IJIRIS: International Journal of Innovative Research in Information Security
Volume 10, Issue 02, May 2023
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Vocal Based Parkinson's Disease Detection

G Sai Pavani , Jemeema A
Student, Computer Science & Engineering,
Vemana Institute of Technology,
Bengaluru-34, India
Mallu Babu Reddy, Puli Ganesh Kumar Reddy
Student, Computer Science & Engineering,
Vemana Institute of Technology,
Bengaluru-34, India

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.09/Issue02/MYIS10080
Research Article | Open Access | Double-blind Peer-reviewed | Article ID: IJIRIS/RS/Vol.09/Issue02/MYIS10080
Received: 10, May 2023 | Accepted: 21, May 2023 | Published: 31, May 2023
Volume 2023 | Article ID MYIS10080 http://www.ijiris.com/volumes/Vol09/iss-02/01.MYIS10080.pdf
Article Citation: Sai,Jemeema,Mallu,Puli(2023).Vocal Based Parkinson Disease Detection. International Journal of Innovative Research in Information Security (IJIRIS), 10, 07-09. **doi:** https://doi.org/10.26562/ijiris.2023.v0902.01
BibTex key: sai2023vocal
Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar, IJIRIS, AM Publications, India

Abstract: James Parkinson described Parkinson's disease as a neurological syndrome, it affects the central nervous system as a result the patients face the issue in talking, strolling, tremor during motion. Parkinson's disease patient typically has a low-volume noise with a monotone quality; this system explores the classification of audio signals feature dataset to diagnosis Parkinson's disease (PD), the classifiers we utilized in this system are from Machine Learning. The significant classification models that are effectively utilized are (K-Nearest Neighbor) KNN, Decision Tree, Logistic regression and eXtreme Gradient Boost (XGboost). The performance analysis of the Parkinson's disease datasets has a variety of features. There are 21 features which play an important role for predicting the best algorithm in the classification. The system has achieved better result in predicting the PD patient healthy or not, XGboost provided the peak accuracy of 92% and the Matthews Correlation Coefficient (MCC) of 78%. MDVP: Jitter (%), Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DPP, MDVP: Shimmer, MDVP: Shimmer (db), Shimmer APQ3, Shimmer APQ5, MDVP: APQ, Shimmer: DDA, NHR are the selected features from the dataset, which are important for the prediction to achieved high accuracy.
Keywords: Parkinson's Disease, monotone quality, neurological syndrome, audio signals feature dataset.

I. INTRODUCTION

Machine learning is a field in computer science; it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or solve problem. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs. Speech or voice data is assumed to be 90% helpful to diagnose a person for identifying presence of disease. In general, Person with PD suffers from speech problems, Several strategies are recorded for early detection of PD based on the different ML techniques. But accuracy in detection and classifying within the time is very important or else, it causes development of more symptoms. There are different kinds of data, brain MRI images, Voice data, posture images, senor captured data, handwritten data, using which we can predict whether person is having PD or not. Out of all those, speech or voice data helps in identifying PD accurately. The Parkinson's disease detection system, based on machine learning, provides the public a platform, the performance analysis of the Parkinson's disease datasets has a variety of features. There are 21 features which play an important role for predicting the best algorithm in the classification. The system has achieved better result in predicting the PD patient healthy or not, XGboost provided the peak accuracy of 92% and the Matthews Correlation Coefficient (MCC) of 78%. MDVP: Jitter (%), Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DPP, MDVP: Shimmer, MDVP: Shimmer (db), Shimmer APQ3, Shimmer APQ5, MDVP: APQ, Shimmer: DDA, NHR are the selected features from the dataset, which are important for the prediction to achieved high accuracy.

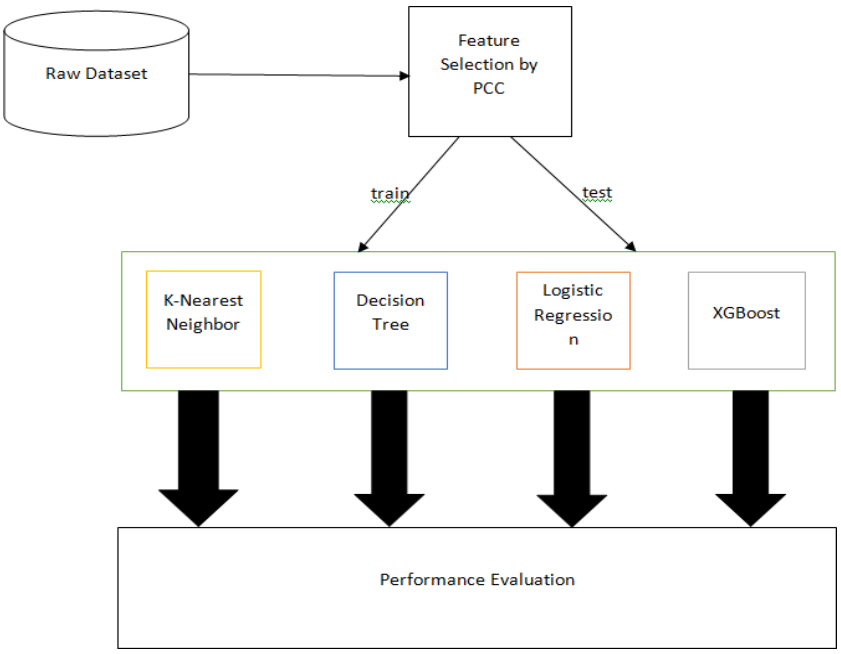
II. SCOPE

The scope of the project is to select the important features from the voice feature dataset which are very useful for detection and splitting the selected features dataset train and test for the classification. Here classification perform on four machine learning algorithms to determine which algorithm provides best result in detection, after training and testing the model, test the model with new dataset to predict the high accuracy and after evaluating all the performance of all four algorithms showing best algorithm for the PD dataset for detecting Parkinson's disease.

III. RELATED WORK

There have been several attempts and projects implemented with respect to multiple disease prediction. [1] the diagnosis of PD consists of three steps: 1) data pre-processing, 2) extraction of features and 3) classification. The speech signals are subjected to filtering so as to eliminate the noise present in the signals. From each segment, several features are extracted in the second step. In the last step, the classification is performed. The performance of the classification method has a strong impact on the method used for feature extraction. Hence, selecting the right classification model is an important issue that needs to be maneuvered in the diagnosis of PD. The models used are like SVM, DNN and KNN. [2] made use of a dataset similar in [1] for prediction of diseases uses the similar steps consisting of Data collection, Data analysis and visualization, Data preprocessing, Model selection and training, Evaluation of model, Comparison and selection of best model out of the employed XGboost and Neural network. [3] KNN and SVM classifiers yielded higher accuracy of 87.8% and 86.6%, respectively. However, k-NN achieved an outstanding Recall score and the fastest processing time. Nevertheless, even with or without the feature selection step, Light GBM still obtains the highest accuracy out of all the algorithms, which is 90.5%, when running the baseline model.

IV. PROPOSED METHODOLOGY



The architecture describes the complete process involved in this model. At first, select the important features from the voice feature dataset which are very useful for detection and splitting the selected features dataset train and test for the classification. Here classification perform on four machine learning algorithms to determine which algorithm provides best result in detection, after training and testing the model, test the model with new dataset to predict the high accuracy and after evaluating all the performance of all four algorithms showing best algorithm for the PD dataset for detecting Parkinson's disease.

2 Dataset

The dataset utilized for Parkinson's disease diagnosis system comprises of nearly 195 samples. It constitutes a total of 147 affected and 48 unaffected Parkinson's disease patients. Dataset created by Max Little of the University of Oxford, high prediction has beenachieved with classification accuracy, which we taken from UCI repository contains 21 features.

V. IMPLEMENTATION

The Parkinson's disease detection system has been implemented using the Jupyter IDE. Jupyter is open-source and cross platform environment. The Stream litframework has been used to run web-application. The programming language used is the Python programming language. The following machine learning algorithms were utilized as part of the multiple disease prediction system:

Decision Tree

Decision Tree Analysis is a general, predictive modeling tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that identifies ways to split a dataset based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

k nearest neighbor

K nearest neighbor is an approach to data classification that estimates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in. KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry.

logistic Regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

XGboost

XGboost is a boosting algorithm, it is statistical learning method and derived from gradient boosting decision tree, it has better performance and optimization. The reason why we used XGboost is it has good efficiency and feasibility, XGboost allows dense and sparse matrix as the input and a numeric vector uses integer starting from 0 for classification.

VI. RESULT

The accuracy obtained in classifying the respective Parkinson disease based on the entered dataset gives 92% for XGboost out of all the three machine learning algorithms that were utilized.

VII. CONCLUSION

For analysis, the common features considered are jittering, shimmering and voice signal frequency. Various algorithms for machine learning have been used in the prediction of Parkinson's disease. The range of prediction accuracy from applied algorithms lies between 75% and 92%, which demonstrates that machine learning algorithms can be used to predict Parkinson's disease by analyzing voice signals. Here the classification accuracy was studied and compared, with good performance and fast implementation, XGboost achieved a high accuracy with 92%. This system provides the comparison between machine learning classifiers of Logistic regression, Decision tree, K-Nearest neighbor and XGboost in Parkinson's disease diagnosis with high dimensional data MDVP: Jitter (%), Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DPP, MDVP: Shimmer, MDVP: Shimmer (db), Shimmer APQ3, Shimmer APQ5, MDVP: APQ, Shimmer: DDA, NHR are the selected features from the dataset to achieve high accuracy in prediction.

VIII. REFERENCES

[1] Iqra Nissar; Waseem Ahmad Mir; Izharuddin; Tawseef Ayoub Shaikh." Machine Learning Approaches for Detection and Diagnosis of Parkinson'sDisease". 2021 IEEE. https://doi.org/10.1109/icaccs51430.2021.9441885
[2] Pooja Raundale; Chetan Thosar; Shardul Rane." Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm".2021 IEEE. https://doi.org/10.1109/incet51464.2021.9456292
[3] Son V T Dao "An Analysis of Vocal Features for Parkinson's Disease Classification Using Evolutionary Algorithms" 2022.
[4] Mohammed S Islam "Parkinson's Disease Prediction Using Machine Learning" 2020.