Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 2, pp. 237–247
DOI : 10.47974/JSMS-1250

Performance evaluation of machine learning algorithms for Parkinson's disease identification model

Shivkumar J. Karale [†]
Nikhil Mangrulkar [@]
Department of Computer Technology
Yeshwantrao Chavan College of Engineering
Nagpur
Maharashtra
India

Sagarkumar S. Badhiye [§]
Symbiosis Institute of Technology, Nagpur Campus
Symbiosis International (Deemed University)
Pune
Maharashtra
India

A. R. Bhagat Patil [‡]
Department of Computer Technology
Yeshwantrao Chavan College of Engineering
Nagpur
Maharashtra
India

Dinesh Bhoyar *
Department of Electrical Engineering
Yeshwantrao Chavan College of Engineering
Nagpur
Maharashtra
India

[†] *E-mail:* **sjkarale@gmail.com**
[@] *E-mail:* **mangrulkar.nikhil@gmail.com**
[§] *E-mail:* **sagarbadhiye@gmail.com**
[‡] *E-mail:* **arbhagatpatil@gmail.com**
* *E-mail:* **dinesh.bhoyar23@gmail.com (Corresponding Author)**

Rajesh Khobragade [#]
Department of Electronics and Computer Science
Shri Rameobaba College of Engineering and Management
Nagpur
Maharashtra
India

Abstract

Though Medical sector is referred to as one of the most advance sector in the Country, there are some implications which shows disorientation of medical sector and early prediction of various ailments. There are some diseases which does not have any preventive measures and are incurable till date. Parkinson's disease is one of those disease that is incurable till date and its clinical diagnosis requires early detection of symptoms which may vary according to age parameter. Parkinson's disease is one of the incurable diseases which affects the Central Nervous System (CNS) of Human brain which leads to various symptoms like barely noticeable tremor on limbs, slowness in movement, improper balance and posture of body, Loss of automated movement and sometimes change in vocal frequency. With theadvent of Machine Learning (ML) and Artificial Intelligence (AI) technique are used for diagnosis of brain diseaseswhich is helping the doctor for early detection of the diseases.Considering the performance evaluation parameters, we have given equal importance to every performance measures used. The dataset for the Parkinson's disease detection is taken from Oxford UCI Machine repository. Analysis and Comparative study show that all the algorithms can detect the person is affected by the Parkinson's disease or not. Out of them random forest algorithm yielded highest performance measures.

Subject Classification: *68Q32.*

Keywords: *Parkinson's disease (PD), Machine learning (ML), Random forest (RF), Xtreme gradient boosting (XGBoost), False positive (FP).*

1. Introduction

Early diagnosis and treatment can significantly improve outcomes for patients for Parkinson's diseases, but the current diagnostic methods are often subjective and rely on clinical evaluation by specialists. In recent years, there has been growing interest in using machine learning algorithms to assist in the diagnosis of Parkinson's disease from various types of medical data, including vocal recordings.

In this study, we aimed to develop a machine learning model for detecting Parkinson's disease from vocal recordings and implement it in a web application. We used a dataset of vocal recordings from patients with

[#] *E-mail:* **khobragaderh@rknec.edu**

and without Parkinson's disease and tested on machine learning algorithms likes Xtreme Gradient Boosting, Logistic Regression, Random Forest, Support Vector Machines, Simple Decision Tree, and K-Nearest Neighbors. We evaluated the performance of each algorithm using a confusion matrix and several performance metrics, including accuracy, precision, recall, F1 score, FP rate and specificity with error validation. Our results indicate that the Random Forest algorithm performed the best out of all the algorithms tested, with an accuracy of 93.22% and a precision of 1.0 and F1 score of 95.65%.This study provides important insights into the potential of machine learning algorithms for assisting in the diagnosis of Parkinson's disease and highlights the importance of developing accurate and reliable diagnostic tools for this debilitating disorder.

The dataset used in this study is a voice modulation dataset obtained from 31 patients were 23 of the patients are having Parkinson's disease. This dataset is publicly available on UCI oxford machine learning repository donated by max little. It has a total of 195 entries and 23 attributes.

2. Proposed Methodology

Parkinson's disease symptoms include loss of motor control along with voice degradation, Dopamine level count. Dataset of voice degradation is enough for this disease detection as 90% of patients shows this symptom. We will be using voice sample data to train the algorithm and judging out the best out of them. Dataset for Parkinson's disease detection is obtained from Oxford UCI Machine Repository. After importing the dataset series of operations like Data Pre-processing, feature selection and splitting of data is being done to get training and testing data from dataset. After training the model various machine learning algorithms is being implemented to obtain the performancein terms of accuracy score, f1 score precision, recall, FP rate and specificity. Error validation must be performed on the algorithms as an important step for accurate results. Based on all selected performance metrics the algorithm which yields highest performance measures is considered as the best among all the implemented algorithms. Figure 1 shows the flow diagram of proposed system.

3. Implementation

i. Importing the dataset.

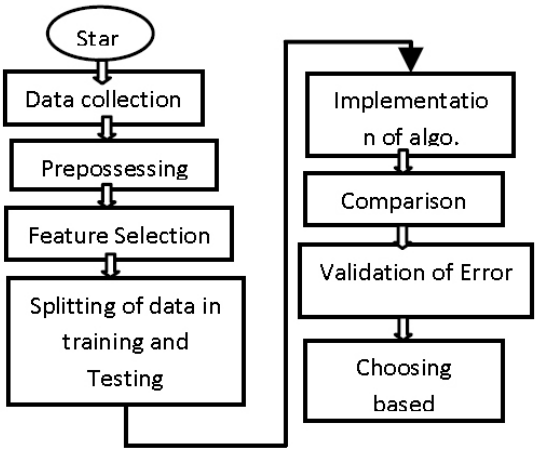

Figure 1
Flow diagram of proposed system

ii. Checking the imported dataset using df.head() to check for the first 5 rows.

iii. Importing necessary libraries with filtering of features followed by extracting labels.

iv. Number of Parkinson's patients in the dataset.

v. Apply Min-Max scaler, transforming and fitting of data and finally splitting data set into training and testing

vi. Defining functions to calculate accuracy and confusion matrix parameters.

vii. Implementation of model classifiers with calculating training, testing and validation error.

viii. Displaying confusion matrix, PR – curve and AUC.

ix. Calculate performance driven metrices (Precision, Recall, Accuracy, F1 score, FP rate and specificity).

x. Apply this for all other algorithms as well.

xi. Selecting one algorithm by considering all performance evaluation parameters.

Precision Recall Curve of Implemented models:

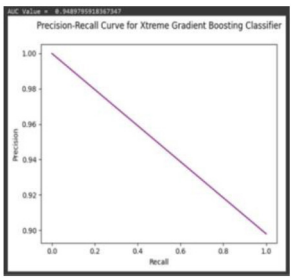

Figure 2

PR curve for Xtreme gradient boosting classifier

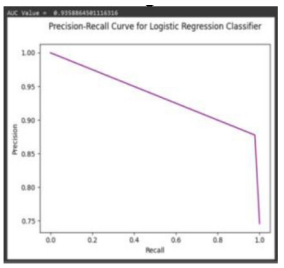

Figure 3

PR Curve for Logistic Regression Classifier

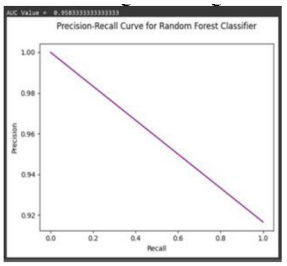

Figure 4

PR Curve for Random Forest Classifier.

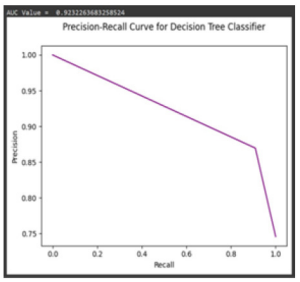

Figure 5

PR Curve for Simple Decision Tree Classifier

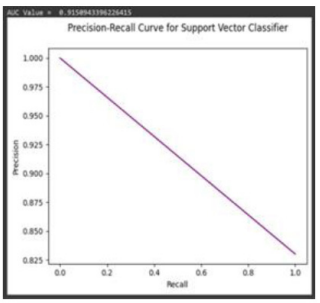

Figure 6

PR Curve for Support Vector Classifier

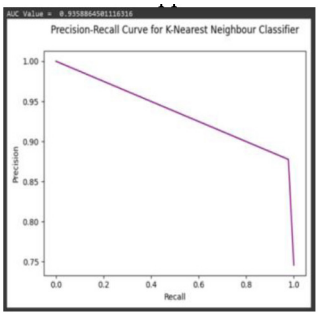

Figure 7

PR Curve for K- nearest neighbor Classifier

4. Results, Discussions and Findings

The results for the True Positive, False Positive, True Negative, False Negative values shown in below Table 1.

Table 1
Values for confusion matrix for different models

ML Algorithm	CM Parameter			
Xtreme Gradient Boosting	44	10	0	5
Logistic Regression	43	9	1	6
Random Forest	44	11	0	4
SVM	44	6	0	9
Simple Dicision Tree	40	9	4	6
K-NN	43	9	1	6

From Table 2, it is evident that Random forest Classifier has the highest value of True Positive that means this algorithm correctly classifies the patients of Parkinson's disease. Utilizing the values from Table 2, Precision, Recall, F1-Score, Accuracy, FP rate and specificity can be calculated. The results are tabulated in Table 2 and represented in Figure no. 2, 3, 4, 5, 6 and 7."

Table 2
Results Obtained By Various ML Models

ML Algorithm	PEPs					
	Accuracy	Precision	Recall/ TP Rate	F1 score	FP Rate	Specificity
Xtreme Gradient Boosting	91.52	1.00	0.89	0.95	0.00	1.00
Logistic Regression	88.13	0.97	0.88	0.92	0.10	0.90
Random Forest	93.22	1.00	0.92	0.96	0.00	1.00
SVM	84.75	1.00	0.83	0.91	0.00	1.00
Simple Decision Tree	83.05	0.91	0.87	0.89	0.31	0.69
K-NN	83.14	0.98	0.88	0.92	0.10	0.90

From Table 2 it is evident that Random forest algorithm has the highest accuracy of 93.2203% and f1 score of 0.9565 followed by Xtreme gradient boosting algorithm.

Accuracy : Random forest classifier yielded highest accuracy of 93.2203% followed by xtreme gradient boosting classifier with 91.5254% as Shown in Figure 8.

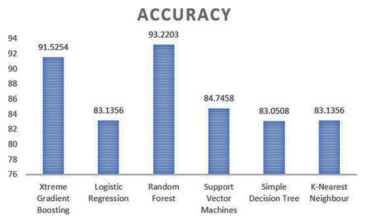

Figure 8

Comparison of accuracy

Precision: Random forest classifier, Xtreme boosting classifier and support vector classifier resulted in equal precision of 1.0.

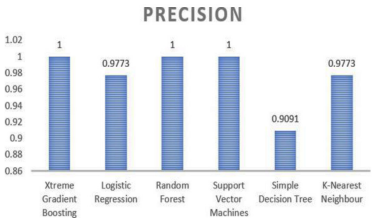

Figure 9

Comparison of precision

Recall/TP rate: Random forest classifier resulted highest TP rate of 0.9167 shown in Figure 10.

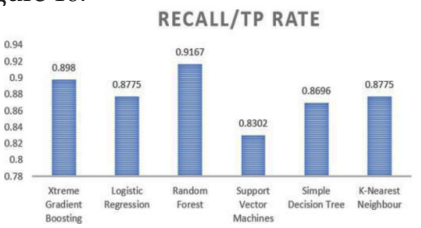

Figure 10

Comparison of Recall/TP rate

F1 Score : 0.9565 was the F1 score obtained using random forest classifer and is greater than every other implemented models (Figure 11).

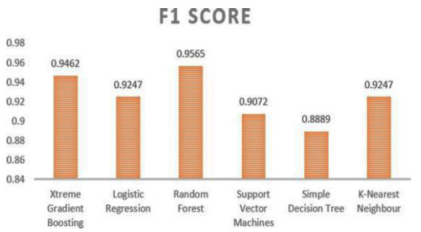

Figure 11

Comparison of F1 score

FP Rate : Nil false positive rate was obtained in random forest classifier, Xtreme boosting classifier and support vector classifier.

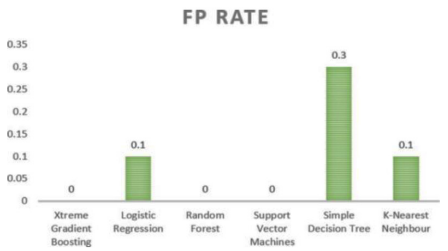

Figure 12

Comparison of FP rate

Specificity: Complete specificity was obtained in random forest classifier, Xtreme boosting classifier and support vector classifier

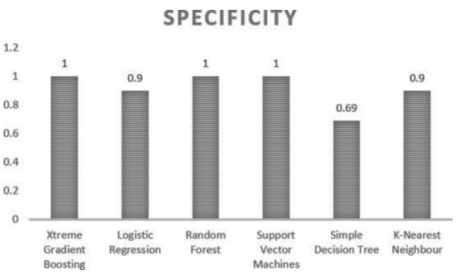

Figure 13

Comparison of specificity

Findings: In this study we intended to give equal importance to every performance evaluation parameters used. Results of precision, FP rate and specificity are the key findings of this study.

1. From Figure 9, It is relevant that complete precision has been obtained by three algorithms random forest, xtreme gradient boosting and support vector machine by using follow ups of error validation. From reference [2], highest precision recorded was 0.9655 by random forest classifier.
2. From Figure 12, FP rate was obtained which came out to be nil in same three algorithms (random forest, xtreme gradient boosting and support vector machine), Simple decision tree showed the worst FP rate.
3. From Figure 13, specificity was obtained for all six implemented algorithms, out of which random forest, xtreme gradient boosting and support vector machine resulted in complete specificity rate with simple decision tree to result in least specificity rate of 0.69.

5. Future scope and Conclusion

This paper demonstrates the ability of machine learning to accurately predict the presence of Parkinson's disease. The project also provided insight into features that could be used to predict disease. Current models use several noise parameters such as MVDP, jitter, and various frequency tuned sounds to predict the presence of Parkinson's disease. However, there may be other factors or characteristics associated with predicting disease. For example, research has shown that cognitive impairment is a symptom of Parkinson's disease, and knowing more about patterns can improve their accuracy. Another area of the future for the could be the development of higher standards and accuracy. Although the current model uses various machine learning algorithms, many advanced machine learning techniques such as deep learning and clustering can be explored to improve the performance of samples. In addition, the integrated method combining multiple models can be explored to improve the performance of the model. The development of time tracking can also be used for future research.. This will lead to early detection and timely intervention in patients with Parkinson's disease. In addition, the use of remote monitoring tools helps to provide personalized care and reduce the burden of healthcare services.

The "Performance Evaluation of Machine Learning Algorithms for Parkinson's Disease Identification Model" project aimed to predict whether a person is suffering from Parkinson's disease or not, using various supervised machine learning algorithms such as logistic regression, KNN, simple decision treeThe results showed that the random forest classifier achieved the highest accuracy of 93.22% , precision of 1.0, recall score of 0.9167, F1 score of 0.9565, FP rate of 0 and specificity of 1.0, making it the best algorithm to predict Parkinson's disease for the given dataset.

Based on the outcomes of this project, we conclude that the use random forest algorithm is a reliable and efficient method for predicting Parkinson's disease. The model can be used by medical practitioners to identify individuals with a high likelihood of developing Parkinson's disease and provide early intervention and treatment.

However, it is essential to note that this project's success was primarily dependent on the quality and quantity of data used for training the machine learning models. Therefore, future work should focus on obtaining more diverse and comprehensive datasets to improve the accuracy of Parkinson's disease diagnosis.

References

[1] Md. Toukir Ahmed, Md. Nazrul Islam Mondal, Mohammed Sowket Ali, Md. Moshaheb Hossain, Mahabuba, "Parkinsons Disease Detection Using Machine Learning Algorithm", IJRASET44340, 2022-06-15

[2] Shreevallabhadatta G, Suhas M S, Vignesh, Manoj C, Rudramurthy V C, Bhagyashri R Hanji, IRJET-V9I6322 June (2022).

[3] P. Raundale, C. Thosar and S. Rane, "Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm, 2021 2nd International Conference for Emerging Technology (INCET), pp. 1-5 (2021), doi: 10.1109/INCET51464.2021.9456292.

[4] Harshvardhan Tiwari, Shiji K Shridhar, Preeti V Patil, K R Sinchana and G Aishwarya, "Early Prediction of Parkinson Disease Using Machine Learning and Deep Learning Approaches" (2021).

[5] Nissar, Iqra, Waseem Ahmad Mir, and TawseefAyoub Shaikh. "Machine Learning Approaches for Detection and Diagnosis of Parkinson's Disease-A Review", 2021 7th International Conference

on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE (2021).

[6] Nagasubramanian, Gayathri, et al. "Parkinson data analysis and prediction system using multi- variant stacked auto encoder." IEEE Access 8 : 127004-127013 (2020).

[7] Wang, Wu, et al. "Early detection of Parkinson's disease using deep learning and machine learning", IEEE Access 8 : 147635-147646 (2020).

[8] Sharma, Vartika, et al. "A fast parkinson's disease prediction technique using PCA and artificial neural network." 2019 International conference on intelligent computing and control systems (ICCS). IEEE (2019).

[9] Patra, Amit Kumar, et al. "Prediction of Parkinson's disease using Ensemble Machine Learning classification from acoustic analysis. *Journal of Physics: Conference Series.* Vol. 1372. No. 1. IOP Publishing (2019).

[10] Aich, Satyabrata, et al. "A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of Parkinson's disease." 2019 21st International Conference on Advanced Communication Technology (ICACT). IEEE (2019).

[11] Sapkal, Vinod, Gupta, Praveen & Khan, Aboo Bakar. Deep learning based phishing website identification system using CNN-LSTM classifier, *Journal of Information and Optimization Sciences*, 44:3, 315–330 (2023). DOI: 10.47974/JIOS-1343.