iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** X **Month of publication:** October 2022

DOI: https://doi.org/10.22214/ijraset.2022.46272

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue X Oct 2022- Available at www.ijraset.com

Parkinson's Disease Detection using Machine Learning Algorithms

Amreen Khanum D[1], Prof. Kavitha G[2], Prof. Mamatha H S[3]
[1, 2]*Dept, of Computer Science and Engineering, University B.D.T College of Engineering, Davanagere-577004,Karnataka ,India*
[3]*Assistant Professor, DOS in Mathematics, University B.D.T College of Engineering, Davanagere-577004, Karnataka , India*

Abstract: ***The main goal of the study is to inspect the performance of the different Supervised Algorithms for the improving the Parkinson Disease diagnosis by detection. We have used Five machine learning techniques for the detection of Parkinson Disease datasets. KNN,LR,DT,NB and XGBoost were used for the prediction of Parkinson Disease. The Performance of the classifiers is evaluated via, precission, Accuracy, F1-Score, Recall and Support. Where after computing with different classifier we got result as KNN shows the accuracy level 96% for Parkinson Disease. XGBoost achieved the second highest classification accuracy of 91%. Moreover, in the terms of accuracy for analyzing the Parkinson Disease dataset ,NB achieved the lowest accuracy of 76%. In our study has emphasized the current Parkinson Disease research trends and scope in relational to clinical research fields by machine learning technique .That will be effective impact in field of Parkinson Disease.***
Keywords: classification, KNN, Decision tree, XGBoost, Naïve bayes , logistic regression.

I. INTRODUCTION

Parkinson's diseases is a Neurodegenerative disease It is a progressive pathology that affects the brain and the nervous system, leading. to the death of nerve cells. The most known and frequent ones are Alzheimer's and Parkinson's diseases Parkinson's disease, is particularly linked to the loss of dopamine producing neurons in the basic ganglia. which has human, social and financial impacts, on a personal, professional and social level. The medication fees for Parkinson's disease are very expensive. At the moment, no cure has been found. Medication is limited to treatments, at an early stage, to improve the patient's quality of life. Several methods have been used to detect the symptoms of Parkinson's disease, but most of them require motor actions that appear only in an advanced state of the disease. Most used traditional methods for the determination of the disease, are costly invasive methods namely SPECT and CT tomography's which are effective, essentially, in the mature stage of the disease. Besides classic methods, practitioners adopt several diagnostic paths. Some of them were based on handwriting by considering the relationship between handwriting and nervous system problems. Others have relied on peripheral biomarkers for early detection of PD.

The detection of Parkinson's disease is based on the use of different classifiers. The distinction between them is based on measurement criteria, namely classification accuracy, F1-score , Recall, Precision, R2-score ... etc. Each of these measurement criteria has formulas to calculate it and conclude which is the most qualitatively adequate classifier for the study. Before defining these criteria, we must focus on the confusion matrix Called a contingency table, it is a tool for measuring the performance of a learning model, checking how far its predictions are correct, compared to reality in classification problems.

II. METHODOLOGY

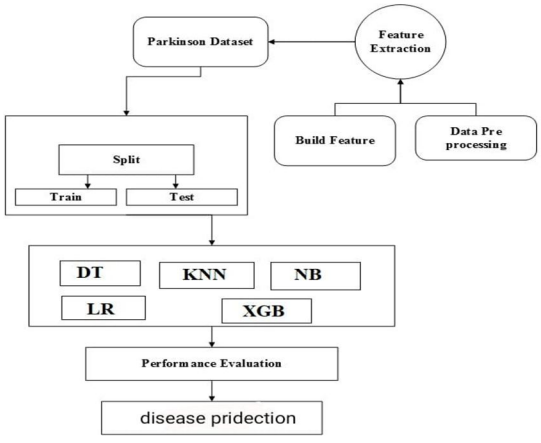


Fig 1 : Architectural design for predicting Parkinson Disease

A. Data Collection

Parkinson Disease Datasets

In this Project, we used the Parkinson disease data from provided by the UCI Machine Learning Repository this dataset is consisting of 195 patient data with 23 features of voice data set is used.

'https://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/parkinsons.data'

B. Data Pre-processing

In this section, firstly we extracted features from the Parkinson disease datasets. We picked the 23 columns and 195 entries of data. Then missing data and huge number of functions in the data collection, the data is very challenging to use effectively. This technique eliminates duplicate information, conflicting data, noisy, incomplete data, and. To boost the network's classification efficiency and redundant values. They are 23 features from the dataset which were we collected. There is no duplicate data is present in this dataset and no columns of the dataset contains any Missing Values.

Table Head	Table Column Head
MDVP: RAP	Relative Amplitude Perturbation
MDVP: PPQ	Five-point Period Perturbation Quotient
Jitter: DDP	Average absolute difference of differences between cycles, divided by the average period
MDVP: Shimmer	Shimmer Local amplitude perturbation
MDVP: Shimmer (db)	Local amplitude perturbation (decibels)
Shimmer: APQ3	3-point Amplitude Perturbation Quotient
Shimmer: APQ5	5-point Amplitude Perturbation Quotient
MDVP: APQ	11-point Amplitude Perturbation Quotient
Shimmer: DDA	Average absolute difference between the amplitudes of consecutive periods
NHR	Noise-to-Harmonics Ratio
HNR	Harmonics-to-Noise Ratio
RPDE	Recurrence Period Density Entropy
D2	Correlation Dimension
DFA	Detrended Fluctuation Analysis
Spread1	Fundamental frequency variation
Spread2	Fundamental frequency variation
PPE	Pitch period entropy

Table Head	Table Column Head
MDVP: F0 (Hz)	Average vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
MDVP: Flo (Hz)	Minimum vocal fundamental frequency
MDVP: Jitter (%)	Fundamental frequency perturbation (%)
MDVP:Jitter (Abs)	Absolute jitter in microseconds

Fig 2 : Attribute Details For Diagnosis Of PD Patient

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue X Oct 2022- Available at www.ijraset.com

C. Machine learning Thecniques

1) *Decision Tree Classifier:* Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.
2) *Logistic Regression:* Logistic Regression was mostly used in the biological research and applications in the early 20th century (Jr, Lemeshow, & Sturdivant, 2013). Logistic Regression (LR) is one of the most used machine learning algorithms that is used where the target variable is categorical. Recently, LR is a popular method for binary classification problems. Moreover, it presents a discrete binary product between 0 and 1. Logistic Regression computes the relationship between the feature variables by assessing probabilities (p) using underlying logistic function.
3) *Naive Bayes:* Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a highdimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
4) *KNN Classifier:* K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.
5) *XGBoost Classifier:* XGBoost is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models. XGBoost may be formally defined as a decision tree-based ensemble learning framework that uses Gradient Descent as the underlying objective function and comes with a lot of flexibility while delivering the desired results by optimally using computational power.

D. Evaluation Criteria

In this work, we used five supervised learning techniques for the detection of Parkinson disease. Therefore, the performance measurements of the classifiers are evaluated by different statistical procedures. Such as Recall, Precision, F1-Score and R2-Score etc.

Hence, the computation method of the measurement considerations are as follows.

$Precision = TP/\ TP + FP$

$Accuracy = TP + TN/\ TP + TN + FP + FN$

$Recall\ or\ sensitivity = TP\ /\ (TP + F\ )$

$Fscore = 2 * precision * sensitivity/\ \ precision + sensitiviy$

III. RESULTS AND DISCUSSION

Here we have evaluated the five-machine learning supervised algorithms for detection of Parkinson Disease. The analysis of five classification techniques were evaluated and we have got different accuracy with different algorithm.

In this section we have lead a comparison between five different machine learning classifier where we have compared the Accuracy, Precision ,F1-Score,Recall and R2-Score values which we have got by this analysis.

	Metric	DT	LR	NB	KNN	XGB
0	Accuracy	0.932203	0.830508	0.762712	0.966102	0.915254
1	F1-Score	0.920000	0.782609	0.650000	0.960000	0.909091
2	Recall	0.884615	0.692308	0.500000	0.923077	0.961538
3	Precision	0.958333	0.900000	0.928571	1.000000	0.862069
4	R2-Score	0.724942	0.312354	0.037296	0.862471	0.656177

Table 1 : Comparison Table

IV. ANALYSIS OF RESULTS

According to the performance measurements of three classification algorithms are presented in figure. The results clearly show that the KNN reached to the highest accuracy (96%) and DT achieved the second highest accuracy. where as Naive Bayes has achieved lowest accuracy (76%), respectively. Finally, KNN is the highest performer by overall performance.

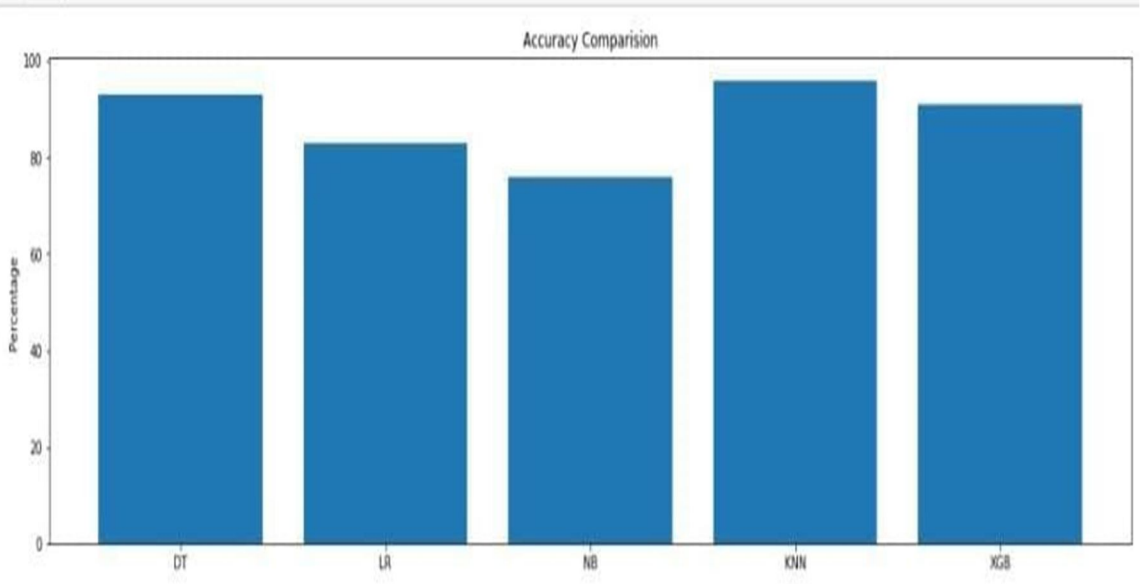


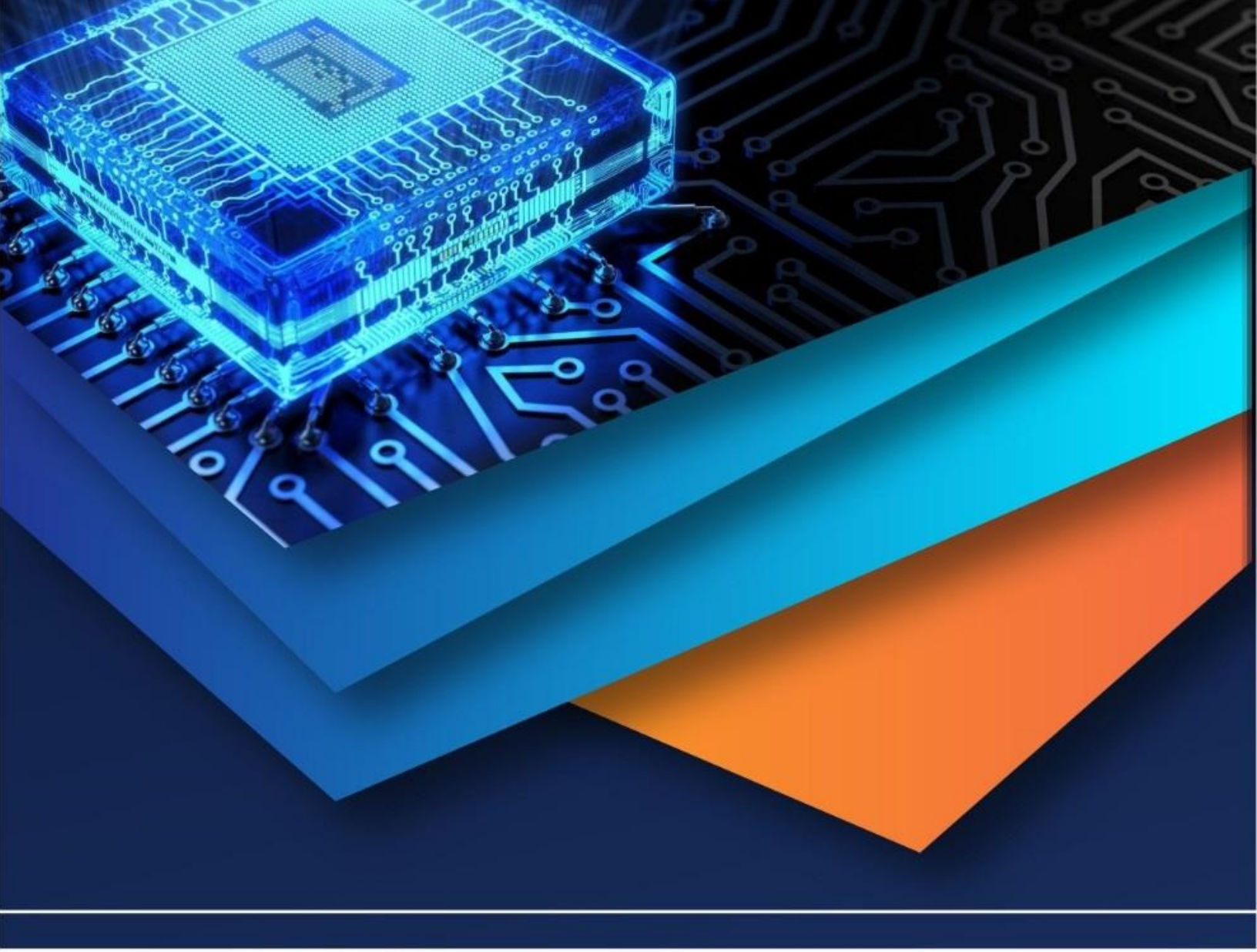
Fig 3 : Classification performance graph

V. CONCLUSION

In this analysis, we have illustrated five supervised learning machine learning approaches. Afterwards, we evaluated the performance of the five classifiers which are used in the prediction of Parkinson disease and assessed their performance using different statistical methods. The tentative performance shows that the KNN have achieved the highest performance than the other four classifiers within the Parkinson datasets. It is 96%.hence we can say that KNN perform good comparative to other classifier . This analysis has used five machine learning techniques for the detection of Parkinson disease based on several parameters. this application will be able to detect in Parkinson disease in very few minutes and notify dangerous probability of having disease. In very minimal steps we can predict the person is healthy or PD patient.

REFERENCES

[1] Basil K Varghese, Geraldine Bessie Amali D*, Uma Devi K S "Prediction of Parkinson's Disease using Machine Learning Techniques on Speech dataset" year- Feb 2019.
[2] Timothy J. Wroge1 , Yasin Ozkanca ¨ 2 , Cenk Demiroglu2 , Dong Si3 , David C. Atkins4 and Reza Hosseini Ghomi "Parkinson's Disease Diagnosis Using Machine Learning and Voice".
[3] Antonio Suppa1,2†, Giovanni Costantini 3†, Francesco Asci 2 , Pietro Di Leo3 , Mohammad Sami Al-Wardat 4 , Giulia Di Lazzaro5 , Simona Scalise6 , Antonio Pisani 7,8 and Giovanni Saggio3 * " voice in parkinson's disease: A Machine learning study " year- Feb 2022 .
[4] Shrihari K Kulkarni1, K R Sumana2 "Detection of Parkinson's Disease Using Machine Learning and Deep Learning Algorithms" year- Aug 2021.
[5] Md. Sakibur Rahman Sajal1,2* , Md. Tanvir Ehsan1,2, Ravi Vaidyanathan3 , Shouyan Wang4 , Tipu Aziz5 and Khondaker Abdullah Al Mamun1,2 "Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis" year-2020.
[6] "Parkinson's Disease Diagnosis: Detecting the Effect of Attributes Selection and Discretization of Parkinson's Disease Dataset on the Performance of Classifier Algorithms"- Gamal Saad Mohamed. Year- Nov 2016.
[7] R. Prashanth and S. D. Roy, ''Early detection of Parkinson's disease through patient questionnaire and predictive modelling,'' Int. J. Med. Informat., vol. 119, pp. 75–87, Nov. 2018.
[8] H. Gunduz, ''Deep learning-based Parkinson's disease classification using vocal feature sets,'' IEEE Access, vol. 7, pp. 115540–115551, 2019.
[9] S. Lahmiri and A. Shmuel, ''Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine,'' Biomed. Signal Process. Control, vol. 49, pp. 427–433, Mar. 2019.
[10] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, ''Automatic detection of Parkinson's disease based on acoustic analysis of speech,'' Eng. Appl. Artif. Intell., vol. 77, pp. 148–158, Jan. 2019.
[11] V. Illner, P. Sovka, and J. Rusz, ''Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease,'' Biomed. Signal Process. Control, vol. 58, Apr. 2020, Art. no. 101831



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY