Parkinson's Disease Detection Using Xgboost Classifier Machine Learning

Nandan BH[*1],Nisarga SC[*2],Prarthana[*3] ,Uday Kumar KP [*4]

Mrs. Nethravathi[*5]

[1,2,3,4]*Student,* [5]*Professor,*
[1,2,3,4]*Department of ECE, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, india*

ABSTRACT
Parkinson'sDisease(PD)isaprogressivedisordercentralnervoussystemthataffectingthevariousmovements. Symptoms are different Tremors are common, but the disorder is also commonly causes stiffnessand slowing of the movement. Using machine learning algorithm it's easy to use analyze the voice patternvariations to predicting the existence PD patients. This paper proposes the predictive model that effectivelydiagnosesPDwithmaximumaccuracyusingthedatasettoextrapolateddatafromvoicerecordingsofParkins on's patients. xgboost is a new machine learning algorithm to design with speed and performance inmind. xgboost stands for extreme gradient boosting and is based on decision tree. This algorithm is used topredictthe disease.
Keywords: *MachineLearningAlgorithmxgboost, ConvolutionNeuralNetwork,SupportVectorMachine.*

I.   INTRODUCTION

Parkinson'sDiseaseisachronic,progressivediseasewhichaffectsthemovementthroughoutthebody.Parkinso n'sDiseaseSymptomscanbedifferentofeveryone.TherearemanysymptomsofPD,includingtremors, muscle rigidity, changes in facial expressions, handwritten drawings. Current diagnostic tests for PDare limited. There is no standard diagnostic tests, such as blood test so a clinician's opinion is required toneurological test make a diagnose. Parkinson's Disease has 5 stage to it and affects the more than 1 millionindividuals every year in India . This is a chronic and has no cure yet. PD is a neurodegenerative affectingdopamine producing neurons in the brain. Parkinson's disease is the second most common neurodegenerativediseaseaffecting1%of thepopulation over 55yearsof age

MachineLearning

Machine Learning is the data science process that allows computer to gain information insight to the patternand existing data to predict the outcomes and trends of the data to programmed identified them. This can maketaskslikediagnosingParkinson'sdiseasemoreautomatedefficient,andaccuratewitharobustmachinelearning model and also identify patterns and characteristics in the data in ways that humans may not noticed.Machine learning algorithm to train data input and use statistical analysis in order to create specific output.Machinelearninghasvariousalgorithmtousedpredict thedisease.

II.   DESIGN AND IMPEMENTION

Methodology:
ThissectionexplainedthedifferentstepstoachievethepredictionofParkinson'sdiseaseusingvariousmachinelearning. DataGathering,Datapreprocessing,Modelselection,Training,Evaluation,prediction.

Figure1: WorkflowOfModel

DataGathering:
Thefirststepisdatagathering.Thisstepisveryimportantbecausethequalityandquantityofthedatayougatherwilldirectly affectsthelevel ofyour prediction model.

DataPreparation:
Inthisstepdataisvisualizedwelltospottherelationshipbetweentheparameterspresentinthedatasoastotakethe advantageof aswell astogetthe data.

ModelSelection:
In thismodelselectiontherearevariousmodelisused totilldatebyresearchandscientist. Somearemeantbyimageprocessing,someforsequencesliketext, numbersorpatterns.

Training:
Trainingthedatasetisthemaintaskofmachinelearning.Wewillapplythedatatoprogressivelyimprovetheselectedmode ltopredictthebetteractualresultshould beapprox.topredict one.

Evaluation:
ThemetricswehavecalculatedareROC,Accuracy,Specificity,Precisionetc.whichwillhighlightsthebestalgorithmam ongall.

Prediction:
InthisphasewefinallygetthemodelreadytodetectthepredictionofParkinson'sdiseasebasedonthegivendataset.

Algorithm:
XGBoost(ExtreamGradientBoosting):
XGBoost is an optimized distributed gradient boosting library designed tobehighly efficient, flexible and portable.Itis used inprediction problemsinvolving unstructureddata (images,text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. XGBoost provides aparallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurateway.It implementsmachinelearning algorithmsundertheGradientBoostingframework.
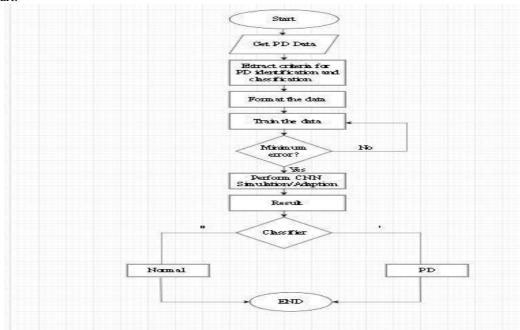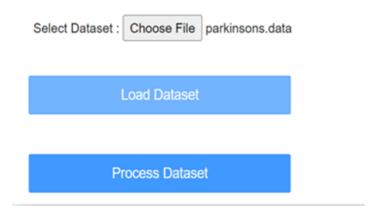
Flowchart:

Figure2:Flowchart

III. RESULTSANDDISCUSSION

Initially the parkinson's dataset was loaded just by tapping on the choose file which is shown in the figure below by that the required dataset to be in the format of csv because it should read the data which is the format of csv file then after that next procedure is to load the dataset into the program which is co-merged with the web application which designed so that the data will be loaded to the machine learning program then the next process is to process the dataset which is included in the web application by tapping the data will process as per the code. After the processing of data the output will be in the format of 1 and 0's which one include the patient has Parkinson disease and also zero indicating that healthy person which process by combining with xgboost algorithm which helps to predict then finally the predicted will be vomit as a result along with patients data with the serial number.

Performance analysis

Here the evaluation of model can be done so that it helps to know about the accuracy of the model and this model has the accuracy of around 94.87 percentage. which is evaluating with respect to true labels with the predicted models. Which process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Model evaluation is important to assess the efficiency of a model during initial research phases, and it also plays a role in model monitoring.

```
[26] acc_score = accuracy_score(Y_test, Y_predict)*100

print(f"Accuracy_score: {acc_score}%")

Accuracy_score: 94.87179487179486%
```

IV. Conclusion

This project aims to provide a better prediction for Parkinson's disease by optimizing and tuning the parameters of KNN, SVM and Random Forest Algorithms. This project also provides a reliable health monitoring system for those suffering from Parkinson's disease. The proposed web application also provides areal time appointment and feedback system which helps in gathering valuable data which aids the current research to find a cure for Parkinson's disease.

The present study is the first review which included results from all studies that applied machine learning methods to the diagnosis of PD. Here, we presented included studies in a high-level summary, providing access to information including (a) machine learning methods that have been used in the diagnosis of PD and associated outcomes, (b) types of clinical, behavioural and biometric data that could be used for rendering more accurate diagnoses, (c) potential biomarkers for assisting clinical decision making, and (d) other highly relevant information, including databases that could be used to enlarge and enrich smaller datasets. In summary, realization of machine learning-assisted diagnosis of PD yields high potential for a more systematic clinical decision-making system, while adaptation of novel biomarkers may give rise to easier access to PD diagnosis at an earlier stage. Machine learning approaches therefore have the potential to provide clinicians with additional tools to screen, detect or diagnose PD.

Reference

[1].   B. E. Sakar et al., "Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings." in IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4. pp. 828-834, July 2013, doi: 10.1109/JBHI.2013.2245674

[2].   Benba, A., Jilbab, A. &Hammouch, A. Analysis of multiple types of voice recordings incepstral domain using MFCC for discriminating between patients with Parkinson's diseaseand healthy people. Int J Speech Technol 19, 449-456 (2016)

[3].   Ö. Eskidere. A. Karatutlu and C. Ünal, "Detection of Parkinson's disease from vocal features using random subspace classifier ensemble." 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO), Almaty. 2015, pp. 1-4. doi: 10.1109/ICECCO.2015.7416886

[4].   Nissar, Iqra& Rizvi, Danish & Masood, Sarfaraz & Mir. Aqib. (2018), Voice-Based Detectionof Parkinson's Disease through Ensemble Machine Learning Approach: A Performance Study.EAI Endorsed Transactions on Pervasive Health and Technology.162806. 10.4108/cai.137-2018.162806

[5].   Y. Li, C. Zhang, Y. Jia, P. Wang, X. Zhang and T. Xie. "Simultaneous learning of speech feature and segment for classification of Parkinson disease," 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom). Dalian. 2017,pp. 1-6, doi: 10.1109/HealthCom.2017.8210820

[6].   Caliskan, Abdullah &Badem, Hasan &Basturk, Alper&Yüksel, Meltem. (2017). Diagnosis of the Parkinson disease by using deep neural network classifier. Istanbul University - Journalof Electrical and Electronics Engineering. 17.3311-3318.

[7].   Berus L. Klancnik S, Brezocnik M. Ficko M. Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks. Sensors (Basel). 2018:19(1):16. Published 2018 Dec 20. doi:10.3390/s19010016