



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 9, September 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Review on Parkinson's Disease Prediction Using Machine Learning

Ajay Y P, Srinivasulu M

Student, Department of Master of Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India

Assistant Professor, Department of Master Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India

ABSTRACT: A complicated neurological condition, Parkinson's disease (PD). For the disease's first management, an error free diagnosis in the untimely stages is critical. This study analyzes the approaches used in PD diagnosis that were created using machine learning methodologies. To conduct the comparison analysis, we depend on clustering and prediction having to learn through this approach. To be more explicit, we employ support vector regression ensembles and several clustering approaches for PD data clustering to forecast Motor-UPDRS and Total-UPDRS. The end results are then contrasted with those secured by using other prediction learning techniques such as several linear regression, nerve analysis, and regression with support vectors. To conduct the contrast analysis, its depend on clustering and prediction having to learn methodologies. To forecast Motor-UPDRS and Total-UPDRS more precisely, we use support vector regression ensembles together with a number of clustering techniques for PD data clustering. The final result are then compared to those discovered using further prediction learning methods, such as multiple linear regression, neural analysis, and regression with nodes.

I. INTRODUCTION

The second most prevalent and complicated neurodegenerative ailment globally is Parkinson's disease (PD). PD can be brought on by polygenic and biological factors. It has been discovered that a single gene contributes to the development of the illness in 1%–2% of PD cases (most of which are familial). Along with additional symptoms including insomnia (nonmotor characteristics), cardiac arrhythmia, and constipation, the chief symptoms of PD include slowed movement (motor features), muscular rigidity, and quake. One of the characteristics of PD is the alteration of voice and speech. The essential form for tracking PD symptom development is the Unified Parkinson's Disease Rating Scale, or UPDRS, which shows the existence and severity of indicator. The UPDRS is regarded as the most applied clinical rating scale and quite well tool for PD patients. The 4 components of the UPDRS—UPDRS I, UPDRS II, UPDRS III, and UPDRS IV—are used to assess the severity of treatment-related problems as well as solid motor functions evaluated in PD and detected by physical examination. This scale is often based on the Total-UPDRS, which has a range of 0-176 (176 for total disability and 0 for health), and the Motor-UPDRS, that also represents the UPDRS' motor portion and has a range of 0-108 (108 indicating severe motor impairment and 0 indicating healthy state).

Machine learning procedure have appear that they can hold wider range of medical datasets and provide sensible guidance. By modifying large datasets of patient information, ML-based solutions may prevent infections, improve the standard of medical treatment, lower healthcare costs, and complement doctors' efforts. ML techniques are frequently used to attribute and illnesses. Gadekallu and others. Inspected how to diabetic retinopathy using machine learning techniques. For the categorization of the collected features from the dataset for diabetic retinopathy, the authors employed a Deep Neural Network (DNN) model based on PCA and the Grey Wolf Optimization (GWO) method. The accuracy, recall, sensitivity, and specificity assessment metrics were used to assess the approach and to compare it to the support vector machine (SVM), naive Bayes classifier, decision tree (DT), and XGBoost. Comparing their approach to the SVM, DT, and XGBoost methods, they found that their method was more accurate overall. A technique for categorization a multimodal stroke dataset with skewed data was developed by Bhattacharya et al. The Antlion optimization technique was applied by the authors to the DNN model to quickly choose the best hyperparameters. This technique has the advantage of using just 38.13% of the training time on the stroke dataset. One of the most important methods for classifying and predicting diseases is an artificial neural network. Clinical decision support systems

(CDSSs), according to Brindle, are unique instruments created to assist medical practitioners in their decision-making while taking into account certain conditions or diseases. To assist medical experts in making precise and timely forecasts, ML methods might be used to create efficient CDSS. A significant role has been performed by CDSSs created employing machine learning techniques in determining the presence or severity of the illness.

Unsupervised methods are employed in machine learning techniques to reduce the dimensionality of data, allowing the illness to be detected. Additionally, these methods enable data manipulation, noise removal, similarity calculations, and data segmentation. The ultimate categorization, prediction, and diagnosis of the disease, on the other hand, are made possible via the application of supervised learning techniques. Although ML has demonstrated its value, its effective use requires significant work from human experts because no single technique can deliver acceptable results in all conceivable circumstances. Researchers have access to medical data, but the best use of these sources may be constrained by a lack of experience with handling large quantities of data. Additionally, even if a number of algorithms have been utilised to forecast diseases using various real-world medical datasets, the choice of the deployed methodology should take accuracy enhancement and computation time efficiency into account.

This study compares several machine learning (ML) methods for following the course of Parkinson's disease remotely. Based on clustering and prediction learning techniques, the comparison investigation. In the last stage of the suggested strategy, this study employs ensemble learning to further increase the UPDRS prediction's accuracy. Prediction challenges have shown to benefit from ensemble learning techniques. Ensemble learning techniques have not been widely used in research to create illness diagnostic systems. To determine how well these techniques perform in UPDRS forecast, more research is required. As a result, we employ ensembles of support vector regression and several clustering strategies for PD data sets. The outcomes are then contrasted with different modeling learning techniques, including support vector regression, multiple linear regression, nervous techniques, and deep belief networks (DBN).

II. LITERATURE REVIEW

A literature review goes over books, academic papers, and other noteworthy sources pertinent to a certain topic, field of study, or theory in order to give a description, summary, and necessary distinction of these works in connection to the problem-solving goal. The goal of literature reviews is to explain to your audience how your research fits within a larger field of study and to give a summary of the sources you have consulted. A literature evaluation in the social sciences frequently follows an organizational structure and incorporates every precis and synthesis, typically within specific conceptual categories. A literature overview might also include a summary of all important sources. A synthesis is a reorganization, or rearranging, of those records in a way that informs how you are preparing to present yourself at a search problem. A summary is a recap of the main records of the source

III. EXISTING AND PROPOSED SYSTEM

Existing System

Parkinson's disease is diagnosed by a neurologist, a doctor who specializes in disorders of the nervous system, based on a neurological and physical examination, an analysis of your signs and symptoms, and your medical history. Every time, this system is incredibly laborious.

Proposed System

The initial step after pre-processing and loading the dataset is Exploratory Data Analysis to comprehend the data at hand and find the key aspects for developing precise models. This is accomplished through the use of feature importance analysis, a group of methods for assigning scores to identify the key features from the input data, each of which is utilized to make a prediction. The dataset can be better understood by feature importance analysis. Which features are more important to the target are highlighted by the relative scores. A predictive model fitted to the dataset determines the most significant scores. Various modelling strategies are put into use after the data have been prepared and valuable insight into significant features have been gained from the data.

IV. RESEARCH METHODOLOGY

For this review paper, we used to develop a model using XGBClassifier, various Python libraries like sciKit-learn, NumPy, pandas, and xgboost were used. We used these procedures and followed this flow to finish this project:

1. In the first step loaded the necessary info.
2. Based on loaded info get some features and labels .
3. In the next step scaled the features
4. After that split the dataset
5. Based on data set built an XGBClassifier
6. Based on XGBClassifier, determined the model's accuracy is the final step.

XGBoost

To solve the Parkinson's disease detection problem, we used the XGBoost algorithm. It is very useful if the data set is large. In order to provide better results with fewer computer resources and in the shortest period of time, it is the ideal mix of software and hardware optimization approaches. It is basically an extreme gradient boosting method introduced by data scientist Tianqi Chen in 2014. Since then this algorithm has gained big popularity and compared to other machine learning (ML) techniques it is quite simple. Its accessibility and advanced features make it a versatile algorithm for Windows, Linux, and OS X. To implement XGBoosting gradient boosting is required. This is a machine learning algorithm for regression, classification, and ranking. Gradient boosting requires a loss function (a function that maps an event or values of one or more variables onto a real number), a weak learner, and an additive model (a nonparametric regression method) to calculate loss and modify decision trees to minimize the error. The XGBoost algorithm supercharges gradient boosting tasks. When applied, this open-source software library untangles regression, classification, and ranking issues. It's very fast, accurate, and accessible, so it's no wonder that it has been adopted by numerous companies, from Google to start-ups. XGBoost is designed within the framework of the decision tree algorithm of gradient boosting. It is scalable. The algorithm can produce billions of outcomes quickly. Along with the tree learning algorithms, XGBoost can solve linear models, too. That means it can handle parallel computations on one machine.

XGBoost Accessibility

The XGBoost software library is accessible from many common interfaces, including, but not limited to:

- Command Line Interface
- C++
- Scikit-learn / Python
- Caret package / R
- Julia
- Java and JVM languages
- Apache Spark, Hadoop, and Flink

Specifications of XGBoost:

- **Regularization:** composite models can be disciplined by XGBoost using both L1 and L2 variational. Overfitting can be prevented by regularisation.
- **Processing the data approach:** one-hot encoding or missing values can make data sparse. To address various sorts of sparsity patterns in the data, XGBoost uses a sparsity-aware split discovery method.
- **Weighted quantile sketch:** When the weights of the data points are balanced, the generality of ongoing tree-based algorithms can locate the split sites (using quantile sketch algorithm). They cannot handle weighted data, nevertheless. To efficiently handle weighted data, XGBoost offers a distributed weighted quantile sketch approach.
- **Block structure for parallel learning:** The CPU's numerous cores can be used by XGBoost to speed up computation. This is made feasible by the block structure used in the system architecture. Blocks are the in-memory units where data is sorted and kept. In contrast to other techniques, this allows the data layout to be reused over consecutive rounds rather than having to be computed from scratch.
- **Cache awareness:** Getting the gradient statistics by row index in XGBoost necessitates non-continuous memory access. Consequently, XGBoost has been developed to make the best use of hardware. The gradient statistics are saved in internal thread buffers that are allocated for this purpose.

V.DISCUSSION

Effective PD identification is essential since prompt diagnosis and treatment can prevent the onset of symptoms and challenges related to this condition. Although quick diagnosis of PD is important, doing so is difficult since existing diagnostic methods are frequently focused on subjective symptoms. Additionally, nonmotor symptoms including sadness, sleep issues, and rapid eye movement are more commonly recognised than motor symptoms in the early stages of Parkinson's disease (PD), which slows down the diagnosis.

Recently, ML has been utilised to diagnose medical conditions, including PD. This can be explained by the practical functionality and precise outcomes of ML approaches. Predictive modelling is important in the classification of illnesses. It is regarded as a significant data mining method since it groups the population according to a specified standard. In order to determine which classification technique offers the highest performance, it is critical to compare the results of several categorization methods. Therefore, the primary objective of this study is to evaluate various methods used for PD prediction and categorization. The assessment of ML techniques based on different datasets makes it difficult to create a meaningful comparison between the deployed approaches, despite the fact that ML methods have been evaluated in several studies individually. Therefore, it is crucial to compare different approaches in a single research using a selected dataset.

The EM and SOM methods were used in this study's initial round of data processing to provide a range of segment counts for the ensembles used in the study's final clustering results. The generated clusters were then utilised for ensemble learning SVR method UPDRS prediction. The findings are then shown for various ensembles of clusters and their numbers. Based on the results and the R2, PA, IA, MAE, and RMSE metrics, we can say that the performance of SOM and EM ensembles using HGPA is the best of the deployed methodologies. Additionally, a number of tests on the PD dataset were used to compare the suggested technique to earlier methods in the literature. Comparing the outcome of the deployed strategy with alternative strategies (SVR, ANFIS, MLR, HSLSSVR, and NN). Compared to previous ensemble learning methodologies, the results showed that the HGPA + EM + SVR ensemble delivered greater accuracy.

VI.CONCLUSION

Parkinson's disease has been affecting people for thousands of years. Many plant-based remedies, some of which are still in use today, have been used to treat this condition, whose consequences on humans are severe, with different degrees of success. It is troubling because there is still no diagnostic test to identify the condition. Using machine learning-based techniques, this research offered the prediction of Parkinson's disease. We now know about XGBoost, a technique for scaling tree boosting that yields cutting-edge results. Other issues relating to machine learning systems can also be resolved using this technique. The sparse data are handled by a sparsity aware method, and a fairly weighted quantile sketch is utilised for approximative learning. With a minimum of resources, XGBoost can resolve issues of a real-world scale.

REFERENCES

- [1] B. S. Carsten Lundby, Paul Robach, "Blood Doping -- Effects and Detection," Br. J. Pharmacol., pp. 1– 25, 2012, doi: 10.1111/j.1476.
- [2] M. Ali et al., "Parkinson ' s Disease Read the Story of a Patient Diagnosed with Disease," no. 2267, p. 60008, 2020.
- [3] J. S. Hawley, "What is Parkinson ' s disease ?," Park. Dis. Improv. Patient Care, vol. 501, no. c, pp. 1–6, 2014.
- [4] J. Browniee, "A Gentle Introduction to XGBoost for Applied Machine Learning," Mach. Learn. Mastery, pp. 1–20, 2016, [Online]. Available: <https://machinelearningmastery.com/gentleintroduction-xgboost-applied-machine-learning/>.
- [5] R. Betarbet, T. B. Sherer, and J. T. Greenamyre, "Animal models of Parkinson's disease. [Review] [89 refs]," Bioessays, 2002.
- [6] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: a practical introduction," BMC Med. Res. Methodol., vol. 19, no. 1, pp. 1–18, 2019, doi: 10.1186/s12874-019- 0681-4.
- [7] S. Bind, A. K. Tiwari, and A. K. Sahani, "A Survey of Machine Learning Based Approaches for Parkinson Disease Prediction," Int. J. Comput. Sci. Inf. Technol., 2015.

- [8] D. Avci and A. Dogantekin, "An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel-Extreme Learning Machine," *Parkinsons. Dis.*, 2016, doi: 10.1155/2016/5264743.
- [9] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, and V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease," *Futur. Gener. Comput. Syst.*, 2018, doi: 10.1016/j.future.2018.02.009.
- [10] P. Piccini and A. Whone, "Functional brain imaging in the differential diagnosis of Parkinson's disease," *Lancet Neurology*, 2004, doi: 10.1016/S1474-4422(04)00736-7.
- [11] "AUTOMATIC RECOGNITION OF PARKINSON'S DISEASE FROM SUSTAINED PHONATION TESTS USING ANN AND ADAPTIVE NEURO-FUZZY CLASSIFIER," *Mühendislik Bilim. ve Tasarım Derg.*, 2010, doi: 10.21923/mbtd.12375.
- [12] M. Ene, "Neural network-based approach to discriminate healthy people from those with Parkinson's disease," *Ann. Univ. Craiova, Math. Comp. Sci. Ser.*, 2008.
- [13] D. Gil and M. Johnsson, "Diagnosing Parkinson by using artificial neural networks and support vector machines," *Glob. J. Comput. Sci. Technol.*, 2009.
- [14] D. Chang, M. Alban-hidalgo, and K. Hsu, "Diagnosing Parkinson's Disease From Gait," *Stanford*, 2015.
- [15] S. Shetty and Y. S. Rao, "SVM based machine learning approach to identify Parkinson's disease using gait analysis," 2016, doi: 10.1109/INVENTIVE.2016.7824836.
- [16] S. V. Perumal and R. Sankar, "Gait monitoring system for patients with Parkinson's disease using wearable sensors," 2016, doi: 10.1109/HIC.2016.7797687.
- [17] F.-T. S., G. N., P. C., H. T., G. L., and H. J.M., "Effect of gait speed on gait rhythmicity in Parkinson's disease: Variability of stride time and swing time respond differently," *J. Neuroeng. Rehabil.*, 2005.
- [18] Bharath S., Hsu M., Kaur D., Rajagopalan S., Andersen J. K. Glutathione, iron and Parkinson's disease. *Biochemical Pharmacology*. 2002;64:1037–1048. doi: 10.1016/s0006-2952(02)01174-7. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [19] Cacabelos R. Parkinson's disease: from pathogenesis to pharmacogenomics. *International Journal of Molecular Sciences*. 2017;18:p. 551. doi: 10.3390/ijms18030551. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [20] Engelender S., Isacson O. The threshold theory for Parkinson's disease. *Trends in Neurosciences*. 2017;40:4–14. doi: 10.1016/j.tins.2016.10.008. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [21] Natale G., Pasquali L., Ruggieri S., Paparelli A., Fornai F. Parkinson's disease and the gut: a well known clinical association in need of an effective cure and explanation. *Neuro-Gastroenterology and Motility*. 2008;20:741–749. doi: 10.1111/j.1365-2982.2008.01162.x. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [22] Krohn L., Grenn F. P., Makarious M. B., et al. Comprehensive assessment of PINK1 variants in Parkinson's disease. *Neurobiology of Aging*. 2020;91:168–161. doi: 10.1016/j.neurobiolaging.2020.03.003. [[PMC free article](#)] [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [23] Eskidere Ö., Ertaş F., Hanilçi C. A comparison of regression methods for remote tracking of Parkinson's disease progression. *Expert Systems with Applications*. 2012;39:5523–5528. doi: 10.1016/j.eswa.2011.11.067. [[CrossRef](#)] [[Google Scholar](#)]
- [24] Nilashi M., Bin Ibrahim O., Mardani A., Ahani A., Jusoh A. A soft computing approach for diabetes disease classification. *Health Informatics Journal*. 2018a;24:379–393. doi: 10.1177/1460458216675500. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [25] Tsanas A., Little M., McSharry P., Ramig L. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *Nature Precedings*. 2009;57(4):884–893. doi: 10.1038/npre.2009.3920.1. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]
- [26] Goetz C. G., Poewe W., Rascol O. Movement disorder society task force on rating scales for Parkinson's disease: the unified Parkinson's disease rating scale (UPDRS): status and recommendations. *Movement Disorders*. 2003;18:738–750. [[PubMed](#)] [[Google Scholar](#)]
- [27] Moguel-Cobos G., Saldivar C., Goslar P. W., Shill H. A. The relationship between social anxiety disorder and motor symptoms of Parkinson disease: a pilot study. *Psychosomatics*. 2020;61:321–326. doi: 10.1016/j.psych.2020.03.006. [[PubMed](#)] [[CrossRef](#)] [[Google Scholar](#)]



INNO  SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN  INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details