International Journal of Modern Agriculture, Volume 10, No.2, 2021
ISSN: 2305-7246

Design And Development Of AnAlgorithm To Detect And Diagnose Parkinson'sDisease

Singa ReddyBhagya Lakshmi[1],Siri SGowda[2], TrishaA[3]

[1,2,3] Student at ECE dept, RV College of engineering

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

Abstract

Parkinson's disease is a neuro degenerative disorder that severely deteriorates the brain cells and leads to trembling , rigidity, and toughness in movement, coordination, and balance. Symptoms of this disease gets worse with time and will also lead to amnesia, a decline in mental health, tiredness, and a change in the sleeppattern. The impairment or death of the locomotor-controllingneurons present in our brains is the major cause of this disease.Machinelearningtechniquesandalgorithmsplayaveryimportant role in recognizing the patterns in the medical science field.These techniques have aided many researchers in categorizing medical images and forecasting the models to have a better understanding of complex medical problems.Multiple kinds of research have been carried out int his field using various classifiers and algorithms,which results out in prediction accuracy ranging from 60percentto99percent.Inthispaper,theXGBoost algorithm is being used for the categorization of patientdatasets to predict the people affected by the disease and normal healthy people.A thorough data cleaning, extraction of the feature,and data analysis are implemented before using this algorithm. A successful classification of the people affected by Parkinson's disease and normal people is done using the XG Boost algorithm to obtain high accuracy. This helps in understanding that this algorithm can derive the most discriminative featuresfromclinicaldata.

Key words: Parkinson,Neurodegenerative,UCIrvine,XGBoost

Introduction

Parkinson's disease (PD) is a neurological disorder that is known toaffect the movement ability of a human being. It occurs whenthe dopamine-producing neurons are impaired or dead. The cause of the death of the cells which are known to produce a chemical called dopamineisyet to be known. Along with this, human beings who are affected by this disease also lose the nerve ending which produces a chemical called norepinephrine.It is the chemical thatcontrols many unconditioned functions of the human system, mainly blood pressure, heart rate. Symptoms of this disease startmildly, that it's barely even noticeable.It primarily starts

with tremors in one hand and slowly progresses into tremorsall over the body. Stiffness of the hands, legs and spinal cord,speed reduction in movement, damage of balance, and also thecoordinationwhichsometimesleadsto the collapse is also the effects of this disease. Some of the not so noticeable symptoms include mood swings, depression,trouble while swallowing and speaking, urinary tract infection,skin problems,and insomnia. The cognitive ability of the brain is permanentlydamaged by this disease, and therefore dementia is also oneof the major complications of this disease. Symptoms oftenstart in one side of the body, like maybe only in one part ofthebodybeforeslowlyaffectingtheentireperson.Currently,no blood or laboratory test can be conducted to detect a non-geneticcauseofit.Thediagnosisshouldbepurely done based on the person's medical history, familybackground, and a thorough neurological examination. Thisis an extremely inefficient way of predicting whether thedisease. Research is still going on in this field tofind a cure for this disease. But some medicines, surgery canoften be seen as temporary solutions. The primary cause of this disease is yetto be known.But it has been found out that factors like genetic mutations, exposure to toxins might beoneofthe primereasonsforthisdiseasetooccur.This diseaseaffects1percentofthepopulation,which is over the age of 65, every single year. According to World Health Organization (WHO) reports, it has been estimated that this disease hasaffected more than 10 million across the globe. Andaccording to a recent study conducted in India, some state thatthis disease has affected 7 million elders. The statistics alsoprovesthatmentend to be affected by this diseasemore whencomparedtowomen.Thelackofaccurateearly prediction of this disease has led to it becoming an untreatablepermanentcentralnervoussystemdisorder,andinmostcases, this disease leads to death.

The death toll has almostincreased to 3 lakhs around the world. The tests required to detectt his disease are also extremely costly and not accurate.This is onereason to haveacheap alternative that is efficient and accurate to detect the disease before it be comes incurable.

Various research studies and laboratory tests are beingconductedbytheexperts, they have stated that 90 percent of human beings who suffer from this particular diseasehas speech and vocal disabilities. Theanalysis of voice measurement is simple compared to anyother way of data collection from the patients. One of theotheradvantagesisthatit'snon-invasive,thuspatientswon'tfeel uncomfortable and protruding while the data is beingcollected.Thus,totracktheprogressionofthedisease,measurementofthe vocalscanbeused.

In this paper, an attempt has been made to explore anefficientmachinelearning-basedmodelfortheearlydetection of the disease. It is based on the voice impairmentdata which has been taken from the UCIrvineMachine Learning Repository.This research provides knowledge about the distinctivefeaturesandusesthemor a part of it asaninput for the algorithm, in order to predict PD.Themain aim is to create a model which is capable to detect the disease in real time.

Background

PD deteriorates the movement functions of humans [1]. There are more than 1 millionpeople that are suffering from it in North America alone.PD was earlier known as shaking palsy. Numerous studies have shown that this number can rise in an aging population because it is often seen in people whose age is over 60[2]. PD is caused by the death of braincellwhich are responsible for producing neuro transmitters that include dopamine,serotonin, andacetylcholine.The loss of dopamine results in the symptoms like anxiety, depression, weight loss, and visual issues. Theother symptoms which will be seen withinpeople withPD are poor balance, voice impairment, andtremor. Various research studies have shown that 90% ofindividualswhoaresufferingfrom the PD have speech and vocal problems which include speech disorder,monotone,and hypotonia [3]. Thus, the degradation of voice is taken intoaccount to be because of the initial symptoms of the disorder. Causeandcureofthis disease are none the less unknown;however, the availability of varied drug therapies offer the significant mitigation of symptoms particularly at its earlier stages, thus improving the life quality of patients and additionally reduces the estimated cost of thePathology.The analysis of voice measure is straight forward and non-invasive.Thus,to trace the progression of PD, vocal measurement can be done [5]. For assessing theprogression of the PD, numerous vocal tests have been created which include sustained phonations and running speech texts.The tele monitoring and telediagnosis systems have been widely used as these systems are based on the signals used. Various researchers have made notable attempts to detectPD. The following is aquick run down of some of the research.

By detecting dysphonia, P. E. McSharry, L. O. Ramig, E. J. Hunte, M. A. Little and L. O. Ramig [6] proposed a novel technique for classifying subjects into Parkinson's diseased and control subjects. Pitch time entropy (PPE), a new robust measure of dysphonia, was introduced in their work. Their methodology was divided into three stages: feature calculation, preprocessing and feature selection, and finally the classification. They used a linear kernel support vector machine for classification (SVM). Their proposed model had a 91.4% accuracy rate.

Bhattacharya, I., & Bhatia, M. P. S [7] used a data mining tool called weka to differentiate the healthy subjects from the Parkinson's disease subjects. For classification, they used SVM, a supervised machine learning algorithm. Data preprocessing was performed on the dataset prior to classification. Using libSVM, different kernel values were used to achieve the highest possible accuracy. The linear kernel SVM had the highest accuracy of 65.2174%, while the RBF kernel and polykernel SVM had 60.8696% accuracy.

B.E Sakar et al [4] proposed a model for distinguishing between normal subjects and subjects who were affected by PD. Data were obtained from 40 people in their report (20 were healthy subjects and 20 were Parkinson's disease subjects). A total of 26 voice samples were taken from each subject, including short sentences, phrases, numbers, and sustained vowels. They used SVM and knearest neighbor to classify the data (k- NN). They used

Summarized Leave-One-Out (s-LOO) and Leave-OneSubject-Out for cross-validation (LOSO). For k- NN, the values 1, 3, 5, and 7 were chosen, while for SVM, linear and RBF kernels were used. The k-NN classifier got 82.50 percentage of accuracy, while the SVM classifier got 85 percent.

The main goal of Jilbab, A, Benba, A, Sandabad S, and Hammouch, A [8] was to distinguish between humans with PD and normal ones. The data used in their study consisted of 34 sustained vowels obtained from 34 individuals, 17 of whom were Parkinson's disease patients. For classification, SVM with various kernel types was used. As a cross-validation method, LOSO was used. On taking the top 12 MFCC coefficients, linear kernel SVM recorded the bestaccuracy of 91.17%.

A method for predicting the PD was proposed by Mathur R, Pathak V, & Bandil D [9]. They used the Weka tool to implement the algorithms for data preprocessing, classification, and analysis of the results on the given dataset. They used Adaboost.M1, bagging, and MLP in addition to k- NN. The best classification accuracy was found to be 91.28% with k-NN + Adaboost.M1.

Artificial neural networks were used by Saritas I, Cinar A C, Yasar A and Sahman M A [10] to diagnose PD. Forty Five properties were chosen as input values and one output for the classification using the MATLAB tool. With an accuracy of 94.93%, their proposed model was able to distinguish healthy subjects from PD subjects.

Dataset

The dataset that is being used in this research is based upon the symptoms taken from the voice feature of the subjects. The dataset being used here is taken from UC Irvine Machine Learning Repository. It is a compilation of multiple theories, databases regarding multiple domains and data reproducers which can be used by machine learning enthusiasts for experimental analysis of algorithm. The dataset for this research is collected from a total of 48 people, each of them giving 3 sample inputs. Out of these 65, 49 people have the Parkinson's disease and 16 people are normal. The data that was collected is in the comma-separated value (CSV) format. Frequency, Jitter, Shimmer, noise to harmonic ratio (NHR), harmonic to noise ratio (HNR), removable partial denture (RPDE) is some of the features present in the dataset**Methodology**

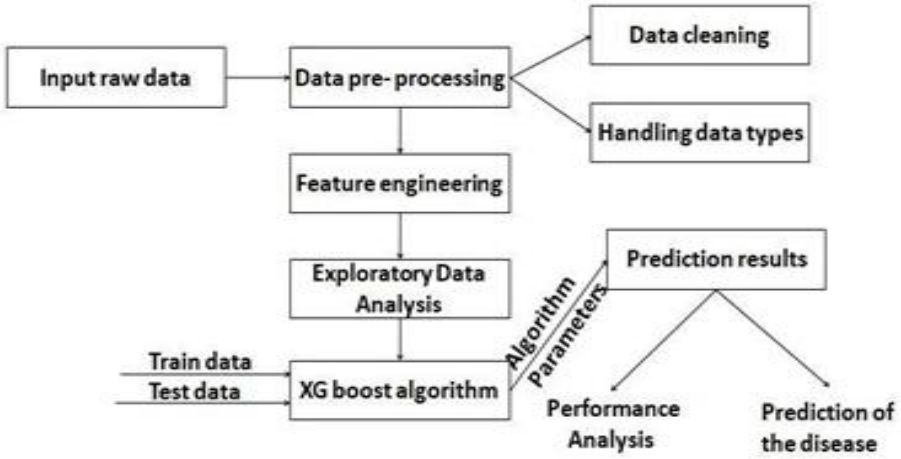

Fig 4.1

The main objective of the project is predicting Parkinson's disease progression over time for a set of patients from a clinical study which is shown in the

The approach towards the project starts by collecting input raw data which usually contains noise and sometimes in an unusable format that prevents them from being used directly in machine learning models. Hence, the application of data pre-processing is required. Data pre-processing is one of the most critical steps which contributes to the accuracy of the model. This step allows us to process and transform the raw data into a clean and usable form of data, making it more meaningful and standardized. Data cleaning is done to pre-

process where the detection and correction of inaccurate records from the database take place. It identifies the incorrect parts of the data and then replace, modify or even delete the data that aren't required

The implementation of feature engineering is required to extract the feature from data in form of numerical representation of an aspect of real-time phenomena or data which is suitable for machine learning algorithms. This process involves feature selection, feature transformation, and feature extraction. After the extraction of feature explanatory data analysis is performed. It is a technique for analyzing and investigating our data set and illustrating the main characters generally in the form of a data visualization technique. This helps to discover the recurring patterns, spot anomalies, and so on, thus providing a better understanding of data set variables and the relationship between them. Once this step is complete, its features can be used for more sophisticated modeling in machine learning.

The existing data set is split into two stages, that is training and testing stage. Starting with the data set, the training set is randomly selected, which forms 80% original data, and the remaining is a testing set. The two sets are displayed to make

sure, it is not completely different. The data is fit into a model and the training of data gets completed. Next, the test set is used for the testing stage to compare if the result obtained is similar. Initializing an XGBoost Classifier and training the model. After our model produces the required result, values are predicted and finally generated. The accuracy of the model is predicted and displayed.

Implementation And Results

Dataset were derived from the UC Irvine Machine Learning Repository. It has feature divided values of voices of both normal subjects and those who are suffering from PD. The machine learning model is based on ANN. The CSV file which has the dataset of the subjects is given as the input to the model. It learns from the non-linear patterns in signal values and trains itself based on it. It is created using keras and tensor flow. The dataset was divided into training set and testing set and was randomly fed to the XGBoost algorithm after the data pre-processing and feature extraction procedure. An accuracy of more than 90 percent was observed for the input given and the precise classification of the people with the Parkinson's disease was done.

Hence the model that has been proposed here is efficient and effective compared to the other algorithms.

Conclusion

Different research papers were thoroughly studied to find an efficient way of detecting PD. There is no blood test or any other laboratory test that can be conducted to find out whether a person is affected or not. Medical history of the patient and thorough neurological examinations should be done by the doctors. But these are not very efficient and are extremely costly. Misdiagnosis or late diagnosis can even lead to the death of the affected person. In such cases, models created using machine learning can be an aid to medical practitioners. Various machine learning algorithms were assessed for classifying the subjects. The conclusion that was landed on is that it is good to implement a machine learning technique on the vocal dataset. The dataset was collected from the UC Irvine Machine Learning Repository. A prediction model for this disease was developed using a machine learning technique, namely the XGBoost algorithm. Testing and training of the data was done using this algorithm and was found out that it is a very efficient model that gives a high accuracy output. It can also be emphasized that the proposed method for detection is fully automated and provides improved accurate performance and therefore it can be used in real-life applications.

In the coming future, it is planned to use a VGFR spectrogram detector along with this to get a much more efficient model with a high accuracy rate to classify the patients. A new algorithm will be designed to reduce the computation timing. This will, in turn, make the system lighter, precise, and more compatible.

References

1. Olanow C W, Stern M B, & Sethi K"The scientificand clinical basis for the treatment of Parkinson disease",Neurology,72(21Supplement4),S1-S136,2009.

2. Langston, J W,"Parkinson's disease: current and futurechallenges", Neurotoxicology,23(4-5),443-450,2002.

3. Rigas G, Tzallas A T, Tsipouras M G, Bougia P,Tripoliti E E, Baga D, &Konitsiotis S, "Assessment oftremoractivityintheParkinson'sdiseaseusingasetofwearablesensors",IEEETransactionsonInformationTechnology inBiomedicine,16(3),478-487,2012.

4. Sakar B E, Isenkul M E, SakarC O, Sertbas A,Gurgen, F, Delil, S &KursunO,"Collection and analysisof a Parkinson speech dataset with multiple types of soundrecordings",IEEEJournalofBiomedicalandHealthInformatics,17(4), 828-834,2013.

5. RahnIIIDA,ChouM,JiangJJ,&ZhangY, "Phonatory impairment in Parkinson's disease: evidencefrom nonlinear dynamic analysis and perturbation analysis",JournalofVoice,21(1), 64-71,2007.

6. M A Little, P E McSharry, E J Hunter, J Spielman, andLORamig,"Suitabilityofdysphoniameasurements fortelemonitoringofParkinson'sdisease,"IEEETrans Biomed.Eng.,vol. 56, no. 4, pp. 1010–1022,2009.

7. BhattacharyaI,&BhatiaMPS, "SVMclassificationto distinguish Parkinson disease patients" In Proceedings of the 1stAmritaACM-WCelebrationonWomeninComputing in India(p.14).ACM,September2010.

8. Benba A, Jilbab A, Hammouch A, &Sandabad S, "Voiceprint's analysis using MFCC and SVM for detecting patientswithParkinson'sdisease",IEEEInternational conferenceonelectricalandinformationtechnologies(ICEIT) (pp. 300-304),2015March.

9. MathurR,PathakV,&BandiD,"ParkinsonDiseasePredictionUsingMachineLearningAlgorithm",Emerging Trends in Expert Applications and Security (pp.357-363).Springer, Singapore,2019.

10. YasarA,SaritasI,SahmanMA,&CinarAC, "Classification of Parkinson disease data with artificialneural networks", IOP Conference Series: Materials ScienceandEngineering(Vol.675,No.1,p.012031).IOPPublishing, 2015, March.

AUTHORSPROFILE

1. SingaReddyBhagyaLakshmi VIIIsemester,ECEdepartment,RVCollege of Engineering, Bangalore

2. SiriSGowda VIIIsemester,ECEdepartment,RVCollege of Engineering, Bangalore

3. TrishaA VIIIsemester,ECEdepartment,RVCollege of Engineering, Bangalore