International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-7, May 2020

Predictive Analytics Algorithms for Clinical Decision Making in Healthcare

P. Selvashankari, P. Prabhu

Abstract: ***Healthcare is major issue and challenge now-a-days for human being in a daily life. Parkinson Disease or P.D is a one of the disorders that affected in the mid of nervous system. Parkinson Disease affected person cannot be act as normal human being. Among the innumerable disease listed so far, the Parkinson disease occupies an alarming position due to its life threaten concern. Early prediction of Parkinson disease from large volume of electronic health records leads to protect various health issues. There are various challenges and issues such as scalability, accuracy, risk factor, time complexity and sparsity in early prediction of Parkinson disease. There are various conventional algorithms have been proposed to solve these issues and challenges and still needs improvement. The present study, systematic predictive analytics using various classification algorithms such as Support Vector Machine (SVM), Random Forest, AdaBoost, Multi-Layer Perceptron (MLP), Naive Bayes, Decision table, J48, Logistic Regression is presented and evaluated using benchmarking Parkinson disease data set which are collected from UCI machine learning repository. The extraction of hidden data present in the dataset is obtained using WEKA environment. The results from the prediction models gives better clinical decision-making support to the doctors in predicting disease earlier and risk level.***

Keywords: ***Decision Making, Healthcare, Knowledge Discovery, Machine Learning, Parkinson's, Prediction, Supervised Classification.***

I. INTRODUCTION

The healthcare systems generate large volume of data day by day. This is mainly due to the improvement in science and technology especially in the field of information technology generates more data from patient's electronic health records. "PARKINSONISM" or "PARKINSONIAN SYNDROME" is a general term of Parkinson disease some scientists also refer the disease as "SYNUCLEINOPATHY". PD is generally classified into two types namely Primary Parkinsonism or Idiopathic Parkinson and secondary Parkinson or Parkinson Plus.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author

P.Selvashankari*, Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India. E-mail: selvashankari95@gmail.com.

P.Prabhu, Assistant Professor in Information Technology, Directorate of Distance Education, Alagappa University, Karaikudi, Tamilnadu, India. E-mail: pprabhu70@gmail.com, prabhup@alagappauniversity.ac.in.

Secondary Parkinson is further classified into Drug induced Parkinsonism, Vascular Parkinsonism', Normal pressure hydrocephalus, Corticobasal degeneration (CBD), Progressive supranuclear palsy (PSP) and Multiple System Atrophy (MSA). A cause of Parkinson disease is unknown but it occurs based on genetic and environmental factors. A common terms of Parkinson diseases that the treatment can help but "It could not cure disease". PD requires various proper medical diagnoses such as arthritis, depression, drug induced parkinsonism, multiple system atrophy, obsession slowness, psychogenic parkinsonism, toxins and vascular parkinsonism. A Parkinson disease cannot be identified or predicted through x-ray or blood test. The symptoms of Parkinson disease are tremors, shaking, bradykinesia (Movement slowness), depression, difficulty in Walking, rigidity and postural instability. A Neurological condition of PD patients affects a wide range of functions. It can change the mental and physical aspects of a personal life and their complications are more. They are speaking, chewing, swallowing, depression, anxiety, sleep, urinary problems, constipation, dementia, pain, blood pressure and sense of smell problems. They have difficulty to speak over time and harder to communicate the ability of thinking. It may hardest to merge with the social activities. Chewing swallowing difficulties occur for PD patients during later stage. It may cause stuck in the throat and shocking while getting food is risk. Depression anxieties occur almost 40 to 50 percent of PD affected person, they experience the depression. Depression has physical symptoms they are mood disturbance, anxiety, sleep problems, the person thinking behavior is changed. PD people mostly have a sleep problem, some people may fall in sleep easily and other people find that hard to sleep. It may other sleep problem include: daytime sleepiness, nightmare, talking during sleep, restless legs, difficulty turning over in bed, walking up to use the bathroom and urinary problem. PD people may have the urinary leakage. It can be due to medicine they take. Two-third of the people who affected PD is commonly have constipation. It may causes some reasons muscle weakness, a reduction in fluid, side effects of medications they take.

PD lead to dementia for Lewy bodies, they are common symptoms of dementia difficulty remembering and focusing, speech is unclear, depression and anxiety, sleep difficulties and daytime sleepiness. PD persons commonly have a pain. The pain comes after the third tremor and stiffness. The normal person's blood flow and blood pressure are controls autonomic nervous system; PD person's nervous system cannot be controlled. An affected person's blood pressure changes throughout the day.

Commonly PD persons have low blood pressure some peoples have very high blood pressure it may lead heart problems. The 95 percent of the PD people lose the smell sense. This is the first symptoms of the PD affected persons. Hence it is a top most symptom.

There are various challenges and issues in prediction of PD.

These include scalability, accuracy, risk factor, time and space complexity.

Electric Health Records (EHR) are commonly used to solve these challenges and issues using data mining techniques and Internet of Things (IoT) for monitoring health care clinical decision support system.

Data mining in healthcare: Data mining used and proven an effective in healthcare. it is efficiency and quality of organizing the data through data mining. the purpose of healthcare in data mining is analyzing a large amount of data. it is more effective in predictive medicine, customer relationship management, fraud detection, managing of healthcare & measuring the effectiveness of treatments, manage healthcare in different levels.it may collect a medical data and extract knowledge hidden.

It used mainly in assist clinicians at the Point of Care .it highly focused knowledge manage of the patient data. The major topics in artificial intelligence in medicine are covered in clinical decision support system. It contains two major types: knowledge based clinical decision support system; non-Knowledge based clinical decision support system.

Knowledge based clinical decision support system: Knowledge based clinical decision support system contains an rules and associations of complied data the form of IF THEN RULES. There are two types inference engine, communicate the mechanism. Non-Knowledge based clinical decision support system: non-Knowledge based clinical decision support system it allows the computer to find the past experiences and the clinical data pattern. there are two types artificial Neural Network and genetic Algorithm. Rest of the work is organized as follows; The chapter 2 deals with related works about classification and prediction of Parkinson's disease using various techniques. The chapter 3 explains about methodology about Parkinson's disease classification algorithms and how they are experimentally evaluated. The chapter 4 presents experimental setup about dataset description, WEKA tool used and measurements for evaluation. The chapter 5 explains about the result and discussion. The chapter 6 conclusion about the disease and the future work about the Parkinson's disease.

II. RELATED WORKS

There is various research work was carried out in the literature using datamining and machine learning techniques especially supervised and unsupervised learning methods to support clinical decision making to doctors and patients. Artificial Intelligence also play vital role in prediction especially in the field of getting knowledge from large datasets.

Indira Rustempasic and Mehmet Can [1] Voice measurement has shown a great progress in the advancement of Parkinson Disease detection. About 90% of people with Parkinson's disease present some kind of vocal deterioration. The goal is to divide the dataset in such a way that objects belonging to the same cluster are as similar as possible, whereas objects belonging to different clusters are as dissimilar as possible. In non-fuzzy or hard clustering, data is divided into crisp clusters, where each data point belongs to exactly one cluster. Prabhu P and Selvabharathi S, [2] presented a model based on deep belief networks for prediction of diabetes types. This method is implemented using real-world datasets. The method outperformed with many conventional methods in terms of classification accuracy.

Basil K Varghese et.al. [3] A tell-tale marker of this disease is a decrease in the dopamine levels in the brain which could be attributed to the degeneration of dopaminergic neurons. The onset of the disease may be suggested by tremor, rigidity, slowness of movement and postural instability. The motor symptoms aren't severe enough for detection. Therefore, monitoring progress of the patient and effective diagnosis requires persistent visits by the patient. Ahmad Almogren[4] PD administration requires normal clinical visits for evaluation and close checking of PD side effects. In any case, it is difficult to monitor separation of PD side effects between center visits. A considerable measure of consideration and there are some dynamic endeavors exist in the writing. Ramzi M. Sadek et.al. [5] Parkinson's disease is a long-term degenerative disorder of the central nervous system that mainly affects the motor system. The symptoms generally come on slowly over time. An ANN model was presented for Parkinson's disease prediction to help specialist in the field. The obtained accuracy is 100%. Alexander Yu. Meigal (Petrozavodsk State University, Russia), et.al [6] discussed Ambient Assisted Living At-Home Laboratory for Motor Status Diagnostics in Parkinson's disease Patients and Aged people. Analyzing the health condition of the old aged people that affected Parkinson's disease.

Vojtech illner et.al [7] Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smart phone in Parkinson's disease. It recognized the vocal folds frequency of speech impairments in Parkinson's disease that exist the current performance in smart phone-based evaluation and robustness against background noise. Cynthia M.fox and lorranie Olson ramig[8] Vocal Sound Pressure Level and Self-Perception of Speech and Voice in Men and Women With Idiopathic Parkinson Disease. American journal of speech language pathology, PD were examined on additional variables, such as time post diagnosis and stage of PD. Taha Khan, et.al [9] Assessing Parkinson's disease severity using speech analysis in non-native speakers. Elsevier cepstral separation difference (CSD) features to quantify dysphonia and dysprosody accurately distinguish the severity of speech impairment. Paulraj Prabhu., and Neelamegam Anbazhagan[10] proposed method for improving Business Intelligence Based on Frequent Itemsets generated from candidates set using k-Means Clustering Algorithm. K-means method is used to group the object of users to form cluster of users to classify them.

III. METHODOLOGY

The aim of this research methodology is to propose a method for discriminate healthy people from PD using various machine learning, supervised classification techniques.

This research work prediction model architecture is shown fig 1. In this work classification techniques such as SVM, Random Forest, Ada Boost, Multilayer perception, Naive Bayes, Decision table, J48 and Logistic Regression are analyzed for predicting Parkinson's disease patients with greater accuracy.

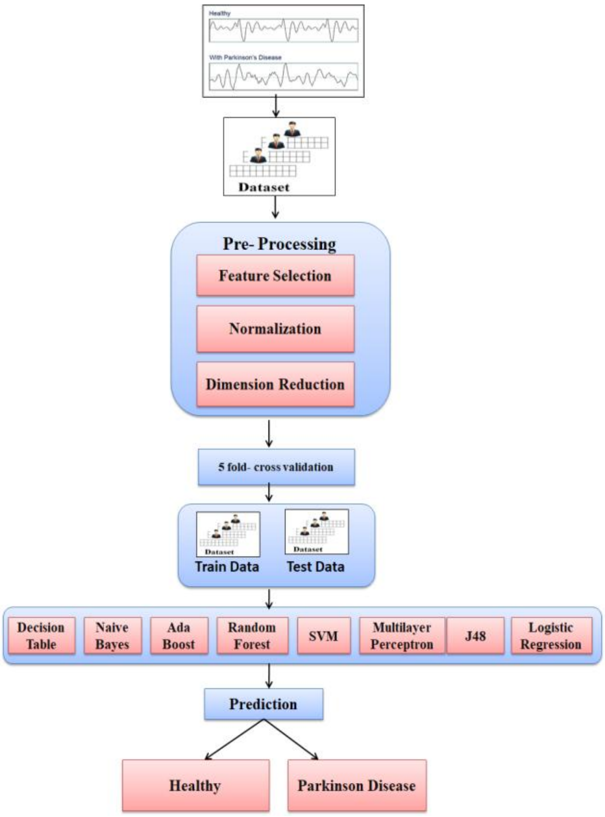

Fig. 1.Architecture of predictive analytics model

This architecture works in three phases; pre-processing, modeling and prediction. Initially, the dataset collected from patients are pre-processed. During preprocessing data relevant to process are identified using feature selection technique. Normalization converts data into a uniform scaling. Dimension reduction techniques can be applied to reduce the dataset to manage scalability problem. Second, pre-processed datasets are then modeled using classifier algorithms. Finally, the test datasets are used to predict the disease earlier.

A. Naïve Bayes

Bayesian classification is based on Bayes Theorem. Bayesian classifiers are the statistical classifiers. Bayesian classifiers can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Firstly, in general, the result P(A|B) is referred to as the posterior probability and P(A) is referred to as the prior probability.

- P(A|B): Posterior probability.
- P(A): Prior probability

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) is supervised learning models that can be associated with learning algorithm to analyzing the data used for regression analysis and classification. SVM algorithm construct hyper plane in infinite dimensional space. SVM is effectively perform a non-linear classification. It categorizes for unlabeled data used widely in clustering algorithm in industrial application

C. Random Forest

Random Forest is an Ensemble Learning method for classification and regression. It is like a Bootstrapping algorithm with Decision tree model. Random Forest tries to construct a multiple CART models with a different initial variables and different samples. Random Forest gives more accurate predictions. It is a classifier which consist of a tree structured classifier collection.

D. Ada Boost

Ada Boost "Adaptive Boosting" is a machine learning meta algorithm. It can be used in conjunction. Adaboost is an adaptive sense that subsequent weak learners. It is sensitively noisy data and outliers. Adaboost refers to a training a booster classifier in particular method. It focuses on classification problems and aims to convert a set of weak classifiers into a strong one. Adaboost is a successful booting algorithm that developed for binary classification understanding booting is the best starting point.

E. Decision Table

Decision Table concise for visual representation to perform action in the given condition in the table classification. The table corresponds variable relation and predicate the possible values among the alternative conditions. A Decision table is balanced or complete consideration. Input variables combination is included in the decision table. An Output is a set of action in the Decision table algorithm.

F. J48 Classifier

J48 is an algorithm that used to generate a decision table classification. It is referred as statistical classifier. A dataset with a predictions or independent variables list and a dependent variables or targets list. The J48 classifier is implantation of ID3 algorithm. The algorithm used to create univariate decision table.

G. Logistic Regression (LR)

Logistic Regression is a method of statistical algorithm this method is used to predict a data value that based on prior observation of a dataset. Logistic Regression is an important tool. It predicts a dependent data variable by analyzing the relationship between one or more existing independent variables. It plays vital role in data preparation activities.

IV. EXPERIMENTAL SETUP

A. Software tool used

In this work WEKA software tool is used to test and simulate the prediction model using real-word datasets in in i7 processor with windows 10 Operating system. This tool is designed at University of Waikato in the country of New Zealand. It is data mining software that uses a collection of machine learning algorithms. These algorithms can be applied directly to the data or called from the Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

B. Dataset Description

The Parkinson's dataset is collected from UCI machine learning repository [11] is used to test the performance of classifiers.

This dataset composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease. There are six voice records of each patient with total of 195 from 48 Negative and 147 Positive samples. It contains 22 input attributes such as Signal fractal scaling exponent, ratio of noise, jitter, vocal fundamental frequency and shimmer values and one output attributes for target class. There are no missing values identified in the dataset. The fig.2 shows the five-fold cross validation is performed on the PD dataset.

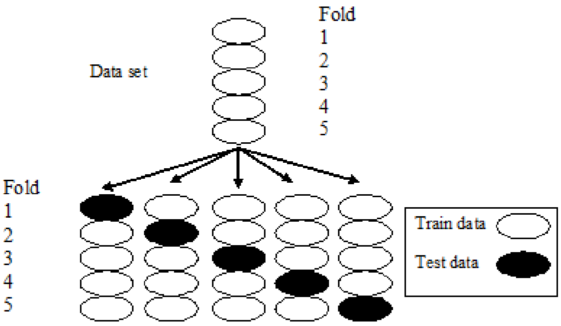

Fig. 2 Five-fold cross validation

C. Evaluation Measures

The prediction model is evaluated based on measures like Recall, Precision and F Measure from the number of True Positives (TP), False Positives (FP) and False Negatives (FN) obtained.

Precision = TP / (TP + FP) (1)

Recall = TP / (TP + FN) (2)

F Measure= (2 x Precision x Recall)/(Precision+Recall) (3)

Kappa statistic = $k = (P_o - p_e) / (1 - p_e)$ (4)

where P_o observed agreement among the raters, p_e hypothetical probability of the raters indicating a chance agreement. MAE can be defined as the following:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \quad (5)$$

Where predicted value can be represented as $\{p_1, p_2, ..., p_n\}$, its corresponding true value can be represented as $\{q_1, q_2, ..., q_n\}$. The Root Mean Square Error (RMSE) can be calculated as;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}} \quad (6)$$

Root Relative Squared Error (RRSE) can be calculated as follows;

$$E_i = \sqrt{\frac{\sum_{j=1}^{n} (P_{(ij)} - T_j)^2}{\sum_{j=1}^{n} (T_j - \bar{T})^2}} \quad (7)$$

where $P_{(ij)}$ is the value predicted by the individual model i for record j (out of n records);

T_j is the target value for record j; and $\bar{T}$

V. RESULTS AND DISCUSSION

The classifier algorithms are simulated and experimentally evaluated using datasets and their performance using various measures such as confusion matrix and statistical measures are observed.

A. Confusion Matrix

The confusion matrix is used to test the performance classifier when its true values are known in advance. The fig. 3 shows confusion matrix obtained from the experimental results using various classification techniques.

Decision Table	P	N
P	134	13
N	19	29

Naïve Bayes	P	N
P	91	56
N	4	44

Ada Boost	P	N
P	137	10
N	19	29

Random Forest	P	N
P	145	2
N	22	26

SVM	P	N
P	145	2
N	21	27

MLP	P	N
P	139	8
N	20	28

J48	P	N
P	129	18
N	20	28

LR	P	N
P	135	12
N	14	34

Fig 3. Confusion Matrix of various classifiers.

The result show that SVM and Random Forest classifier outperformed with 145 True Positives out of 147 positive samples. SVM and Random Forest gives 27 and 26 False Negatives respectively from 48 negative samples.

B. Classification Errors

The Table I. shows result obtained thru various accuracy measures namely Correctly Classified Instances, Incorrectly Classified Instances, Kappa Statistic, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Root Relative Squared Error (RRSE) and Total Number of Instances for the various classifiers. The SVM classifier result in less MAE with 0.1179. The kappa statistic value of kappa 40-59 is poor, 60-79 is moderate, 80-90 is strong and above 90 is perfect. Here, SVM gives maximum kappa value of 0.6333 and minimum RAE of 31.65.

C. ROC and PRC Curve

A Receiver Operating Characteristic Curve (ROC) it is a graphical plot that illustrate the ability of a binary classifier system threshold. The curve created to plot the positive rate against the false positive rate of various classifier threshold. ROC curve is the comparison of two operating characteristics True Positive Rate (TPR) and False Positive Rate (FPR) as the criterion charges.

The Precision and Recall Curve (PRC) gives the values between the TPR and Positive Predicted Value PPV. Fig. 4.a shows results of positive threshold curve and 4.b shows negative threshold curve of SVM classifier. The x axis shows false positive rate and y axis shows true positive rate.

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-7, May 2020

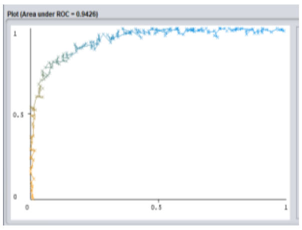

Fig. 4. (a) ROC Positive

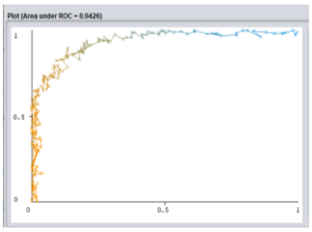

Fig 4.(b) ROC Negative

Fig. 5. Shows the comparison of ROC and PRC values obtained using various classifiers.

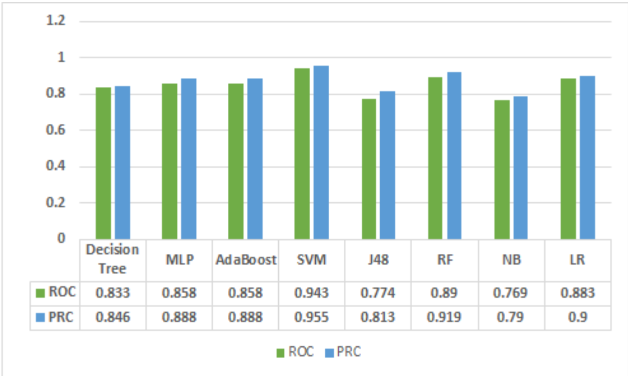

Fig. 5 Comparison of ROC and PRC on classifiers

The SVM classifier gives high values 0.943 and 0.955 for ROC and PRC respectively.

D. Prediction Results

The classifier models are simulated with dataset for testing positive and negative separately. The average of positive and negative results using various classifiers are tabulated. Table II shows average Precision, Recall, F measure and Matthews Correlation Coefficient (MCC) calculated from various classifiers namely Decision Table, Naive Bayes, Ada Boost, Random Forest (RF), Support Vector Machine (SVM-SMO), Multilayer perceptron (MLP), J48, Logistic Regression (LR). The MMC has values between -1 and 1. MCC has -1 value fully wrong binary classifier and 1 indicates fully correct binary classifier. The SVM method gives the maximum MCC value as 0.665 which indicates high accuracy.

Table- I: Classification Errors on various classifiers

	Classification Algorithms							
Measure	DT	NB	AB	RF	SVM	MLP	J48	LR
Total Instances	195	195	195	195	195	195	195	195
Correctly Classified Instances	163	135	166	171	172	167	157	169
Incorrect Classified Instances	32	60	29	24	23	28	38	26
Kappa Statistic	0.538	0.3925	0.5723	0.6142	0.6333	0.5775	0.4674	0.63
Mean Absolute Error (MAE)	0.2368	0.3079	0.1743	0.2433	0.1179	0.1601	0.2019	0.179
Root Mean Squared Error (RMSE)	0.3629	0.5453	0.3328	0.3097	0.3434	0.3642	0.4265	0.3364
Relative Absolute Error (RAE)	63.54	82.61	46.76	65.29	31.65	42.96	54.16	48.04
Root Relative Squared Error (RRSE)	84.21	126.56	77.23	71.88	79.70	84.52	98.99	78.08

Table- II: Accuracy on classifiers

	Measure			
Classifier	Precision	Recall	F-Measure	MCC
Decision Tree	0.830	0.836	0.832	0.540
Naïve Bayes	0.830	0.692	0.713	0.462
AdaBoost	0.830	0.692	0.713	0.462
Random Forest	0.883	0.877	0.865	0.649
SVM-SMO	0.888	0.882	0.871	0.665
Multilayer Perceptron	0.850	0.856	0.849	0.587
J48	0.802	0.805	0.804	0.468
Logistic Regression	0.865	0.867	0.866	0.636

The predictive analytics result shows that among these classifiers, SVM-SMO classifier outperforms with highest precison of 0.888, best recall value of 0.882 and F measure 0.871 when compared with other classifiers.

VI. CONCLUSION

Parkinson is a neurodegenerative disease which leads to the so many risk factors to the patients when it is not predicted earlier. Hence in this work the novel predictive analytics model is presented. These classification and prediction analytics models are analyzed for prediction of disease earlier using Parkinson's disease dataset.

The results show that SVM method provides high accuracy 0.871 in prediction when compared with other methods discussed. The utilization of Dee Learning Machine, Extreme Learning Machine classier and other optimization techniques such as Ant colony and Particle Swarm Optimization can be considered as a future work for improving the accuracy of disease prediction. Furthermore, real-world datasets collected from hospitals using sensors such as triaxial accelerometers, magnetometer and gyroscopes for PD and other chronic condition diseases will be tested for performance benchmark.

ACKNOWLEDGMENT

This research work was carried out with the financially support of RUSA-Phase 2.0 grant sanctioned vide Letter No. F24-51 / 2014-U, Policy (TNMulti-Gen) Dept. of Edn. Govt of India, Dt.09.10.2018 at Alagappa University, Karaikudi, Tamilnadu, India.

REFERENCES

1. Indira Rustempasic, Mehmet Can, Diagnosis of Parkinson's Disease using Fuzzy C-Means Clustering and Pattern Recognition, *Southeast Europe Journal of Soft Computing*, pp. 44-49, March 2013.
2. Prabhu, P., Selvabharathi, S. Deep Belief Neural Network Model for Prediction of Diabetes Mellitus. *In 2019 3rd International Conference on Imaging, Signal Processing and Communication, ICISPC 2019* (pp. 138–142). Institute of Electrical and Electronics Engineers Inc, 2019.
3. Basil K Varghese, Geraldine Bessie Amali D, Uma Devi K S. Prediction of Parkinson's Disease using Machine Learning Techniques on Speech dataset. *Research J. Pharm. and Tech 2019*; 12(2):644-648.
4. Almogren, A. An automated and intelligent Parkinson disease monitoring system using wearable computing and cloud technology. *Cluster Computing* 22, 2309–2316 (2019). https://doi.org/10.1007/s10586-017-1591-z.
5. Ramzi M. Sadek, Salah A. Mohammed, Abdul Rahman K. Abunbehan, Abdul Karim H. Abdul Ghattas, Majed R. Badawi, Mohamed N. Mortaja, Bassem S. Abu-Nasser, Samy S. Abu-Naser, Parkinson's Disease Prediction Using Artificial Neural Network, *International Journal of Academic Health and Medical Research (IJAHMR)* Vol. 3 Issue 1, January – 2019, Pages: 1-8.
6. Meigal, A. Y., Korzun, D. G., Moschevikin, A. P., Reginya, S., & Gerasimova-Meigal, L. I. (2020). Ambient Assisted Living At-Home Laboratory for Motor Status Diagnostics in Parkinson's Disease Patients and Aged People. In S. Balandin, & E. Balandina (Eds.), Tools and Technologies for the Development of Cyber-Physical Systems (pp. 176-201). Hershey, PA: *IGI Global*. doi:10.4018/978-1-7998-1974-5.ch007.
7. VojtechIllner, PavelSovka, JanRusz, Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease, *Biomedical Signal Processing and Control* Volume 58, January 2020, 101831.
8. Cynthia M.fox and lorranie Olson ramig. Vocal Sound Pressure Level and Self-Perception of Speech and Voice in Men and Women with Idiopathic Parkinson Disease, *American Journal of Speech-Language Pathology*, vol 6. No.2 pp.85-94 May 1997.
9. Khan, T., Lundgren, L., Anderson, D. G., Novak, I., Dougherty, M., Verikas, A., Aharonson, V. (2019). Assessing Parkinson's disease severity using speech analysis in non-native speakers. *Computer Speech & Language* (Print), 61. Published. https://doi.org/10.1016/j.csl.2019.101047.
10. Paulraj Prabhu., Neelamegam Anbazhagan. Improving Business Intelligence Based on Frequent Itemsets Using k-Means Clustering Algorithm. In: Meghanathan N., Nagamalai D., Rajasekaran S. (eds) Networks and Communications (NetCom2013). *Lecture Notes in Electrical Engineering, vol 284. Springer, Cham,* pp 243-254 2014.
11. Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig, Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, *IEEE Transactions on Biomedical Engineering*,2008.

AUTHORS PROFILE

P. Selvashankari is pursuing M.Phil Computer Science, Department of Computer Applications, Alagappa University, Karaikudi, India. She has presented research papers in various International and National conferences.

Dr. P. Prabhu is working as Assistant Professor in Information Technology, Directorate of Distance Education, Alagappa University, Karaikudi, India. His research areas include Data Mining, Machine Learning, Neural Networks and Bigdata Analytics. He has published good number of Publications in Reputed Journals. He is Reviewer Member for various Reputed peer-reviewed journals. He is a life member of Indian Science Congress Association, Kolkatta, India.