Parkinson's Prediction Techniques: A Comparison

Agrajalenin, Suniyya PP, AryaIndraja, DhanyaS, NavneethaKA

Dept. of Biomedical Engineering KMCT College of Engineering for Women NIT Campus post, Kalanthode, Kozhikode,
Dept. Of Biomedical Engineering KMCT college of engineering for women NIT Campus post Kalanthode Kozhikode,
Assistant professor, Dept.of Biomedical Engineering KMCT college of engineering for women NIT campus post Kalanthode, Kozhikode,
Dept. of Biomedical Engineering KMCT College of Engineering for Women NIT Campus post, Kalanthode, Kozhikode, Kerala
Dept. Of Biomedical Engineering KMCT College of Engineering for Women NIT Campus post Kalanthode Kozhikode, Kerala

ABSTRACT-- Abstract--Parkinson's disease is a medicalcondition of the central nervous system that affectsmotion , often including quivering.Nerve cell damageinthebraincausesdopaminelevelstodrop,leadingtothe symptoms of Parkinson's. Which include Tremorin the hands, limbs, stiff muscles, difficulty standing,difficulty walking, difficulty with bodily movements,unintentional movements, muscle become unbending,problemswithintegration,throbbimgmusclecontractions,gentlybodilymovement,orslowshuffling gait early awakening, nightmares, restlesssleep,fatigue,dementia,ordifficultythinkingandsmall handwriting, trembling, unintentional writhing,orweightloss.itmayalsosymptomsofotherdisease.Thus, we aim to use comparison Techniques(Decision Tree, SVM, NB) with some preprocessingtechniquestopredict theParkinson's disease

Keywords—SVM—Support Vector Machine -NB-Naivebye

I. INTRODUCTION

Parkinson's disease is a medical condition in whichcentral nervous system that affects the motion, oftenincluding quivering. Nerve cell damage in the braincausesdopaminelevelstodrop,leadingtothesymptoms of Parkinson's. Which include Tremor inthe hands, limbs, stiff muscles, difficulty standing,difficultywalking,difficultywithbodilymov ements,involuntarymovements,musclerigidity,problemswithcoordination,rhythmicmusclecontractions,slowbodilymovement,orslow shuffling gait early awakening, nightmares, restlesssleep, fatigue, dementia, or difficulty thinking andunderstanding, difficulty speaking, soft speech, orvoice box spasms fear of falling, loss in contrastsensitivity,necktightness,smallhandwriting,trembling,unintentionalwriting,orweightloss.Advancements in science and technology has madethe process of accessing and understanding datasetsmuch easier .Multiple datasets based on Parkinson'sdisease are also openly available.This paper dealswith mainly three PD datasets which are voice PD,handPDandspeechPDdataset.Anumberoffeaturesaretakenintoconsiderationforcategorizingthepeople intoaffectedorhealthy.Thedatasetrelatedto the different features cancontribute toincrease the accuracy level of the detection process.But in the past few decades it was observed that thevolumeofthedatasetshasbeenincreasingdrastically and has led to the drop in performance ofvarious models, thereby degrading theresults. Thus,it is essential to process the data by choosing only asubsetoftherelevantfeatures.Recently,thepowerofoptimization algorithms is being used in advancedfeature selection methods, to produce results withhigheraccuracybyremovingnotpertinentandneedless attributes from datasets, whichdoes notmakeabighandouttotheaccuracyofamodel.Grey WolfOptimizationisonesuchoptimizationtechnique

which is used for feature selection .Thereducedsubsetoffeaturesobtainedafterfeatureselection is further processed by a machine learningalgorithmforproducingafinalpredictionondiagnosis of Parkinson's disease. Here in this paperdealswithadeeperstudyonthedifferenttypeof prediction techniques like SVM, Naive byase ,anddecision tree.The process of disease prediction isdivided into two steps feature selection and diseaseprediction.Theprocessofselectingfeaturesisa veryimportanttaskasitplaysacrucialroleindetermining goodtimecomplexityaswellasproducingbetteraccuracyforthemodelthereybyincreasing the performance. The total set of featuresare examined closely for selecting a reduced subsetof features, this paper uses GWO for the same.Thenext major step is disease prediction. The reducedsubset of features is further provided as input tovarious machine learning algorithms like svm, navebiase, Decision tree etc to produce an accurate andconcrete result. It also focuses on comparing themethodsofpredictionsinthecomingsections.

II. LITERATURE SURVEY

Fordiseasepredictionweusedifferenttypesofclassificationalgorithms,thesearemostcommoninmanyhealthcares.Manyoftheseshowshighclassification accuracy;

1) Dr.AnupamBhatiaand Raunak Sulekh ,"PredictiveModelforParkinson'sDiseasethroughNaive Bayes Classification" In this study, Naive Bayeswas applied to predict the performance ofthe dataset. Rapid miner 7.6.001 is a tool,whichisusedtoexplore,statisticallyanalyze, and the data. The Naive Bayesmodelperformswith98.5%accuracy,and99.75%of precision.

2) MehrbakhshNilashietal,"Ahybridintelligent system for the prediction ofParkinson'sDiseaseprogressionusingMachineLearningtechniques"Inthissystemamethodwasproposed fortheUPDRS (Total-UPDRS andMotorUPDRS)predictionusingmachine learning.ISVRisusedtopredicttheTotal-UPDRSandMotor-UPDRS.SOMandNIPALSwereusedforclusteringandddatadimensionalityreduction.TheresultsshowthatthemethodcombiningSOM,NIPALS,andISVRtechniqueswaseffectiveinpredictingtheTotal-UPDRSandMotor-UPDRS.

3) M.Abdar and M. Zomorodi-Moghadam , "Impact of Patients' Gender onParkinson'sdiseaseusingClassificationAlgorithms"Inthissystem,SVMandBayesian

Networks were used forclassificationofdatabasedonthegenderof the patient. The accuracy for SVM was90.98% and Bayesian network was88.62%.ThistestprovedthattheSVM algorithm had a great ability to identify apatient'sgendersufferingfromPD.6)DraganaMiljkovicetal,"MachineLearningandDataMiningMethodsforManagingParkinson'sDisease"Inthissystem,based ontheinitialpatientsexaminationandmedicationstaken,thePredictorpartwasabletopredicteachParkinson's Disease symptom separatelycovering 15 different Parkinson's Diseasesymptomsintotal.Theaccuracyofpredictionrangesfrom57.1%to77.4%dependingonthesymptom wherethehighest accuracy is achieved from tremordetection.

4) Md. Redone Hassan et al[, "A KnowledgeBase Data Mining based on Parkinson'sDisease"Inthissystem,theresultsandoutputofthevectorsupportmachine(SVM), K nearest neighbor and the outputfiguresforthedecisiontreealgorithmswere shown in the output section of thetraindata. The decisiontree offeredthehighestprecisionof 78.2%

5) SatishSrinivasan,MichaelMartin& AbhishekTripathi,"ANNbasedDataMining Analysis of Parkinson's Disease"Inthisstudy,itwasintendedtounderstandhow the different types of pre-processingsteps could affect the prediction accuracyoftheclassifier.IntheprocessofclassifyingtheParkinson'sDiseasedatasetusingtheANNbasedMLPclassifierasignificantly high prediction accuracy wasobservedwhenthedatasetwaspreprocessed using both the Discretizationand Resample technique, both in the caseof 10fold cross validation and 80:20 split.Whereasinthe70:30splititwasfoundthatthe combination of the pre-processing stepsnamely Resampling and SMOTE on thedatasetresultedtowardsthehigherpredictionaccuracyusingtheMLPclassifier.Onan80:20splitofthepreprocessed (Discretized and Resampled)datasettheANNbasedMLPclassifierachieveda 100% classificationaccuracywith F1-score andMCCbeing100%.

6) RamziM.Sadeketal,"Parkinson'sDisease Prediction using Artificial NeuralNetwork" In this system, 195 samples inthe dataset were divided into 170 trainingsamples and 25 validating samples. Thenimportingthe dataset inthe Just NeuralNetwork (JNN) environment, we trained,validatedtheArtificialNeuralNetwork

model.Themostimportantattributescontributing to the ANN model were madeknownof.TheANNmodelwas100%accurate.

III. METHODOLOGY

Data Mining isa process of extracting usable datafrom a much larger set of any raw data. It meansstudying data patterns in large batches of data usingoneormoresoftware.Datamininghasimplementationfieldsofscienceandresearch.Asanimplementation of data mining, medical science canlearnmore aboutthe diseasesanddevelopmoreeffective Master planto combat those diseases andin turn Support resources in a more optimal andinsightfulmanner.Dataminingnecessitateddatacollectionandwarehousingaswellascomputerprocessing. For segmenting the data and evaluatingthe probability of future events, data mining usessophisticated mathematicalalgorithms.

3.1   DATAMININGTECHNIQUES
A)   Association

Association is used to make an association betweenitemsandisoftenusedtostudyingsalestransactions.Theattemptsofassociationdatamining to establish a relationship between an itemthatoccurssortoutinagivendataset.Indatamining, association rules are used for studying andguessingthemedicalhealthpredictiontogetabetter diagnosis.

B)   Classification

Classification is a model used to predict the futurebehaviour of the data through classifying the recordsintopredefinedclasses.Theclassificationalgorithmis measured in terms of exactness and recall metricstoestimatetheperformanceofclassificationalgorithm.

Therearevarious dataminingclassifierssomeofthem are listedbelow:

- NaiveBayes

NaiveBayesinthehugedatasetpresentedacceptablespeedandaccuracy,buttheeffectisextremely unfortunate in the case of a small dataset.TheNaiveBayesclassifieristheprobabilistical gorithm that calculates a set of probabilities bycounting the frequency and groupings of values in agiven record.

- SupportVectorMachine

TheSupportVectorMachine(SVM)wasfirstformed by Vapnik and has since involved a highgradeofconcentrationinmachinelearning.SupportVectorMachineisaconstantalgorithmcomparedtoother algorithms that are neural networks, decisiontrees.

- DecisionTree

Decisiontreesarethemostforcibleandpopulartoolfor classification and prediction. A Decision tree isaflowchartliketreestructure,whereeachinternal node denotes a test onanattribute, each branchrepresents an outcome of the test, and sdfghl nodeholdsa classlabel.

3.2   PROPOSEDSYSYTEM
The proposed system includes a training and testingprediction modelthrough followingsteps;
A)   DataCollection
A data set (or dataset) is a collection of data. In thecaseoftabulardata,adatasetcorrespondstooneormoredatabasetables,whereeverycolumnofatablerepresentsaparticularvariable,andeachrowcorresponds to a given record of the data set inquestion.ThedatasetcanbecollectedfromKaggle ,Googledatasearch etc.
B)   Pre-processing
Donebynumpy,NumPyisaPythonlibraryusedforworkingwitharrays.Italsohasfunctionsforworkingindomainoflinearalgebra,Fouriertransform,andmatrices.Iti sanopen-sourceprojectand youcanuse itfreely.
C)   Featureextraction
Features are extracted from the dataset by followingways:
Thefit()methodtakesthetrainingdataasarguments, which can be one array in the case ofunsupervisedlearning,ortwoarraysinthecaseof supervised learning. Note that the model is fittedusing X andy,andoneincludrX andy.
Thefeaturesofeachsegmentedregionsareextractedby taking visual content of image. System workswell on the most relevant statistical parameters ortexture features computed by using Gray Level Co-occurrenceMatrix(GLCM).Thenconstructafeature vector to represents all feature values includecontrast,velocity,andentropy,energy etc.
D)   Output
Predictivemodellingusesscientificallyprovedmathematical statistics to predict events outcomes.Most often one event that a mathematician wants topredict or apply predictive analysis on it is in thefuture(alsoherephysicsandmathematicalnotionof future can be applied), but predictive modelling canbe applied to any type of mathematically stated as"unknown" event, (almost) regardless of when itoccurred.

Fig3.1Showsconfusionmatrix

Andcomparedtheoutputtofindtheaccuracy,usingconfusion matrix. A confusion matrix is a table thatisoftenusedtodescribetheperformanceofaclassification model (or "classifier") on a set of testdataforwhichthetruevaluesareknown.Theconfusionmatrixitselfisrelativelysimpletounderstand,butthe relatedterminologycanbeconfusing.

IV. RESULT AND DISCUSSION

Parkinson's Disease is a very grave disease and hasno cure till are.following it affects the gesticulationof the parts of the body, the speech also be foundaffected. Here, the system strive to provide a way ofdiscover Parkinson's Disease which will result in aquickactiontominimizeorevendelayitfromaffecting the complete body. This system intent tomakethisprocessofupprehensionacaseofParkinson 'sattheearliestbyboth,thepatientaswellas medical professionals. Hence, the aim is to usevarious data mining techniques like SVM, DecisionTree,Naivebiasforgettingthemostaccurater esult.Hereusing the techniquesaccuracyisimproved. $A=\pi r^2$
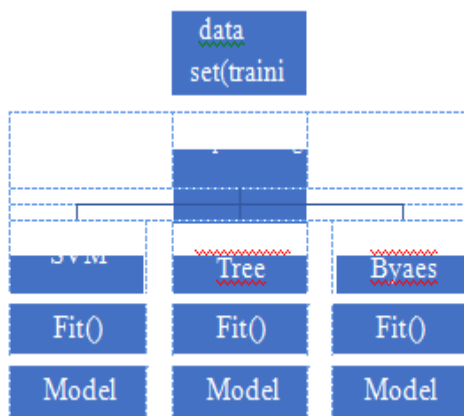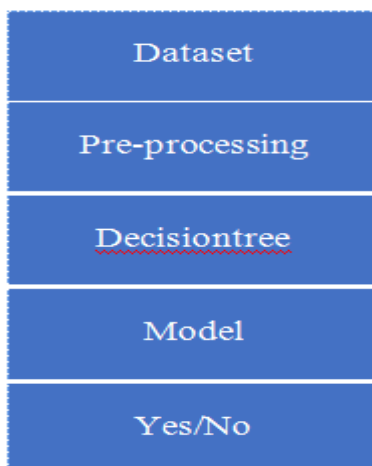

Fig3.2 showstrainingmodelofPDprediction

Fig3.3 TestingmodelforPDprediction