Journal of Green Engineering (JGE)

Volume-10, Issue-3, March 2020

Diagnosis of Parkinson Disease Using Sensor Data and Machine Learning Approach in Mobile Cloud

[1]N. Janani and [2]S.Kanagaraj

[1]Student,Department of Information Technology, Kumaraguru College of Technology, Coimbatore, TamilNadu, India. E-mail: janani.18mds@kct.ac.in
[2]Assistant Professor ,Department of Information Technology, Kumaraguru College of Technology, Coimbatore, TamilNadu, India. E-mail: kanagaraj.s.it@kct.ac.in

Abstract

In reality, the utilization of sensors is developing each day to improve the personal satisfaction by giving medicinal services data on clinical diagnostics. There are various sensors like detecting advance sensors, "electronic gadgets" physical sensors have been effectively shown in the field of biomedical applications because of sensor great working ability. Cloud computing technology is used to accommodate vast amounts of data. Such medical applications also use the cloud computing platform to store and access the data protected. Parkinson's disease (PD) could also be a chronic long condition of the central systema nervosum that primarily affects the motor system of the patient. The side effects of Parkinsons' sickness incorporate muscle rigidity, tremors and modifications in speech. The objective of the framework is to analyse the parkinson's disease in voice detection. As the indications are worsen, the patients motor and non-motor system will get failure. The framework is used for diagnosing the parkinson's disease. The proposed system concentrates on improving the Parkinson's disease diagnosis using voice signals from patient detected from different

Journal of Green Engineering, Vol. 10_3,684-696.

sensors and uploaded to the cloud for processing . This frame work concentrates on improving disease diagnosis with experimental results using Support Vector Machine, Logistic Regression, Random Forest and eXtreme Gradient Boosting.Machine learning algorithm mainly concentrates on improving the prediction of parkinson's disease diagnosis.

Keywords: Cloud computing, Parkinson's Disease, Support Vector Machine, Random Forest, XG Boost, Logistic Regression

1 Introduction

Cloud computing gives a promising approach to satisfy the IT wants in a favourable way. With the fast healthcare and economic development, and more clinical documents are generated.Key issues towards the incorporation of biosensing frameworks incorporate the interest for expensive costs.The motivation to enhancing the level of present day data administration with the aid of using adavanced technology is increased.Cloud Computing emerges in a range of institutions, companies and man or woman users, to share a massive amount of public and private information and statistics in a extra low cost way and dependable IT structure.

More specifically, the scientific and stastics computers are fully cloud based, to understand sharing of medical data and health information, coordination of scientific service, along with the cost-containment medical statistics device infrastructure by using the implementation of a dispensed and high-integrated platform[5]. New innovations,such as cloud computing used in the scientific subject will be developed in the future technology[6]. Health care scientific subject is the primarily important position in cloud computing.

Virtual hospitalization is executed by means of the cloud computing technologies. In which healthcare providers, doctors and patients can also get right of entry to their scientific information through cell devices. The healthcare enterprise makes use of another technology is Wearable Sensor Technology. The wearable gadgets used due to their size, like computation strength and practical power capabilities. The wearable devices loaded with sensors to display user's every day behavior. The wearable devices used due to their size, reasonable computation electricity and realistic power capabilities. In sensible implementation, as an alternative of wearable sensors the wi-fi sensors are used due to their cost. Wearable Sensors are used to track home environments for monitoring and collecting the data.

Parkinson's issue (PD) is a nervous system disease. Though then, on James Parkinson's surname i.e. his Parkinson is used as a name for this ailment. It is normal at certain point in the world and attacks patients older

than 60 years[12].It is a neurodegenerative issue that influences neuron-creating dopamine basically in a specific locale of the Genius called substantia nigra. It prompts a bringing down in a synthetic called cerebrum dopamine.

Dopamine plays an important part of movement synchronization. It's like a confectioner who transmits messages inside the brain[8]. Patients experience the ill effects of movement issue because of the loss of these cells. Person with Parkinson's disease also loses the nerve endings containing norepinephrine, the sympathetic nervous system's primary chemical messenger,that controls many of the body's automatic functions including heart charge and blood pressure.

2 Related Works

Many scientists grouped the Parkinson's disease with the guide of various techniques. Abishek M.S et.al is primarily based on an strategy to disease analysis the usage of voice dataset. The first step is the extraction of points the usage of the genetic algorithm and the foremost aspect analysis and the classification algorithm is used after function extraction. Support vector machine and KNN is done for the person affected with the disease or not. The best classification accuracy is acuqired using this method is 97.5% [1].

Chaitra B.R et.al depends on the beginning period expectation of parkinson's ailment utilizing neural system calculations. Artificial Neural system works about 93.2% precision is accomplished Deep learning neural system accomplishes gives 73.84%[2].

Satyabrata Aich et.al based on a method for different feature selection for voice data set using supervised learning models. SVM gives the best result of accuracy 97.5% after the feature selection with Genetic Algorithm and Principal Component Analysis[4].To analyzes Parkinson's sickness utilizing voice recording dataset. The dataset is pre-prepared with missing information and afterward divide into preparing set and test set for the model. For this AI method of tree based calculations are utilized. The Fold system is utilized to get the best exactness inside the models.

A new formula supported voice analysis to diagnose Pd,the primary step is to use genetic formula to select optimum properties of all extracted characters. To search out the bottom error rate the Support Vector Machine, Extreme learning machine and K-Classifiers have done .The disease classification the usage of data mining methods[3]. They used characteristic relevance with more than one algorithm. Here the Random Tree fabricates the grouping rule principally dependent on this three ascribe highlights to achieve the zero blunder rate.

As of late, the utilization of distributed computing has gotten well known in medicinal services structures, as distributed computing empowers a system to utilize a lot of computational assets based on request. Cloud processing are portrayed by on-request self-administration, made sure about get to, asset pooling, a wide scope of availability, and quick Flexibility[5].

3 Materials and Method

3.1 Proposed Method for Detection

The structure includes a few segments:The frame work includes sensors, however we can catch voice signals through voice recorders, and compact PCs or tablets. In wake of handling, the outcomes are forwarded to a doctor, who recommends drug to patient through cloud. As a rule, cloud incorporates both equipment and programming. The programming is principally on the patient side, while the equipment is for the most part on the organization side. The cloud offers[7] some essential kinds of assistance: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). IaaS incorporates registering assets, capacity, correspondence systems, virtual machines, and servers. The PaaS includes plan, displaying, improvement, testing, databases, and web servers. SaaS incorporates UI, virtual work area, and email and office applications. The patients are associated with SaaS by means of an internet browser.

The patient transfers voice sign to the cloud. The cloud chief validates patient records and gets the sign. The chief at that point imparts the sign to the component extraction and arrangement server,which separates and chooses a few highlights from the sign, which is then delegated PD or sound dependent on the chose highlights. When the sign is arranged, the cloud chief builds up an association with an accessible enrolled specialist, sending the outcome and the voice test to the specialist. The specialist at that point sends solution back to the cloud,and cloud director informs patient with decision and remedy. In light of the solution, the patient acquires treatment.

3.2 Machine Learning Algorithms in Disease

Artificial Intelligence depends on information science technique that empowers PCs to pick up understanding into the examples and attributes of existing information so that data outcomes and trends can be predicted without specific programming to classify them. Machine learning algorithm has strong experience of diagnosing illness and preventing them[9]. A huge amount of reasearch had been published showed application of machine learning algorithm in clinical field, for example,diagnosis of

disease,prediction of disease,survivability and identification of disease.This can make tasks like Parkinson's disease more automated,efficient, and accurate with a robust machine learning model shows and recognize examples and qualities that people may notice[11]. This paper will essentially use regulated supervised learning through a current dataset of voice recording data.

3.2.1 Support Vector Machine

Support vector machine is a regulated machine learning[10] of calculation which can be utilized for both ranking and regression difficulties. In any case, it is used in grouping issues for the most process. In this calculation, we plot each data issue as a point in n-dimensional space where n represents number of features, whereby we estimate each variable as a specific facilitate. At this point, we perform classification by seperating the hyper-plane and its separates the two training classes.The model builts test image to predict whether the image is in the same class or another.The algorithm needs a vast amount of data to get a efficient boundary and computational cost is very high for detection.It is a learning algorithm for classification[4].It tries to find optimal separation between the hyperplane such that expected classification error for unseen pattern is reduced.

Regardless, there are conditions where a nonlinear area can confine the events even more singnificant. SVM deals with this by utilizing a part (nonlinear) capacity to outline data to an elective space where a hyperplane can be utilized to do the separation. This implies the non-direct structure is discovered by a straight learning machine in a high-dimensional part space while the constraint of the system is constrained by a parameter that doesn't depend upon the dimensionality of the space. The algorithm aims to focus on best separating hyperplane by the equation Y=Wx+b that maximizes distance between two classes.

3.2.2 Logistic Regression

Not many of measurable calculations are straight segregate investigation, least mean square quadratic, piece, Logistic Regression and k closest neighbours. This algorithm is stastical classifiers that are utilized for examination of information. It is a type of linear regression used for analysing binary or multi class dependent variable. Logistic Regression[10] is a method used for characterization technique that has been utilized broadly in many applications including document classification, computer

vision,natural language processing.It is mathematically defined as Pr(G=K|X=x).

3.2.3 Random Forest

Random Forest is an addition of bagging strategy, is a type of ensembling method. Due to an dataset collection containing m tests, one example is selected and placed into sample collection and sample is then put back into original data collection,so that sample can still be chosen at next sampling time. In this way we get a sample get a sample collection of m samples after m random operations. A few examples in the underlying preparing set show up in the resampling set a few times. T tests containing m training tests are chosen, at that point an essential test is prepared dependent on each example set, and afterward these fundamental samples are consolidated. Bagging typically utilize of voting a task. The base learner of Random Forest is a decision tree, and random attributes selection is incorporated in to the decision tree process. RF overcomes the issue over fitting. In RF, parameters can be set easily and eliminates the need for tree pruning. RF is simple, intelligible,cost-effective and has achieved powerful success in present world activities. RF been utilized in gene selection, remote detecting, and disease prediction and has achieved better results. Random Forest procedure is shown in the accompanying calculation.

Algorithm for Random Forest

Step 1: Take a new bootstrap from the training package.
Step 2: Create research on an unpruned tree at this bootstrap.
Step 3: Select (m try) randomly at any internal node, and decide the best split.
Step 4: Where every tree is fully grown. Cannot do pruning.
Step 5: Production calculation overall, as all trees vote by plularity.

3.2.4 XG Boost

Extreme Gradient boosting is new gradient boosting learning strategy for the group. It applies an machine learning calculation beneath the angle boosting worldview, and has high productivity, ease of use and conveyability. It is an exceptionally proficient and ordinarily utilized type of machine learning that improves the learning of the troupe. Boosting is a calculation family which can make weak learners and good learners.By means of gradient boosting the decission tree is supported. The training

requires the entirety of xgboost's principle segments to give the best forecast model.

Algorithm for XG Boost

Stage 1: Input - load the numeric qualities in the info parameter.
Stage 2: Target variable - The goal variable or vector is either 0 or 1 for characterization question
Stage 3: objective - The objective of double grouping is strategic.

Target work = preparing misfortune + regularization i.e Obj = L + Ω

In xgboost, target work is advanced by the slope plunge

$$\text{Obj}^{(t)} = \sum_{i=1}^{n}[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)] + \Omega(f_t)$$

Stage 4: Number of emphases - Number of trees to use in the arrangement.
Stage 5: Early halting – To prevent over fitting the dataset into our model during approval this component will stop the emphasis when it doesn't see improvement in the precision and it will likewise indicate the no of cycle to be halted before over fitting

4 DataSet

The dataset contains 197 instances and 23 attributes. These are the features with various fundamental frequency, pitch and amplitude permutation. We viewed multiple signs of patients such as speech and key stroke information for disease prediction.

Table 1 List of attributes

Attribute Name	Attribute description
Name	ASCII subject name and recording number
MDVP: Fo(Hz)	Average vocal fundamental frequency
MDVP: Fhi(Hz)	Maximum vocal fundamental frequency
MDVP: Flo(Hz)	Maximum vocal fundamental frequency
MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: RAP , MDVP: PPQ Jitter: DDP	Several measures of variation in fundamental frequency
MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP:APQ Shimmer:DDA	Several measures of variation in amplitude
NHR,HNR	Two measures of ratio of noise to tonal components in the voice
Status	Health status of the subject (one) Parkinson's (zero)-healthy
RPDE,D2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent
Spread 1, Spread2, PPE	Three nonlinear measures of fundamental frequency variation

5 Result And Discussion

The frame work is carried out in Java programming language. The research includes 195 instances for patients with Parkinson's disease. This framework used a machine learning algorithms such as support vector machine, practical regression, random forest, and XG Boost are implemented. To predict Parkinson's disease, we used the learning algorithm.

The result shows Parkinson's detection using random forest algorithms is efficient.

The table 2 shows the comparison of four different algorithms used and the accuracy is predicted.

Table 2 Experimental Results of proposed system

Algorithm	**Accuracy (%)**
Support Vector Machine(SVM)	84.6%
Logistic Regression(LR)	79.4%
Random Forest(RF)	92.3%
XG Boost	91.2%

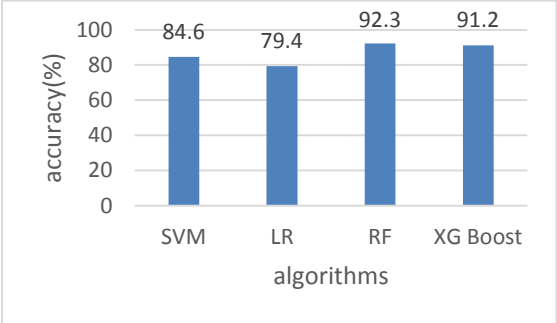

Fig 1 The Parkinson Disease detection accuracy using various machine learning algorithms.

Figure:1 represents the result of accuracy done.In this framework, four different algorithms were used.From the figure:1 ,the combination of classification algorithms are used to get better performing algorithm.This frame work gives random forest is acuquiring better performance than the other algorithms.

In this table 3, it shows the actual and predicted values for precision,Recall and F1 score.To get value of precision, we divide the whole variety of correctly labeled effective examples with the aid of the total wide variety of expected tremendous examples. Recall can be described as the ratio of the total wide variety of correctly categorised examples divide to the

complete wide variety of positive examples.The f score is a measure of test's accuracy it has both measures of precision and recall test score.

Table 3 Performance scores for all models

Model	Precision	Recall	F1 score
Support Vector Machine	0 - 1.00 1 – 0.83	0 – 0.40 1 – 1.00	0 – 0.57 1 – 0.91
Logistic Regression	0 – 0.67 1 – 0.52	0 – 0.40 1 – 0.93	0 – 0.50 1 – 0.87
Random Forest	0 - 1.00 1 – 0.91	0 – 0.70 1 – 1.00	0 – 0.82 1 – 0.95
eXtreme Gradient Boosting	0 – 0.95 1 – 0.83	0 – 0.65 1 – 0.98	0 – 0.76 1 – 0.83

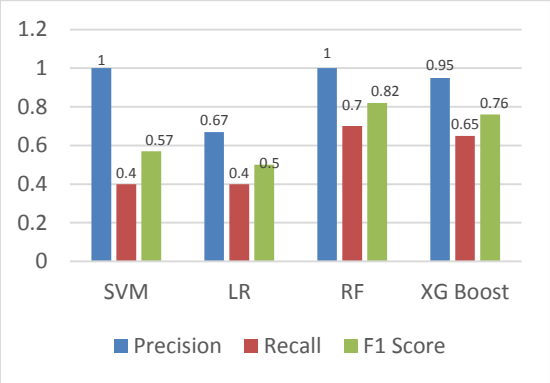

Fig 2 Representation of actual values(0)

The fig 2 represents the actual values are predicted from the dataset in the form of 0's.The values are calculated for all measures.

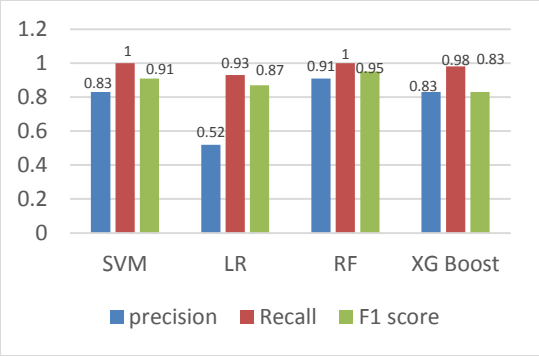

Fig 3 Representation of Predicted Values(1)

The fig 3 represents the values are predicted from the dataset in the form of 1's.The values are calculated for all measures.

6 Conclusion and Future Work

This research discussed about absolute most recent machine learning technique. The technology work addresses cloud computing's important problems and solutions, Artificial intelligence is a way to deal with separate between the benchmark group and parkinson's patient.Above all else, diagnosing parkinson's disease will absolutely accomplished utilizing the voice datas.This work can be Improvised by including real time based scenarios and more voice datas and evaluating different models of machine learning.

Reference

[1] Abishek M.S, Chethan C.R,et.al, "Diagnosis of Parkinson's disorder through speech Data using Machine Learning Algorithm",IJITEE,Vol.7, No.3,PP.69-72, 2020.

[2] Chaithra B.R1, Gowthami A.L1, Harshitha U1, Keerthana N1, Asha V.G, "Early Stage Prediction of Parkinson's Disease using Neural Network", IRJET,Vol.6,No.5, 2019.

[3] Akshay.S, Kiran Vincent, "Identification of Parkinson Disease Patients Classification using Feed Forward Technique Based On Speech Signals",IJEAT,Vol.8,No.5,PP.1769-1778, 2019.

[4] Satyabrata Aich, Hee-Cheol Kim, Kim younga, Kueh Lee Hui, Ahmed Abdulhakim Al-Absi and Mangal Sain , "A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease", ICACT Transactions on Advanced Communications Technology ,Vol.7,No.3,PP.1116-1121, 2018.

[5] K.A. Al Mamun, M. Alhussein, K. Sailunaz,et.al, "Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications",Future Generation Computer Systems,Vol.66,PP.36-37, 2017

[6] G. Muhammad,et.al, "Smart health solution integrating IoT and cloud: a case study of voice pathology monitoring," IEEE Communications Magazine,Vol.55,No.1,PP.69-73, 2017

[7] Musaed Alhussein, "Monitoring Parkinson's disease in smart cities",IEEE acess,Vol.5, 2017.

[8] Satyabrata Aich,et al. "A mixed classification approach for the prediction of Parkinsons disease using Non linear feature selection technique based on voice recording", International Conference on Inventive Computing and Informatics (ICICI), 2017.

[9] Kamal Nayan Reddy Challa et al., "An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques", International Conference on Signal Processing, Communication, Power and Embedded System, 2016.

[10] R. prashanth et al, "High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning", international journal of medical infomatics, pp.13-21, 2016.

[11] Mohammad Shahbakhi, Danial Taheri Far, et al, "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine",JBiSE,Vol.7,No.4, 2014 .

[12] R.C Helmich, et.al, "Cerebral causes and consequences of Parkinsonian Resting Tremor", Brain,vol.135,No.11,PP.3206-3226, 2012.

Biographies

Janani N was born in Coimbatore, India. She received her Bachelor degree in Information Technology from Kumaraguru College of Technology, Coimbatore, in 2018. She is currently pursuing a Master degree in Data Science at Kumaraguru College of Technology, Coimbatore. Her research interest includes Machine Learning, Cloud Computing and Data Mining.

Mr S. Kanagaraj completed his B.E in Computer Science from PGP College of Engineering and Technology, Namakkal and M.E in Computer Science from Kumaraguru College of Technology. He worked for an IT and ITES organization, Chennai as System Administrator for a period of 3 years, He joined as an Lecture in Kumaraguru College of Technology, Coimbatore in 2008. Currently he is continuing as an Assistant Professor in Kumaraguru College of Technology, Coimbatore. He is passionate to educate, teach and motivate the student community in the field of Networking and IoT. He has authored 5 technical papers at various International Journals. Now he is IEEE student branch counsellor at Kumaraguru College of Technology and motivating students to take part in various IEEE activities.