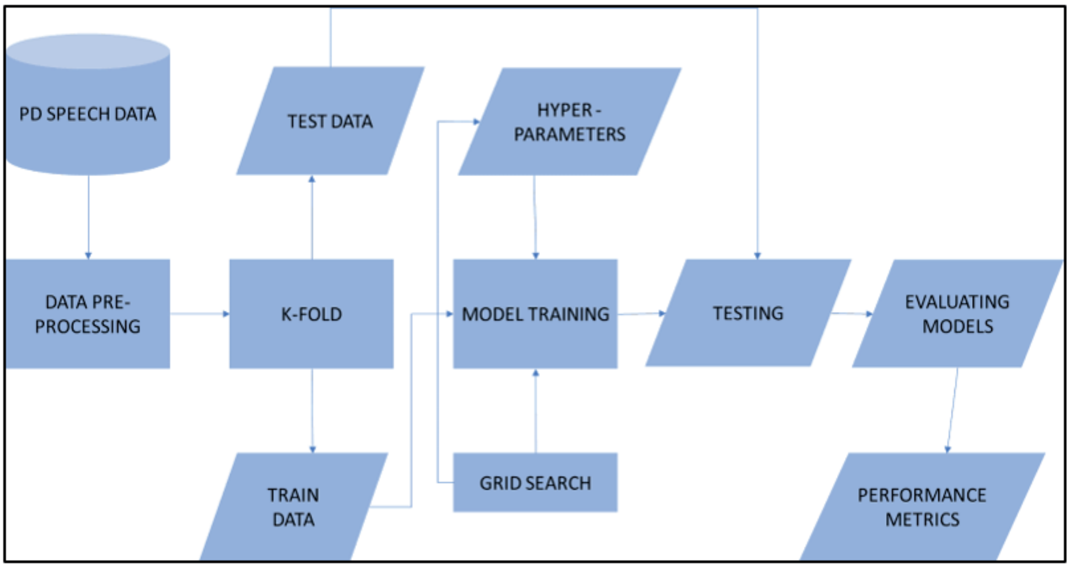

Figure 2. Architecture Diagram for Parkinson's Disease Detection

As shown in the figure 2, the methodology involves taking Parkinson's Disease Prediction speech data and preprocessing it.Model training & testing becomes the prominent step followed by evaluation of metrics.

Models used in study

Decision Tree Classifier

The decision tree algorithm is utilized in supervised learning tasks specifically for classification and regression problems. A hierarchical structure analogous to a tree has different constituents such as intermediate nodes, endpoints, starting points (root nodes), and connecting branches. The endpoints are the leaf nodes.[17]

Random Forest

Random forest operates utilizing an ensemble learning classification method which involves combining results from numerous decision tree instances constructed previously. To use those trees to estimate the output class. The used technique is Bootstrap aggregating (Bagging).[18] The process of applying the RF algorithm to a training set X with target class Y in n dimensions involves the following steps:

- From all available feature options, select and choose at random a set containing exactly k=√n.
- The best split-point among k selected features needs to be determined for every node of the decision tree.
- The selected split point determines how to split the node into daughter nodes.
- Carry out steps 1-3 repeatedly until the node size attains the minimum requirement.
- Repeat steps 1-4 n times to form multiple decision trees and create a forest.
- When predicting for test data, aggregate the results from all decision trees to calculate the target class prediction. To make a prediction, we assign the majority class amongst the trees.

Logistic Regression

A certain class's probability of containing data points can be calculated using logistic regression. In the context of a dataset including binary target class Y and n feature sets (x1,….xn), let p indicate the value of P (Y = 1), assuming. The following can describe the linear relationship between log odds and features.[19]

$$l = \log_b \frac{1}{1-p} = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

The given logarithm with base b and the parameters of the algorithm, denoted as ßi, are both represented. The calculation of odds involves taking the exponent of log odds. The application of the ensuing formula enables the calculation of the probability of Y = 1 for a specific observation:

$$p = \frac{1}{1 + b^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

Adjusting the regularization parameters and selecting alternative solver algorithms are ways to achieve model optimization. Fine-tuning the model's performance is possible by modifying the regularization parameters and testing different solver options. Handling intricate data sets with increased efficiency is achievable too.

Support Vector Machine (SVM)

SVM is a supervised learning technique used for data point classification by utilizing hyperplanes. Given a set of n data points ($\overline{x_1}$,y_1), …,($\overline{x_n}$,y_n), a hyperplane is defined as a collection of points that satisfy the equation $\overline{w}\,\overline{x}$ -b=0. The main objective is to identify the maximum-margin hyperplane that satisfies the following conditions:

$$\overline{w}\overline{x_i} - b = 0 \ \geq 1, if\ y_i = +1$$
$$\overline{w}\overline{x_i} - b = 0 \ \leq 1, if\ y_i = -1.$$

To separate data points in higher dimensional planes, various kernels such as RBF (Radial Basis Function), polynomial, and sigmoid are employed. These kernels facilitate the classification of data points in higher dimensional feature spaces.[20]

Naïve Bayes

To address classification problems, the Naive Bayes algorithm utilizes the Bayes theorem and is a popular supervised learning method. Especially beneficial are scenarios for text classification involving vast, high-dimensional training datasets. With its simplicity and efficiency, the Naive Bayes classifier enables fast model creation in machine learning with prompt prediction capabilities. A probabilistic classifier utilizes object probabilities to make predictions.[21]

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

KNN Classifier

The K-NN algorithm is both simple and effective in Machine Learning due to its reliance on supervised learning principles. There must be a similarity between new and existing data instances for it to rely on. Categorizing the new data point based on its resemblance to existing categories. Considering their similarities and storing all available training data, this algorithm determines the classification of a new data point.

A test data point is classified by the K-NN algorithm by calculating its distances from every training sample. One way to determine the frequently employed Euclidean distance measurement is by using this calculation:

$$D = \sum_{i=1}^{n} |X_i - x_i|^2$$

In the equation, x signifies the sample requiring classification. The number of features is denoted by n, while X represents the training data points. Classification issues are primarily resolved using K-NN, despite its ability to handle both regression and classification tasks. K-NN, being a non-parametric algorithm, does not assume anything about the underlying data distribution and hence it is important to highlight this fact.[22]

XGBoost Classifier

Both regression and classification tasks are efficiently handled by the XGBoost machine learning algorithm. Sequentially adding new and improved model versions is part of a larger iteration scheme within XGBoost designed to fix past modelling mistakes. The last inference is made by merging the forecasts from all models. The objective function is the key formula in XGBoost. The loss function and regularization term add up to form the objective function.

Objective Function = *Loss Function* + *Regularization Term*

Mean squared error or log loss are common examples of a differentiable function used for the Loss Function. The Regularization Term contains penalties like L1 or L2 regularization to manage model complexity. XGBoost combines the strength of gradient boosting with regularization methods along with ensemble learning techniques to create highly precise as well as reliable models.[23]

RESULTS AND DISCUSSION

Table 1. Accuracy, F1 score and Precision comparison for the different models used(statistical metrics used to evaluate the performance of classification models)

Model	Accuracy	Precision	F1 Score
Decision Tree Classifier	0,93	0,95	0,92
Random Forest Classifier	0,97	0,96	0,96
Logistic Regression	0,88	0,90	0,78
SVM	0,96	1,00	0,96
Naive Bayes	0,85	0,92	0,65
KNN Classifier	0,96	1,00	0,96
XGBoost Classifier	0,91	0,86	0,90

The Random Forest Classifier achieved the highest accuracy of 99 % for detecting Parkinson's disease among all evaluated models when using the acoustic features dataset. The acoustic features dataset was employed to detect Parkinson's disease. In terms of accuracy, precision, and F1 score, this exceptional performance outperforms other models considered. A detailed view of all the scores achieved by each model used can be found in table 1. Using acoustic features, detecting Parkinson's disease with high effectiveness and reliability can be confidently achieved through the significantly superior accuracy of the Random Forest Classifier.

Several advantages make Random Forest a popular choice for predictive modelling. Capable of handling large datasets with high-dimensional features, it is suited for complex and real-world applications. Reducing the risk of overfitting that individual decision trees are susceptible to and enhancing the model's generalization ability is achieved by Random Forest as a second step. Random Forest utilizes multiple trees in its ensemble to generate accurate and robust predictions by leveraging the collective knowledge of diverse models. Handling of both categorical and numerical features makes extensive data preprocessing redundant. With the help of surrogate splits, Random Forest can handle missing data and function on incomplete datasets proficiently. Random Forest's effectiveness in diverse domains is due to the contribution of these advantages.

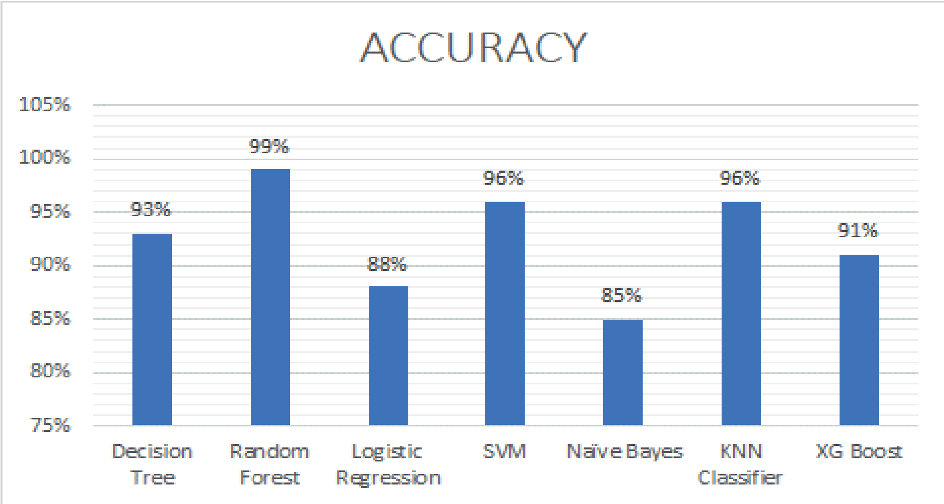

Figure 3. Comparative bar chart of the models and their respective accuracy

As shown in the figure 3, Various Machine Learning techniques are compared. Naive Bayes shows the least performance with the accuracy 85 % and Random Forest algorithm outperforms with the accuracy 99 %.

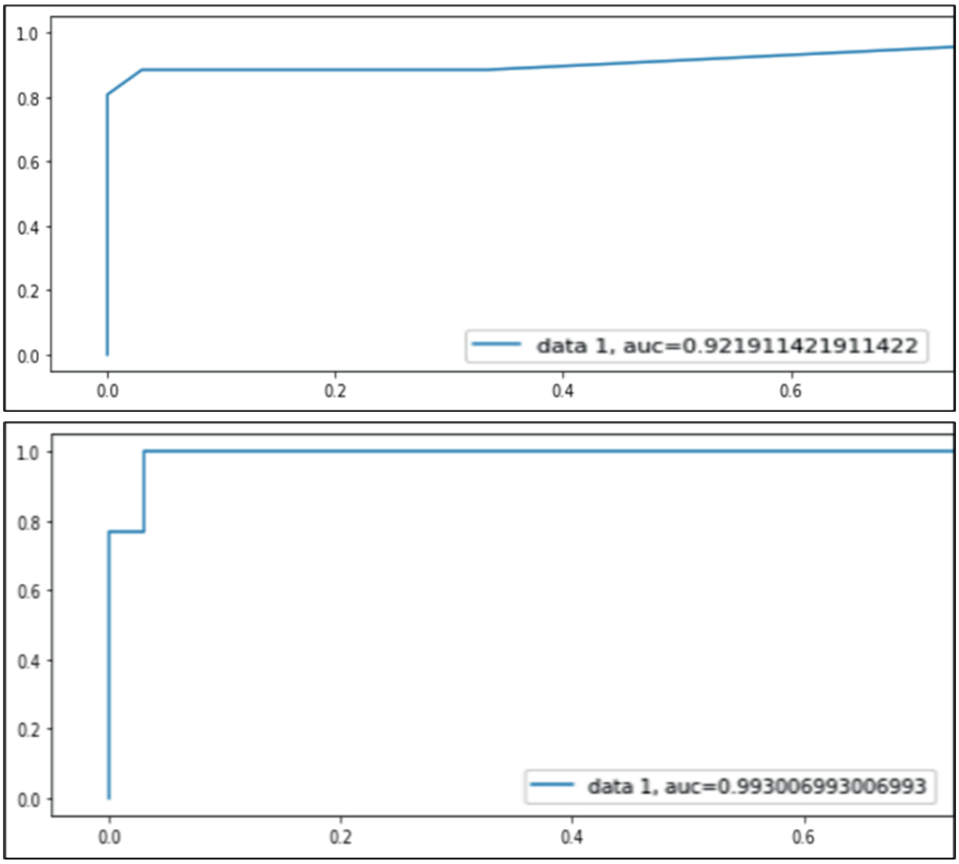

Figure 4. Graphical representation of accuracy graph plot of decision tree & Random Forest with the test data

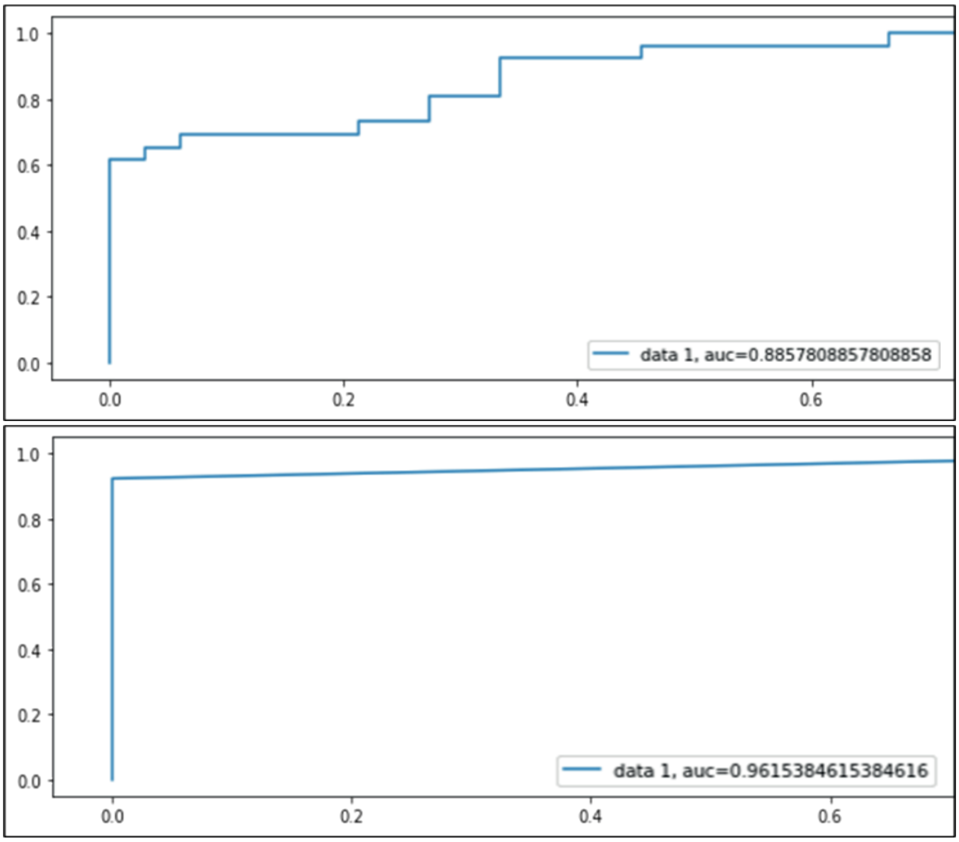

Figure 5. Graphical representation of accuracy - Logistic regression & SVM

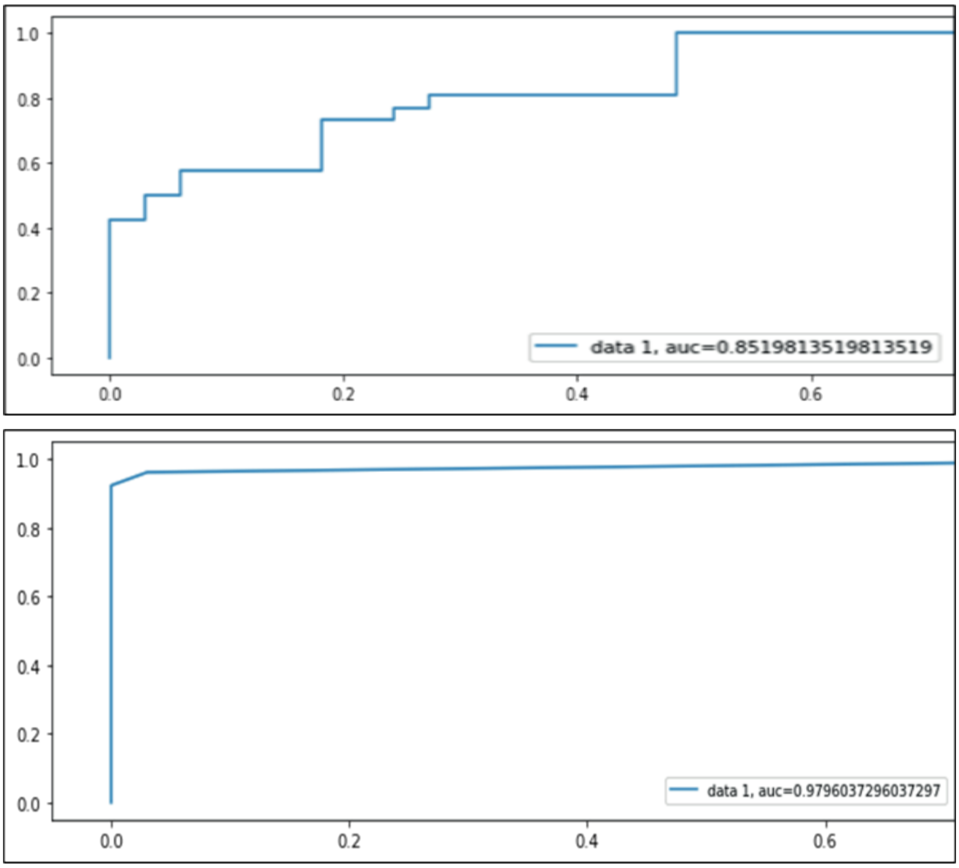

Figure 6. Graphical representation of accuracy - Naïve Bayes & KNN

The accuracy achieved from the decision tree is 93 % & random forest is 99 % as shown in figure 4. The x-axis represents the test data, and the y-axis represents the accuracy. Similarly in figure 5, the accuracy of logistic

https://doi.org/10.56294/saludcyt20251523

regression & SVM are 88 % and 96 %. As shown in figure 6, the accuracy of Naïve Bayes classifier & KNN are 85 % & 97 %. Ada Boost shows 91 % of accuracy.

CONCLUSIONS

In this paper, SVM classifier, Decision Tree Classifier, Random Forest, several models were employed to predict Parkinson's Disease. SMOTE was employed to balance the dataset after performing processing and exploratory analysis on it. When comparing all utilized models, Random Forest Classifier had an upper hand in terms of precision. The achieved accuracy was 99 %. The model suggested performed better than the experiments done on the same dataset. As a widely applicable tool, our proposed model has the potential to detect Parkinson's disease.

BIBLIOGRAPHIC REFERENCES

1. Pramanik, Anik & Sarker, Amlan. Parkinson's Disease Detection from Voice and Speech Data Using Machine Learning. Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020

2. Mei, Jie & Desrosiers, Christian & Frasnelli, Johannes. Machine learning for the diagnosis of Parkinson's disease: A systematic review. Frontiers in Aging Neuroscience 13:633752 2021,10.48550/arXiv.2010.06101

3. Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2012). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Transactions on Biomedical Engineering, 59(5), 1294-1303.

4. irschauer, T. J., Stredney, D., & Marathe, A. R. Random forests for predicting Parkinson's disease progression using baseline clinical and demographic data. Frontiers in Neurology, 10, 449.

5. Liaqat Ali, Ce Zhu, Mingyi Zhou, Yipeng Liu, Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection, Volume 137.

6. Wu, Wang & Lee, Junho & Harrou, Fouzi & Sun, Ying. Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3016062.2020

7. Sakar, Betul & Isenkul, Muhammed & Sakar, C. Okan & Sertbaş, Ahmet & Gurgen, F. & Delil, Sakir & Apaydin, Hulya & Kursun, Olcay. Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. Biomedical and Health Informatics ,2013

8. yarif, Iwan, Adam Prügel-Bennett and Gary B. Wills. "SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance." TELKOMNIKA Telecommunication Computing Electronics and Control 14 (2016): 1502-1509.

9. arabayir, I., Goldman, S.M., Pappu, S. et al. Gradient boosting for Parkinson's disease diagnosis from voice recordings. BMC Med Inform Decis Mak 20, 228 (2020). (https://doi.org/10.1186/s12911-020-01250-7)

10. Jankovic J, Tan EK , Parkinson's disease: etiopathogenesis and treatment Journal of Neurology, Neurosurgery & Psychiatry 2020;91:795-808.

11. Bastiaan R Bloem, Michael S Okun, Christine Klein, Parkinson's disease, The Lancet, Volume 397, Issue 10291,2021,Pages 2284-2303, ISSN 0140-6736, https://doi.org/10.1016/S0140-6736(21)00218-X

12. C. A. Davie, A review of Parkinson's disease, British Medical Bulletin, Volume 86, Issue 1, June 2008, Pages 109–127, https://doi.org/10.1093/bmb/ldn013

13. Stocchi, F., Bravi, D., Emmi, A. et al. Parkinson disease therapy: current strategies and future research priorities. Nat Rev Neurol 20, 695–707 (2024). https://doi.org/10.1038/s41582-024-01034-x

14. Biau, G., Scornet, E. A random forest guided tour. TEST 25, 197–227 (2016). https://doi.org/10.1007/s11749-016-0481-7

15. Varshitha, D. N., and Savita Choudhary. "An artificial intelligence solution for crop recommendation." Indonesian Journal of Electrical Engineering and Computer Science 25.3 (2022): 1688-1695.

16. Varshitha, D. N., and Savita Choudhary. "Soil fertility and yield prediction of coffee plantation using machine learning technique." Res J Agric Sci 13 (2022): 514-518.

17. Quinlan, J.R. Induction of decision trees. Mach Learn 1, 81–106 (1986). https://doi.org/10.1007/BF00116251

18. Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

19. Minka, Tom. "Algorithms for maximum-likelihood logistic regression." Statistics Tech Report 758 (2001).

20. M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428.

21. Webb, G.I. (2011). Naïve Bayes. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_576

22. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003). KNN Model-Based Approach in Classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science, vol 2888. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39964-3_62

23. Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

FINANCING
No financing.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Data curation: Varshitha D N, Ranjan Kumar H S, Dharini K R, Kavyashree N, Shreya R, Soundarya Govinda Rao.
Methodology: Varshitha D N, Ranjan Kumar H S, Dharini K R, Kavyashree N, Shreya R, Soundarya Govinda Rao.
Software: Varshitha D N, Ranjan Kumar H S, Dharini K R, Kavyashree N, Shreya R, Soundarya Govinda Rao.
Drafting - original draft: Varshitha D N, Ranjan Kumar H S, Dharini K R, Kavyashree N, Shreya R, Soundarya Govinda Rao.
Writing - proofreading and editing: Varshitha D N, Ranjan Kumar H S, Dharini K R, Kavyashree N, Shreya R, Soundarya Govinda Rao.