JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Predictive Modeling for Parkinson's Disease Diagnosis: A Comparative Study of Machine Learning Algorithms

[1]Mr.Mubin Jamadar, [2]Mr.Pranay Dhangar, [3]Mr.Tejas Chougule, [4]Mr.Aditya Hulikire

[1]Student, [2] Student, [3] Student, [4] Student
[1]Sharad institute of Technology, college of Engineering, Yadrav
Ichalkaranji, India

Abstract : Parkinson's disease (PD) persistent consideration is limited by lack, inconsistent symptom checking, infrequent access to mind, and limited experiences with human services professionals, leading to subpar clinical dynamics and subpar patient wellbeing related results. The promise of dramatically altering the indicative, observing, and helpful detecting in PD has enabled target and distant checking of impaired motion function. Parkinson Disease (PD) is an incurable neurological condition that is progressive in nature. Delaying the PD's progression is greatly helped by early diagnosis. Almost 90% of PD patients experience dysphonia, which is the most noticeable early symptom. For PD patients, precise, non-invasive, and reliable forecasts are crucial, and voice features-based early diagnosis with the integration of artificial intelligence plays a significant role in this regard. Many upper limb functional tests, including the Motor UPDRS, were used to illustrate our point. In this review, it was recommended that managed grouping algorithms like help vector machines (SVM), credulous Bayes, k-closest neighbors (K-NN), and artificial brain organizations (ANN) can actually analyze abstract diseases whose proposed strategies for diagnosis incorporate element choice in view of separating, wrapping, and order processes. A technique like this might cut down on the time and costs involved in PD screening because just a small number of clinical test characteristics would be needed for the diagnosis.

IndexTerms - Machine learning, Algorithms, Parkinson’s disease, Diagnosis, Artificial Intelligence.

I. INTRODUCTION

Parkinson's disease is the leading cause of death and disability worldwide. The Parkinson's Institute estimates that by 2020, he will have one million Parkinson's patients in the United States[C. Marras et al., 2018]. Neuropathology and Histopathology support the clinical administration of Parkinson's disease[D. Gelb et al., 1999] (2012) [H. Hazan et al.]. Based on the affectability and specificity of the hallmark Parkinson illness highlights, medical diagnostic detection of Parkinson Disease should be attainable across the board. As a result, clinical, pathologic, and nosology research based on the occurrence of events, attributes, and including danger components of tests are predicted to investigate Parkinson Disease [D. Aarsland et al., 2003]. (1999) D. Gelb et al. Parkinson's disease normally influences a huge part of patients more than 50 around the world, and has done as such as of not long ago. Parkinson disease as of now has no perceived etiology, yet it is probable ready to free side effects in the beginning phases from the emotional patients deliberately [N. Singh et al., 2007]. As can be seen from the overview, 90% of affected patients had a weak voice. Treatment for Parkinson's disease is considered very expensive. Because of this, most people cannot afford to pay for Parkinson's disease treatment. Today, a key factor in accurately diagnosing Parkinson's disease for clinical professionals is Parkinson's disease expectations. Parkinson's disease can now be identified using a comprehensive hierarchy based on machine learning, a phenomenal activity. Advances in some clinical areas, including medical services, have greatly expanded the amount of clinical information.

Parkinson's disease affects the nervous system. Movement problems are the main symptoms. The "substantia nigra" is a region of the brain where dopamine is produced. As dopamine levels fall in Parkinson's disease, the substantia nigra's cells begin to die. When they have decreased by 60 to 80 percent, Parkinson's symptoms start to emerge.

It hazy causes Parkinson's disease explicitly. It very well may be affected by both inherited and natural elements. A couple of scientists think infections may likewise cause Parkinson's. Parkinson's disease has been connected to low levels of the synapse dopamine and the chemical norepinephrine, which controls dopamine. Lewy body protein is also found in the hearts of people with Parkinson's disease. The involvement of Lewy bodies in Parkinson's disease movement is unknown to experts.

The diagnosis of Parkinson's disease can't be made utilizing a particular test. The diagnosis depends on the patient's clinical history, the aftereffects of physical and neurological tests, as well as an examination of the side effects. Another choice is a dopamine carrier (DAT) clear. Albeit these tests can't demonstrate Parkinson's, they can help preclude different ailments and backing the expert's diagnosis.

II. LITERATURE REVIEW

Prashanth et al. zeroed in on utilizing NB, SVM, Helped Trees, and RF to utilize nonmotor signals in the early diagnosis of PD. The outcomes showed that SVM's exactness worth of 96.40% was the most noteworthy. Abiyev and Abizade presented a clever FNS and NN-based PD diagnosis approach. The review's discoveries showed the FNS's powerful presentation when contrasted with elective procedures.

A clever strategy for PD recognizable proof utilizing SVM was accounted for by Singh et al., who guaranteed a 100 percent in general exactness. Utilizing ANN, MLP, and GRNN, imen and Bolat [60] focused on the vocal signs for PD diagnosis. The outcomes showed that GRNN gave the best presentation. Shetty and Rao utilized SVM to zero in on stride signals in the diagnosis of PD and other neurological circumstances. The exactness of the proposed strategy was 83.33%. A cross breed philosophy was introduced by Nilashi, Ibrahim, and Ahani utilizing EM, PCA, ANFIS, and SVR strategies. The review's decisions demonstrated the way that the approach can exactly decide the disease's seriousness.

Machine learning-based Parkinson's disease diagnosis is introduced by Senturk, Z. K. in 2020. Processes for include determination and characterization make up the recommended diagnosis strategy. For the component determination task, the Element Significance and Recursive Component Disposal approaches were thought about. Parkinson's patients in the tests were arranged utilizing Grouping and Relapse Trees, Artificial Brain Organizations, and Backing Vector Machines. It has been shown that Help Vector Machines with Recursive Element End beat different procedures.

Younis Thanoun, M., and T. YASEEN, M. O. H. A. M. M. A. D. proposed utilizing two kinds of Groups learning techniques Stacking Classifier and casting a ballot Classifier, which are feasible methodologies of PD location utilizing machine learning. We then, at that point, thought about the results of the two of them. Casting a ballot classifier was beated by the stacking classifier approach, which had exactness consequences of 92.2% and 83.57%, separately. Higher discovery exactness for clinical applications, like this persistent condition, could result from this correlation study.

In this work, Gracious, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., and Acharya, U. R. utilized the EEG information from twenty patients with Parkinson's disease (PD) and twenty solid workers. This is a thirteen-layer CNN engineering that can swap the requirement for the customary component portrayal stages.

The 2019 Concentrate by Beli, Bobe, Bada, Olaja, Uri-Jovii, and Kosti is a controversial paper on the use of artificial intelligence by machine learning algorithms in the kinematic study of developmental problems, particularly Parkinson's disease (PD). No attempt is made to provide a complete overview. With an accentuation on the latest exploration distributed, we inspected distributions distributed between January 2007 and January 2019 through internet-based data sets, like PubMed and Science Direct. To figure Parkinson's disease, Celik, E., and Omurca, S. I. (2019, April) inspected different grouping algorithms, including Strategic Relapse, Backing Vector Machine, Additional Trees, Inclination Supporting, and Irregular Woods.

In their paper published in July 2019, Wodzinski, Skalski, Hemmerling, Orozco-Arroyave, and Nöth describe a method for Parkinson's disease diagnosis that makes use of vowels with sustained phonation and a ResNet architecture that was designed for picture classification in the first place. We determined the audio recordings' spectrum and sent that information into a ResNet architecture that had already been trained using the ImageNet and SVD databases.

III. RESEARCH METHODOLOGY

The essential algorithms of the suggested ML model are SVM, naive Bayes, KNN, and ANN. As they are simple to use and only require a few parameters to be tweaked, these algorithms are frequently utilized in the literature. Creating a model to identify PD in speech recordings entails a number of steps. For better comprehension, pertinent features are first retrieved from the dataset. In the subsequent stage, machine learning algorithms are utilized to sort both sound people and individuals with Parkinson's disease (PD). These procedures depend on acoustic highlights to conjecture the results as visual diagrams and tables with rate precision scores. Because only a few voice features were used in the suggested approach rather than labor-intensive feature extraction techniques like MRI, motion sensors, or handwriting analyses, it is demonstrated that it is more efficient in terms of computational cost than the other procedures. Also, the effectiveness of several well-known classifiers was assessed, and the best classifier for issues with PD diagnosis was discovered to be ANN.

The connection between's the indicator and free factors is determined utilizing a learning calculation, while channel-based approaches are factual systems that are not subject to it. The relevance of the predictor factors to the target variable is assessed. The ML model is then constructed using the variables with better scores. To further develop PD identification, this examination looks to utilize a channel based include determination strategy to recognize the most relevant highlights.

IV. MATERIALS AND METHODS

A. Parkinson's Disease Detection Using Machine Learning and Speech Analysis

A few scientists have zeroed in on the key recurrence (F0) changes welcomed on the vocal action or the acoustic level. In individuals with sensorineural discernibility misfortune and the disease, the impacts of influence range examination of F0 phonation have been contemplated. As far as the pervasiveness and power of the problems, F0's beat was outstanding. Enrichment also showed that the F0 test could be a useful method for revealing neurological disorders.The fundamental frequency of discourse transmission was found using the autocorrelation ability technique. Prosodic abnormalities in Parkinson's disease are usually thought to manifest as distinct neuromotor conditions.

To help the hypothesis, specialists took a gander at how a gathering of patients comprehended and created pitch qualities. Conventional medications, including LDOPA, express that LDOPA is an extremely successful treatment for emotional disease in the beginning phases of PD. In this review, the researchers characterize the patient's discourse information as "extreme" or "not serious" utilizing profound learning. The bound together Parkinson's disease rating scale was the assessment device utilized in this examination (UPDRS). The engine UPDRS rates the patient's engine capability on a size of 0 to 108, while the generally UPDRS offers values going from 0 to 1766.

B. Parkinson's disease classification using an ML classifier

In this technique, the disease will be arranged utilizing a ML classifier. To start with, we decide the objective variable for this report — patient wellbeing status — and count the quantity of patients. Subsequent to assessing the patient's wellbeing, we graphically address the information. A lot of the dataset was used for preparing, while 20% was utilized for testing. The example's 48 sound people are addressed by a score of 0, and the example's 147 Parkinson's disease victims are addressed by a score of 1. There are 147 out of 195 (75.38%) Parkinson's disease patients in the dataset. 48 out of 195 (24.62%) of the members in the dataset were sound.

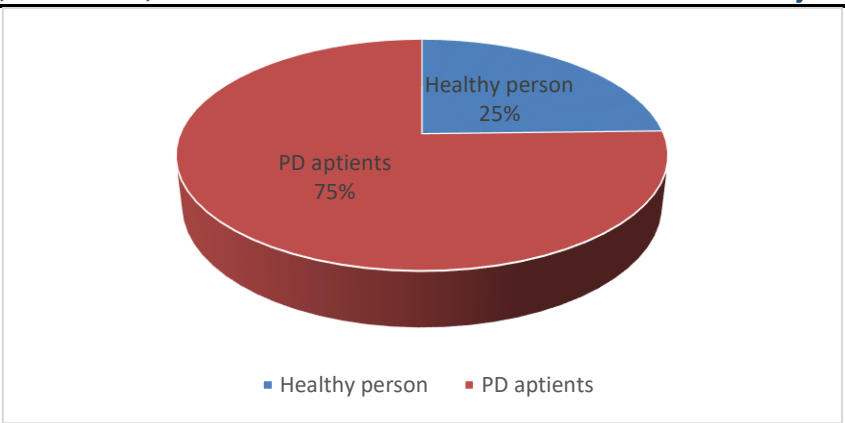

Figure1: Patient's Health Condition with PD.

C. Development of Machine Learning Methods Using Classifier Evaluation Metrics

It is easy to recognize the disease when different classifiers are utilized. It very well may be recognized utilizing characterization awareness, the Matthews' relationship coefficient (MCC), precision, explicitness, the F-score (F-measure), and other estimation boundaries. There is a calculation formula for each of these measurement criteria, which may be used to choose the classifier that is the most qualitatively suitable for the analysis. The confusion matrix must be the primary consideration before these criteria are created.

The precision of a model on a given dataset is addressed by the F1-Score, generally known as the F-Score, as delineated in Condition (1):

$$F - Score = 2 * \frac{precision * sensitivity}{precision + sensitivity}$$

MCC: Used in model evaluation to evaluate the accuracy of dual and multiclass placement as shown in the state(2). It depends on the ideas of bogus negative, misleading positive, and genuine negative, genuine positive. It falls inside the accompanying expressed scope of 1 and 1:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

(−1): Prediction and observation are in conflict

(0): No more accurate than a random prediction

(1): excellent classifier (precise expectation).

V. RESULTS AND ANALYSIS

The proposed work is carried out utilizing JupyterLab and Python 3.7. Here, we portray the four machine learning arrangement algorithms' exhibition in a trial and their exploratory plan.

A. SVM-Classifier

SVM is one of the most utilized classifier models since it yields exact and very solid outcomes. SVM's essential goal is to characterize preparing information by partitioning it into particular classes while completing a different class learning action. It precisely recognizes information examples and offers the best order execution on preparing information. The grouping exactness is ended up being higher in SVM because of its more grounded speculation capacities, and the preparation procedure utilizes a successive minimization technique. Condition is used to ascertain the direct SVM (3).

$$y = f(x) = w^T x - b$$

where w is the vector weight symmetric about the selection hyperplane, b is the hyperplane offset, x is the information, y is the class marker, and T is the rendering manager.

We use the sklearn library for the SVM classifier module for grouping the datasets provided in this review. Results delivered by the SVM classifier are shown in Table 1.

Table 1: SVM classifier

Name	Results
Score of test data accuracy	87.17%
Score of training data accuracy	87.45%
Execution period	0.03112 s
F-1 score	65.18%
MCC	57.58%

B. Naive Bayes Classifier

The credulous Bayes classifier technique is one more key part of ML's classification strategy. Most of results are acquired utilizing the credulous Bayes technique, and it offers proficient order and learning. In view of Bayes' hypothesis, Credulous Bayes ascertains the likelihood of an occasion happening given its conditions. For example, vocal changes are run of the mill of the people who have the condition; subsequently, these side effects are related with the anticipation for this disease's diagnosis. The first Bayes hypothesis, which gives a strategy to computing the probability that an objective event will happen, is broadened and streamlined by the guileless type of the hypothesis. The information incorporates an extensive variety of discourse signal varieties to check the probability of the ailment. The classifier module for the innocent Bayes classification is given by the sklearn Gaussian credulous Bayes calculation. Table 2 shows the classifier's result.

Table 2: Naïve Bayes Classifier Results

Title	Results
Accuracy rate of Taste Data	75.12%
Accuracy rate of training data	77.22%
Execution time	0.0313 s
F-1 score	85.75%
MCC	65.58%

C. Artificial Neural Network

Profound brain organizations' ANN region makes expectations about how the human cerebrum capabilities. The human mind and ANN vary essentially from each other overall. The machine just has a limited complete of processors, though the mind has 'n' equal neurons. Neurons are also weaker and more laid back than computer processors. The brain's capacity to process information on a bigger scale is another significant difference between computer systems and the brain. Synapses or networks that work together make up neurons. The primary goal of this article is to categorize how ANN approaches work in the early diagnosis of this disease, which is based on the following steps:

- Explaining the role that ANN plays in this disease's detection and its function.
- making observations about dataset labels and features.
- Grouping the various studied illness categories according to their symptoms.
- Looking at the dependable outcomes.

These discoveries can likewise be utilized to the clinical field as direction for designers contemplating ANN sending to work on the capability of city wellbeing in light of the explored sickness.

The preparation dataset (70%) and the test dataset (30%) were isolated into two bits for the artificial brain network analyze. The typical precision score, which was the most noteworthy of all the arrangement techniques, was viewed as very high for the artificial brain organization's order discoveries, at 95.8% in Table 3.

Table 3: Results of an Artificial Neural Network Classifier

Title	Results
Accuracy rate of Taste Data	95.8%
Accuracy rate of training data	96.35
Execution time	0.035 s
F-1 score	86.02%
MCC	71.12%

VI. COMPARATIVE STUDY AND ANALYSIS

In contrast with any remaining trial machine learning models and the present status of the craftsmanship, the recommended model — which utilizes four machine learning algorithms — shows improved results. ANN created the best outcomes in the proposed study, beating the other trial algorithms with an exactness pace of 95.6%. Utilizing first class recording hardware, the creators assembled 25 PD and 25 HC discourse datasets, then used KNN and SVM to assess the datasets to distinguish PD. The exactness paces of KNN and SVM classifiers were 58.51% (LOSO) and 67.44% (LOSO), individually. Likewise, the creators utilized various algorithms, including C4.5, C5.0, irregular timberland, and Truck, which depends on choice trees. The analysis was led on 50 records of people, of which 45% had the abstract disease and 55% were sound controls. The most noteworthy typical model exactness for this examination, 65.6%, was reached. ANN was used to find PD. The machine learning library at the University of California, Irvine provided the data set. Forty-five functions were definitely selected in the assembly cycle using MATLAB devices as information values and results. Their proposed model effectively recognized PD cases and sound people with an exactness of 95.92%. In a review, the creators utilized classifiers from irregular timberland, SVM, MLP, and KNN to distinguish PD patients from solid controls. Utilizing the SVM and KNN classifiers, the review's discoveries were 77.8% and 83.1%, separately. The creators of a review contrasted patients and Parkinson's disease (PWP) and solid controls (HC) utilizing a scope of discourse tests. Human variable cepstral coefficients (HFCC) were utilized in this examination. The typical voice print for each voice recording was made utilizing the removed HFCC. SVM was used for the grouping utilizing various bits, including RBF, polynomial, straight, and MLP. The best exactness of 86.4% was made conceivable by the direct portion of the SVM.

The exhibition of the recommended technique is contrasted and comparative ML strategies for PD examination in different conditions and with different kinds of assessed PD datasets notwithstanding the correlations previously talked about. As seen in the above table, the proposed strategy fared better than other comparable contributions of ML methods for diagnosing PD, making it superior to others.

VII. CONCLUSION

After Alzheimer Disease, Parkinson Disorder (PD) is the most prevalent neurodegenerative condition. By combining the voice modality with an Artificial Intelligence (AI) system, the early vocal symptom can be used as a foundation to create an early diagnosis system for Parkinson's disease. PD and HC can be distinguished using automated ML algorithms that use non-invasive speech biomarkers as features to forecast the outcome. In this work, we examine the effectiveness of different machine learning classifiers for disease detection in noisy high-dimensional data. Potential to achieve clinical-level precision with careful element selection. Human voice signals are utilized to analyze the ailment. The got results show that include choice methodologies are compelling when utilized with ML classifiers, especially while working with discourse information where countless phonetic properties can be removed. With the recommended early diagnosis methodology, PD can be precisely recognized in its beginning phases and its extreme side effects can be stayed away from. To accomplish the most elevated level of exactness, various classification algorithms are utilized in the field of clinical imaging. To upgrade classifier execution and get the most noteworthy exactness rating, this examination might be utilized to different machine learning methods and datasets. Future endeavors will make benefit of the all-around existing accounts and increment the quantity of existing highlights to work on the precision of the models created. A few sorts of records handling instruments that are accessible online may likewise be used to look at the obtained information.

REFERENCES

[1] Belić, M., Bobić, V., Badža, M., Šolaja, N., Đurić-Jovičić, M., & Kostić, V. S. (2019). Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—A review. Clinical neurology and neurosurgery, 184, 105442.

[2] C. Marras et al., ―Prevalence of Parkinson's disease across North America, ‖ npj Park. Dis., vol. 4, no. 1, p. 21, Dec. 2018.

[3] Celik, E., &Omurca, S. I. (2019, April). Improving Parkinson's disease diagnosis with machine learning methods. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-4). Ieee.

[4] D. Aarsland, K. Andersen, J. L.-A. of neurology, and undefined 2003, ―Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study, ‖ jamanetwork.com.

[5] D. Gelb, E. Oliver, S. G.-A. of neurology, and undefined 1999, ―Diagnostic criteria for Parkinson disease, ‖ jamanetwork.com.

[6] F.M. Javed Mehedi Shamrat et al, A Comparative Analysis of Parkinson Disease Prediction Using Machine Learning Approaches, International Journal of Scientific & Technology Research Volume 8, Issue 11, November 2019 ISSN 2277-8616

[7] G. Singh, M. Vadera, L. Samavedham, and E. C.-H. Lim, "Machine learning-based framework for multi-class diagnosis of neurodegenerative diseases: a study on Parkinson's disease," Age, vol. 65, pp. 64–94, 2016.

[8] H. Hazan, D. Hilu, L. Manevitz, L. O. Ramig, and S. Sapir, ―Early diagnosis of Parkinson's disease via machine learning on speech data, ‖ 2012 IEEE 27th Conv. Electra. Electron. Eng. Isr. IEEEI 2012, pp. 29–32, 2012.

[9] M. Nilashi, O. Ibrahim, and A. Ahani, "Accuracy improvement for predicting Parkinson's disease progression," Scientific Reports, vol. 6, Article ID 34181, 2016.

[10] N. Singh, V. Pillay, Y. C.-P. in neurobiology, and undefined 2007, ―Advances in the treatment of Parkinson's disease, ‖ Elsevier.

[11] Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. Neural Computing and Applications, 32, 10927-10933.

[12] Parkinson's Disease Information Page | National Institute of Neurological Disorders and Stroke. ‖ [Online]. Available: https://www.ninds.nih.gov/disorders/alldisorders/parkinsons-disease-information-page. [Accessed: 08-Apr-2019].

[13] R. H. Abiyev and S. Abizade, "Diagnosing Parkinson's diseases using fuzzy neural system," Computational and mathematical methods in medicine, vol. 2016, Article ID 1267919, 9 pages, 2016.

[14] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, "High-accuracy detection of early Parkinson's disease through multimodal features and machine learning," International Journal of Medical Informatics, vol. 90, 2016.

[15] S. Çimen and B. Bolat, "Diagnosis of Parkinson's disease by using ANN," in Proceedings of the Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 2016 International Conference on, pp. 119–121, IEEE, Jalgaon, India, December 2016.

[16] S. Shetty and Y. Rao, "SVM based machine learning approach to identify Parkinson's disease using gait analysis," in Proceedings of the Inventive Computation Technologies (ICICT), vol. 2, pp. 1–5, IEEE, Coimbatore, India, August 2016.

[17] Senturk, Z. K. (2020). Early diagnosis of Parkinson's disease using machine learning algorithms. Medical hypotheses, 138, 109603.

[18] Speech impairment in a large sample of patients with Parkinson's disease, ‖ hindawi.com

[19] Wodzinski, M., Skalski, A., Hemmerling, D., Orozco-Arroyave, J. R., &Nöth, E. (2019, July). Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. In 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 717-720). IEEE.

[20] Younis Thanoun, M., & T. YASEEN, M. O. H. A. M. M. A. D. (2020, October). A comparative study of Parkinson disease diagnosis in machine learning. In 2020 The 4th International Conference on Advances in Artificial Intelligence (pp. 23-28).