DESIGN AND DEVELOPMENT OF AN ALGORITHM TODETECTANDDIAGNOSEPARKINSON'SDISEASE

SingaReddyBhagyaLakshmi[1],SiriSGowda[2],TrishaA[3]

[1,2,3]*Student, DeptofElectronicsandCommunication,R.V. CollegeofEngineering,Bangalore,INDIA*

[1]*singareddybl.ec17@rvce.edu.in,*[2]*sirisgowda.ec17@rvce.edu.in,* [3]*trishaa.ec17@rvce.edu.in*

Abstract: *Parkinson's disease is neurological disorder that affects the neurons that produce a chemical substance known as dopamine. The symptoms of this disease are tremors, stiffness all around the body, vocal impairment. If not treated early this disease might even lead to death. This disease doesn't have a treatment,but early prediction might help in the reduction of the progress of the disease. Detecting the disease is muchmore difficult as there is no quantitative test that can be conducted to detect the disease. Voice is one of the primary symptoms of this disease and therefore the features present in the voice can be extracted and can beused to train a model to detect whether the person is suffering from the disease or not. An ensembled learning voting classifier algorithm is designed with a performance accuracy of more than ninety percent and is used for the prediction of the disease using the vocal features extracted from the person. This algorithm is trainedwith the dataset which contains both normal as well as the affected person's voice features. Decision tree classifier algorithm, Logistic Regression and that of the Support Vector Machine Algorithm are used as the input for the voting classifier which is used for detecting the disease.*

Keywords: **Support VectorMachine,VotingClassifier,Parkinson'sDisease,Ensemblelearner**

1. INTRODUCTION

Parkinson's disease is a neural dysfunction that influence the motion of human body.Symptoms are seen slowly, sometimes starting with a hardly noticeable shaking in onehand. Shakiness is common, but the dysfunction is also associated with a constrained movement.The face may show little or no expression in the beginning stage soft the disease.The arm smay not show much movement while walking. Also, the stammering may increase, the voice might become soft. The symptoms become more severe as the disease advances.

This disease causes some specific neurons in the brain to slowly die.Most of the symptom sare caused by the affinity of neurons that are responsible for producing a chemical called dopamine. An unusual brain malfunction happens which result in this symptom when this chemical's level falls.It has been estimated that there are 7 million people in India who are affected by this disease There is currently no blood or laboratory test for this disease. Thediagnose is done by looking at the family history and reviewing the signs and symptoms shown by the patient.

This kind of diagnosis is not very efficient and sometimes this results in the ignorance of the early sign softhe disease.It has been confirmed that Parkinson's Disease can affect the vocal ability of the patient. The speech of diseased patient has change in thefrequency specterintheirvoicebecausetheylossthecontrolofthelimb,whichdecreasethefrequencyoftheaudio.Therefore, themaingoalhereistodetectthediseasebyconsideringspeechastheparameter.Therearemanyparametersofthevocalthatistakenintoconsiderationhere,they are Jitter, Shimmer, Harmonic to Noise ratio and Noise to Harmonic ratio. A total ofsixteenparametersaretakenintoconsiderandthedataset that isutilizedtotrainthemodelistakenfromUCI MachineLearningRepository.

2. LITERATUREREVIEW

The problem of attribute election in case of Parkinson's Disease mechanized recognitionusingadaptive-basedmethodsismentionedinthispaper[1],wherethepartobehighlightedistheaccuracyoveranevaluatingset.Allthesemethodsincreasedtheoutcomesoverthestandarddataset,accordingtothe experiments.

In this paper [2], Convolutional Neural Network were used to investigate Parkinson'sDiseaserecognitionfromavocalwave.Spectrogramsandavarietyofotherattributes wereused as input for Convolutional Neural Network. The influence of each segment on theParkinson's Disease detection output was evaluated and compared to the decision levelfusion of all segments case. One segment had a detection performance of twenty-ninepercent, while the other segment had a detection performance of twenty percent. Thissuggests that some aspects of these recordings are more efficient than others in detectingParkinson'sDisease

In this paper, the author gathered a wide range of vocal samples and different sounds fromPersonwiththediseasetalkingexercises[3].
Itprovidesachancetoexplorethevalidityofalready present models. As a result of the analysis of the dataset, Sustained vowels werefound to have more Parkinson's Disease selective data than separated words, which is inline with the findings published in the literature discussed in this paper. They found thatrepresenting a subject's samples with the mean and Standard Deviation are better thanothers.Thismethodof representationseemedtobe moreprecise.

In this paper, the author explained Parkinson's Disease using various Machine Learningand Deep Learning algorithms to differentiate people into the groups ofhealthy andParkinson'sDiseaseaffectedbasedonnumeroussignsinordertocomeupwithaneffectivewaytodiagnoseParkinson'sDisease[4].Theoutcomesofnumerousstudieswerecontrastedusingvariousmethods,itwasdeterminedthatDeepLearningisthebestmethodforstudyingtwomainsymptoms:twistedwalkingstyleandspeechdisability.Thedatawascollected from the UCIrvine Machine Learning Repository.Twomodules,VascularEndothelialGrowthFactorSpectrogramDetectorusingConvolutionalNeuralNetworkandVoicedisfigurementusingArtificialNeuralNetwork,havebeenimplementedtodifferentiate PD patients based on the signs gait and speech disability, with an accuracy ofeighty eight percent and eight nine percent for the two modules on the testing dataset,respectively,andcomparedwiththreealgorithms, ExtremeGradientBoost,SupportVectorMachineandMultilayerperceptron.

Employingauto-immunevoice biomarkersasattributes,automatedmachinelearningarchitectures candiagnoseandpredictthedisease [5].Theresearchpresentedinthispapercompares the success of different Machine Learning classifiers in disease detection withnoisyandmulti-scaledata.Clinicallevelprecisionispossibleaftercarefulfeatureselection.Thesefindingsareencouragingbecausetheycouldpavethewayfornewwaystousevoicedatatoevaluatepatientwellbeingandneurologicaldiseases.

Inthis paper,theauthorsinvestigatedhowtodifferentiatePersonwithParkinson'sdiseasefromnormalpeopleusingvowelphonationsusingawidevarietyofoldandnewalgorithmsfortesting[6].Inrecentyears,thisbinarydiscriminationissuehaspiquedinterest,withthebest resultsshowingaclassificationaccuracyofaroundninetythreepercent onasubsetoftwenty-twofeatures.Theyshowedthatusingtendefectivemeasures,wecanachievenearlyninety nine percent accuracy. They used a speech data and added several newly suggesteddefects of speech steps that had never been used in this application before. They alsoexplored Radio frequency. They see this research as a first step towards a bigger goal ofdevelopingtechnologiesfortreatmentplaninParkinson'sDisease.Onthebasisofcomprehensive Curriculum vitae tests, the algorithms used in this study tend to be verysuccessfulatdistinguishing diseasedfrompeoplewhoarenotaffectedbyit.

This research looked at a stage-variant, medication-variant category of diseased patients, suggesting that the findings apply to Parkinson's Disease patients at all stages of the disease .

[7].Severalaspects ofspeechdeteriorationhavebeenlinkedtodiseaseprogressionamongpeople with Parkinson's Disease. This paper shows whether nonlinear dynamic studiesreveal a connection between phonatory pathology and the seriousnessofParkinson'sDisease.Overall,thefindingsindicatethatnonlineardynamicanalyse smaybeausefulnewapproach for studying Parkinson's vocal pathology, complementing conventional voiceanalysismethods.

In this paper the mentioned procedure is to try and gain a comparatively better accord andconsistencyintheprocessforpracticalassessmentofpathologicsounds[8].Becauseofthewi de range of methods used to determine functional effects, systematic analysis of theoutcomesofvoicetreatmentsareusuallyconstrained,ifnotimpossible.Amultidimensional set of basic computation is suggested that can be used to diagnose allcommondysphonia.Perception,videostroboscopy,acoustics,aerodynamics,andsubjective patient rating are the five different approaches. Instrumentation is held to aminimum, but it is considered essential for phono surgery professionals. A skilled andcertifiedspeechtherapistwillassisttheEar-Nose-Tonguespecialistsurgeoninconductingthissimpleseriesofmeasurements.

In this paper [9], the author obtained a number of sound tracks from various subjectsutteringthevowelatoimprovetheassessmentofParkinson'sDisease.Manyframesinthe extracted Mel Frequency Cepstral Coefficient from different participants take up the mostprocessingtimeduringtheclassificationprocess,preventingaccuraterecognition.

In this paper, numerous voice track refining algorithms were used to draw out informationfor Parkinson's Disease evaluation in this report, and the obtained attributes were given tothe algorithmsto build accurate resolution supportstructures.Forfeature extraction,TunableQ-factorwavelettransformwasappliedtothevoicetracksof Parkinson'sDiseasepatients [10]. They collected voice recordings from two hundred and fifty-two people anduprooted various characteristic subcomponents from the tracks. Multiple classifiers aregiventotheattributesubcomponents,andtheclassifierprognosisaremergedwithensemble learning techniques. The result of theirthesis shows that Tunable Q-factorwavelettransformisthebetterforPDdiagnosis

3. COMPARATIVEANALYSISOFCLASSIFIERALGORITHMS

The classification algorithms that are taken into consideration to study how they performfor the inputted dataset are Decision Tree Classifier, Support vector Machine, LogisticRegression,NaïveBayesandKNearestNeighboralgorithms.Thesealgorithmsaredesi gned using a software application called as Google Colaboratory. The dataset that istaken from UCI ML Repository,isfirst pre-processed and then isinputted into thisalgorithmtoobtaintheperformanceaccuracy.Sixteenparametersaretakenintoconsideratio nforpredictionpurpose. Theaccuracyvaluethat wasobtainedisshownintheTable1.

Table1.Comparisonofclassifieralgorithmswi threspect toaccuracy

ClassifierAlgorithms	Performance Accuracy(inpercentage)
SupportVectorMachineAlgorithm	89.79
DecisionTreeClassifierAlgorithm	89.79
LogisticRegressionAlgorithm	81.86
Naïve Bayes Algorithm	63.26
KNearestNeighborAlgorithm	80.93

Fromthisitcanbededucedthatforthisparticulardatasetsupportvectormachine,decisiontree classifier algorithm and logistic regression algorithm have outperformed Naïve Bayesand K Nearest Neighbor Algorithm. Therefore, these three algorithms will be used withensemblelearningvotingclassifieralgorithminordertocreateanalgorithmwhichis muchmoreprecise indetectingthedisease withthe givendataset

4. PERFOMANCECOMPARISONOFENSEMBLELEARNERS

Multiple algorithms are combined together and each of their decisions are taken intoconsideration while making the final prediction. Because a multiple combined model willdefinitely perform better than a single classifier algorithm. Diversification always leads tobettermentinaccuracyandpredictionandsamegoesformachinelearningmodels.Diverseset of classifier algorithms taken into consideration to make a stronger model where theerror of wrong prediction is almost zero to none. Three ensemble learners are taken intoconsiderationforthecomparisonpurpose.

Random forest is a bagging technique which has decision tree classifier algorithm as itsbase estimator. AdaBoost algorithm uses Support Vector Machine algorithm as its baselearner and executes the dataset. XGBoost algorithm is also known as Extreme GradientBoostingalgorithmisoneofthemostefficientboostingalgorithmspresentsatthemome nt.Voting classifier involves summing up of the predictions made by the weak learner andtake the average of the said algorithms. For hard voting, majority of the votes is taken intoconsideration.

Voting, Bagging and Boosting are the three types of ensemble learners present. All thealgorithm with their respective and suitable base learner is applied to the dataset and theaccuracy is obtained. This accuracy value shows the performance efficiency of the saidalgorithm.

It is shown in the below Table 2. It can be inferred from the table that Voting classifier isthe most efficient performance wise. Therefore, this algorithm along with its base learner,thatisDecisionTreeClassifier,SupportVectorMachineandLogisticRegressionischose nfor integration with the front end in order to predict whether the person is affected by thediseaseornot.

Table2. Performanceefficiencyofdifferente nsemblelearneralgorithms

EnsembleLearnerAlgorithm	Accuracy(inpercentage)
RandomForest Algorithm	90
AdaBoostAlgorithm	77.55
XGBoost Algorithm	88.59
VotingClassifierAlgorithm	95

5. IMPLEMENTATION

Thecompletedesigningofthealgorithmwas doneusingthesoftwareapplicationcalledasGoogle Colaborative which is provided by the technology company named Google. Thealgorithmwas designedusingthreeclassificationalgorithmwhich werethenensembledtothevotingclassifierwhichpredictstheoutputdependinguponmajorityoft heoutputgivenbytheweakclassifieralgorithm.Allthesethreeclassifiersareconsideredasweakl earners.These three classifier algorithms are combined and are put with theensemble votingclassier.Thevotingmethodthatisusedhereiscalledasthehardvoting.

The dataset that is obtained from UCI ML Repository has the voice measurement, a totalof thirty-one people, out of which twenty-three were affected by the disease. Each columnrepresents a feature of the voice extracted. The last column represents whether the personis affected by the disease or not. If the last column is zero, it means normal person. If thelastcolumnisone,thenitmeansthepersonisaffectedbythedisease.

Data pre-processing is done in order to scale the values of the dataset and to fill in

the missing values if any is present within the file. Importing of the required the libraries are done.Later the entire dataset is divided into testing and training set.Normally itwill be

eighty percent for training set and the remaining is for the test set. The training set isinputted to the model in order to make it understand how exactly the prediction has to bedone.Oncethetrainingisdone,thetestdatasetisinputtedandtheoutputofthemodel,thatis the predicted value is compared with the known outcome of the test dataset and theperformance accuracy value of the algorithm is obtain. The confusion matrix is alsoobtainingtounderstandindetailonhowexactlythedesignedalgorithmisperformingwiththe values given.Theentiremethodologyofalgorithmdesigningisshownin theFigure1.

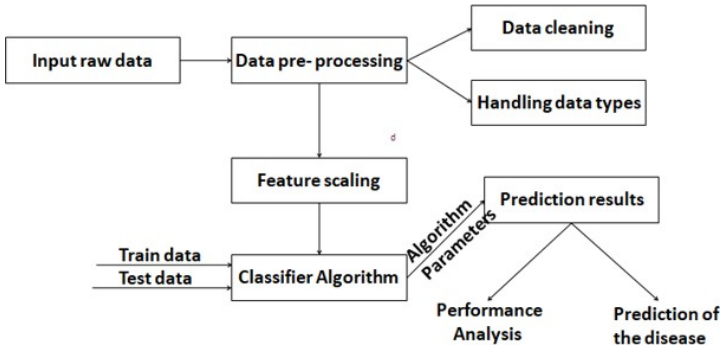

Figure1.Methodology

ThisalgorithmafterthecompletedesignisimplementedinanapplicationcalledasSpyderIDE. Later flask is used for creating a web application which can input the data. The sameapplicationisusedfortheintegrationofthefront-endmodelandthatofthewebpages.Theprediction done by the algorithm for the sixteen features inputted is shown in the Figure 2.Here the binary one means the person is suffering from the Parkinson's disease and binaryzeromeansthepersonisnormal.

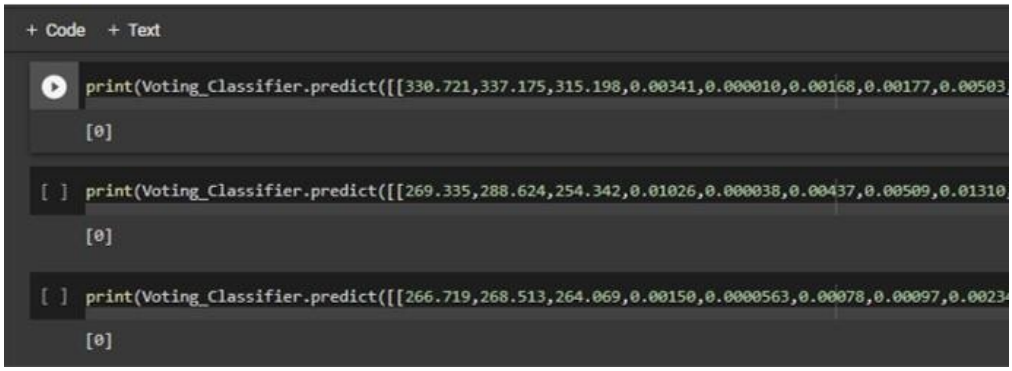

Figure2.Predictionofthedisease

Thevalues ofthevoicefeaturesaretakenfromasoftwarecalledas Praatwhichis usedforthe analysis of the vocal phonetics. It is a free application, where the monotone voice isrecordedmultipletimesandthefeaturesrequired forthealgorithmisextracted.But beforethat voice feature is sampled and the external environment noises that are present areremoved as much as possible. Even though one hundred of it can't be removed, it will stillincrease the accuracy in which the prediction is happening. Below Figure 3 shows the filewhichhasthe extractedvalues.

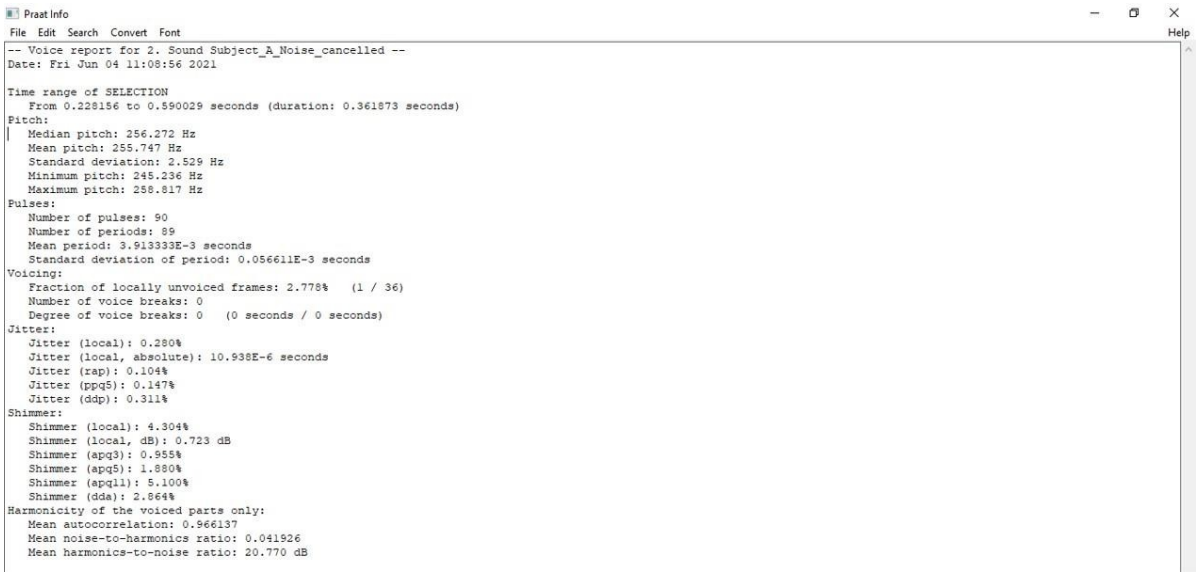

Figure3.Extractedfeaturesvalues

6. CONCLUSIONANDFUTURESCOPE

OneoftheearliestsymptomsofParkinson'sdiseaseisthedistortioninthevoice,itismuchdifferent fromthenormalvoice.Thepitchvariationinthediseasedpersonwillbelesswhencompared to the normal person. Therefore, the features present in the voice are extracted.Adatasetisusedtotrainthedesignedalgorithmandtheextractedvalueisusedastheinpu tfor prediction. The designed voting classifier algorithm with that of the decision tree,support vector machine and logistic regression as the input has a performance accuracy ofninety-five percent. The reason to choose three as the base estimator is because they haveanaccuracyperformancethatbeatstherestoftheweakclassifieralgorithm.Andtherefore,th epredictionwillhappenwithlittletoknowerror.

There are multiple other features that can be taken into consideration for making theprediction. These are all called as non-linear features and their extraction requires a muchmore sophisticated equipment and an almost noiseless environment. A hybrid algorithmcan be designed with two ensemble learners combined together. Wherein depending uponthesubsetof featureschosen,theensemblelearnerwillbechosen.

REFERENCES

[1] A. e A. Spadoto and R. C. Guido, "Improving Parkinson's Disease Identification Through Evolutionary-BasedFeatureSelection.", IEEE,2011.
[2] A. G. EvaldasVaiciukynas, "Parkinson's Disease Detection from Speech Using Convolutional NeuralNetworks.",KaunasUniversityofTechnology,Studentu50,51368Kaunas, Lithuania,2018.
[3] M.E.I.BetulErdogduSakar,"CollectionandAnalysisofaParkinsonSpeechDatasetwithMultipleTypesofSound Recordings.",IEEEJournalofBiomedicalandHealthInformatics,Vol.17,No.4,2013.
[4] A.J.ShivangiandA.Tripathi, "ParkinsonDiseaseDetectionUsingDeepNeuralNetworks.",IEEE, 2019,isbn:978-1-7281-2360-8.
[5] C. D. Timothy J. Wroge, "Parkinson's Disease Diagnosis Using Machine Learning and Voice.", IEEE,2018, isbn:978-1-5386-5917-5.
[6] A.TsanasandP.E.McSharry,"NovelSpeechSignalProcessingAlgorithmsforHighAccuracyClassificationofPa rkinson's Disease.",IEEEtransactionsonbiomedicalengineering,vol.59,no.5,2012.
[7] M. C. Douglas A. Rahn III, "Phonatory Impairment in Parkinson's Disease: Evidence from NonlinearDynamicAnalysisandPerturbationAnalysis.",UniversityofWisconsin-Madison,2007.
[8] P. B. Philippe H. Dejonckere, "A basic protocol for functional assessment of voice pathology, especiallyforinvestigatingtheefficacyof(phonosurgical)treatmentsandevaluatingnewassessmenttechniquesGuid eline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS).", EurArchOtorhinolaryngol,2001.
[9] A. J. AchrafBenba, "Voiceprint's analysis using MFCC and SVM for detecting patients with Parkinson'sdisease.",IEEE,2015.
[10] G. S. C. Okan Sakar, "A Comparative Analysis of Speech Signal Processing Algorithms for Parkinson'sDisease Classification and The Use of The Tunable Q-Factor Wavelet Transform.", Applied Soft Computing(APPLSOFTCOMPUT), 2018.