Parkinson's Disease Prediction Using Multiple Feature Evaluation Approach and Hadoop Framework

D. Siva Sankara Reddy[1] Dr. R. Udaya Kumar[2]

[1]*Research Scholar, Department of Computer Science and Engineering, BIST, Bharath Institute of Higher Education and Research (BIHER), India*

[2]*Research Supervisor, Professor, Department of Information Technology, BIST, Bharath Institute of Higher Education and Research (BIHER) Institution, India*

Abstract - Parkinson's disease (PD) is a critical neurological disorder issue and affects nervous system of people. Over the past many years, the researchers have studied on voice data set to predict PD. It eliminates the patient physically presence at clinic to submit samples. In this paper, the disabilities are illustrated that cause for PD voice disorder and extracts parameters from Voice Signals. In the proposed model, Multiple Feature Evaluation Approach (MFEA) is designed using Principle Component Analysis (PCA) to select best features from recorded patient speech signals, Classification and Regression Tree (CART) is used for best classification and Hoeffeding Tree (HD) is used to find PD status. The proposed PCA can reduce the dimensionality of the feature set with minimal loss of useful information and highly scalable for Big Data sets. The proposed method has obtained the results 98.7% accuracy.

Keywords: Parkinson's Disease, Voice Signals, Multiple Feature Evaluation Approach, Principle Component Analysis, Classification and Regression Trees, Hoeffeding Tree, Big Data

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder which affects the human motor system. It is movement disorder of 60 years of age people and affects nerve cells in brain. It is too difficult to diagnose the disease medically and also costly[1]. Many people are doing research to find inexpensive PD detection techniques. The PD can diagnose with behavioral patterns such as rigidity, tremors, bradykinesia, cognitive impairment and postural instability. PD can be controlled after early diagnosis but not curable. The life expectancy of PD patients falls between 8 to 15 years after diagnosing. The misdiagnosis of PD rate is more than 25% because of death rate increases once after PD was identified and not proper mechanism to give treatment at early period[2]. The PD patient life period can be increased by using more accurate detection methods. The PD patients can maintain high quality of life by using PD detection methods and the necessary pharmacological and surgical intervention [3].

Many researchers have been developed telediagnosis and telemonitoring systems for detecting PD at early stage. Most of the researchers are concentrating in their research to identify biological markers in the PD detection phase. The Machine Learning is thrived in PD diagnosis. The various researchers were used neuroimaging modalities for extracting PD features based on MRI images [4]. The PD detection systems can avoid visiting clinical centers for patients and also reducing work of health workers and clinicians. So many symptoms are appeared in PD patients such as dysphonia, deficiencies, balance, slowed movement, posture etc. The vocal dysfunction is resulted damages vocal quality, vocal instability and loudness [5]. More than 90% of PD patients can get more vocal issues in the beginning of disease. So, the disease can be analyzed based on speech signals at early stage. The recent research on PD detection is emphasized on the vocal disorders of patients. The various classification models with key functions of speech signal processing algorithms are designed for classifying the extracted features based on recorded voice data set. The clinical evaluation system can predict PD based on speech recordings [6]. The clinical evaluation system is mapped with the vocal features of PD using Tele-monitoring system. The speech data can be collected from PD patients and provided as input to the Tele-diagnosis system.

The PD classification has been done on vocal issues using ML techniques such as Random Forest (RF), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost) and Artificial Neural Networks (ANN) [7]. The quality of data, extracted features with relevance and Machine learning methods play very important role in the success rate of PD detection. Many researchers were conducted research on voice data set that consist 23 PD affected instances and 8 healthy instances out of 195 data samples. The features of data set are extracted such as measures of the variation in amplitude, vocal fundamental frequency, measures of the ration of the noise-to-tonal components, measures of the variation in fundamental frequency etc. These are baseline features which are extracted from data set [8]. For detecting PD, the Mel-frequency Cepstral Coefficients (MFCC) and Signal-to-Noise Ratio (SNR) are used. In the recent PD studies, the vocal features were analyzed effectively. The PD is detected easily and accurately using vocal and speech signal processing methods based on voice variations and various symptoms of PD such as tremors, slow movement, balance of deficiency and inertia [9]. In our study, we were adopted feature selection and sparse code classifier to predict the PD.

II. RELATED WORK

To diagnose PD in the early period, the detection methods must be designed carefully. The earlier researchers were conducted research on designing automated tools for PD detection. The various researchers were identified voice data set tool for predicting PD based on different causes. Most of the research prioritizes voice data as tool for detecting PD based on different voice abnormalities. The feature selection methods have been summarized on voice data set for PD classification.

Gupta et al. [10] proposed a method OCFA (Optimized Cuttlefish Algorithm) for selecting best features to predict Parkinson's Disease. It is modified version of Cuttlefish algorithm. It is used to select combination of features than single feature as used in Cuttlefish algorithm. The algorithms are applied on voice data set independently for evaluation of its features. This method is proved better results than the other methods.

Sharma et al. [11] proposed an approach MGWO based on feature selection process to predict Parkinson's Disease. In this approach, the features are selected independently. The authors are tested by combining their feature selection algorithm with different classifiers such as KNN, DT and RF. The MGWO is provided better results than OCFA.

Mostafa et al. [12] proposed a multi-agent system for diagnosing PD based on feature selection process. In this, the 11 features are selected from 23 features in voice data set. The authors are used different classifiers such as Naive Bayes, DT, RF, MLP and RBF-SVM in this model to train the data set with different features. The authors were tested using different metrics such as precision, recall and accuracy. The produced results are proved in best way.

Sakar et al. [13] proposed an ensemble model with different classifiers and also filtering based feature selection algorithms for detecting PD on voice data set. This model includes various classifiers such as RBF-SVM, LR, SVM, RF, Navie Bayes and KNN. The authors were tested model performance using various classifiers and it is proved the results.

Elmehdi Benmalek et al. [14] proposed a system that works on speech data set to predict PD in the initial stage to increase the patient life period. In this model, the various feature ranking algorithms and SVM were used to implement the system. The authors are tested the ranking algorithms independently and produced good performance than other heuristics.

Ali et al. [15] proposed a voice-based prediction model for predicting PD based on feature & hyper-parameter selection algorithm. These algorithms are sorted the features using $\chi 2$-score method. The selected features are determined by the validation loss.

Gunduz et al. [16] proposed a convolutional neural network (CNN) model for detecting PD using voice data set. In this model, the various features are selected with combinations from given data set. The results are produced in best way by applying CNN and SVM on Tunable Q-factor Wavelet Transform (TQWT), Mel-frequency cepstral coefficients (MFCC). The features are generated by concatenating the vocal fold, baseline and time-frequency features.

Emamzadeh, F. N et al. [17] proposed a system using support vector machine, Bayesian, neural network and decision tree for predicting PD on voice data set. The authors were implemented, tested and produced the results as 86.2%, 58.6%, 94.8% and 74% for support vector machine, Bayesian, neural network and decision tree respectively. It is achieved good performance than the other methods.

Mostafa, S. A et al. [18] proposed an approach using several classifiers for predicting PD. This system includes neural network, decision tree and business naïve. The proposed approach was produced accuracies as 87.55%, 86.44% and 74.11% for neural network, decision tree and business naïve respectively.

III. MATERIALS AND METHODS

3.1 Data Set Description

The PD voice data set is collected from University of California Irvine (UCI) machine learning repository system. The data set contains 195 instances of sound recordings of 31 people. The 23 people were affected with PD from 31 people. The age of these people falls between 46 to 85 years old. The voice measures are represented as each column in the table and the 195 voice recordings of each individual are represented in each row which is shown in Table1 [19]. In the table, the status of PD people and healthy people are set to 1 and 0 respectively.

S. No	Attribute ID	NAME of Attribute	DESCRIPTION of Attribute
1	A1	MDVP: Fo (Hz)	Kay Pentax MDVP Average Vocal Fundamental Frequency
2	A2	MDVP: Fhi (Hz)	Kay Pentax MDVP Maximum Vocal Fundamental Frequency
3	A3	MDVP: Flo(Hz)	Kay Pentax MDVP Minimum Vocal Fundamental Frequency
4	A4	MDVP: Jitter (%)	Kay Pentax MDVP Jitter as Percentage
5	A5	MDVP: Jitter(Abs)	Kay Pentax MDVP Absolute Jitter in Microseconds
6	A6	MDVP: RAP	Kay Pentax MDVP Relative Amplitude Perturbation
7	A7	MDVP: PPQ	Kay Pentax MDVP 5 Point Period Perturbation Quotient
8	A8	Jitter: DDP	Average Absolute Difference of Differences between Cycles, Divided By the Average Period.
9	A9	MDVP: Shimmer	Kay Pentax MDVP Local Shimmer
10	A10	MDVP: Shimmer (dB)	Kay Pentax MDVP Local Shimmer in Decibels
11	A11	Shimmer :APQ3	3-Point Amplitude Perturbation Quotient
12	A12	Shimmer :APQ5	5-Point Amplitude Perturbation Quotient
13	A13	MDVP: APQ	Kay Pentax MDVP 11 Point Amplitude Perturbation Quotient
14	A14	Shimmer : DDA	The average absolute difference between consecutive differences between the amplitudes of consecutive periods.
15	A15	NHR	Noise to Harmonic Ratio
16	A16	HNR	Harmonic to Noise Ratio
17	A17	RPDE	Recurrence Period Density Entropy
18	A18	D2	Correlation Dimension
19	A19	DFA	Detrended Fluctuation Analysis
20	A20	Spread1, Spread2	Non-Linear measures of Fundamental Frequency variation
21	A21	PPE	Pitch Period Entropy
22	A22	Status	Health status of a subject, 1-Parkinson's Disease, 0-Healthy

Table1. PD Data Set with Features vector

3.2 Proposed Methodology

As mentioned earlier, the PD is neurodegenerative disease and grows chronically.

It creates problems in the movement and unable to do the regular activities and creates other consequences. The proposed approach MFEA is designed to predict the PD. This method consists of several processing elements such as feature selection and feature classification. The features are selected using PCA in proposed approach. The PD is more accurately analyzed to reduce the size of the feature vector and to select the best features. Among the existing 22 features, the selected 10 features are performed good accuracy in predicting PD after conducting various experiments. This process can reduced the storage space during the diagnosing of PD. The PCA is the best method to diagnose PD based on different type of voice signals. It is very important to identify features, extract and diagnose signal information by displaying signals on a specific basis. The **Figure1** describes the proposed method.

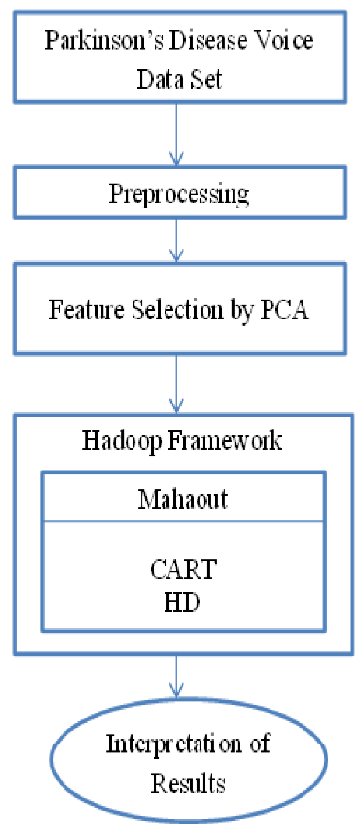

Figure1. Block diagram of proposed model: MFEA

Algorithm: MFEA
Step1: considering initial parameters
Step2: Doing task for every active a_j
Step3: X ← preparing (a_j, C)
Step4: a_j do i until n
Step5: $f_{i,j}$ ← evaluating(a_j, x_i)
Step6: $r_{i,j}$ ← ranking(a_j, x_i)
Step7: $x_{i,j}$ ← searching(a_j,x_i)
Step8: End-do
Step9: X' ← selecting(a_j, $x_{1,2,..j}$)
Step10: End-for
Step11: X' ← collaborate(t updating($a_{1,2,.....t}$), {x^{in}, x^{ex}} filter($a_{1,2;..}$ $x'_{1,2;..}$));
Step12: End

3.3 Principle Component Analysis

PCA is a statistical procedure that it extracts the features from the data set. It is an efficient dimensionality reduction model that is capable of diminishing the wavelet coefficients size from speech signals. This method can combine correlated variables to create an artificial set of variables with small size [20]. The PCA algorithm steps are described as follows.

Algorithm: PCA

Input: UCI Machine Learning Repository Data Set

Output: Reduction of Feature Set

Step1: Y ← Data Matrix with size P x n and vector Y_k per data point.

Step2: Subtracting mean y from Y in each row vector Y_k in Y.

Step3: ∑ ← Generate covariance matrix for Y.

Step4: Generate Eigen vectors and Eigen values for ∑.

Step5: Principle Components ← the Q Eigen vectors with biggest Eigen values.

Step6: Result: Principle Components.

3.4 Framework of Hadoop

The framework of Hadoop is open source software which is used to store and process Big Data using Big Data Analytics. It can be used for improving performance on different clusters. Hadoop is used to perform operations effectively on large data sets which comprises of multiple concepts and modules like Map-Reduce, HDFS, HBASE, Mahaout, HIVE, PIG, SQOOP and ZOOKEEPER. Hadoop produces highest value by using high velocity, high volume and high variety of data [21].

3.5 Mahaout

Apache Software Foundation was developed Mahaout which is a part of Hadoop that it uses MapReduce model. It can use for building scalable applications using machine learning algorithms which are focused in the areas of collaborative, clustering and classification. It contains various Java libraries and Java collections to perform statistical operations [ ].

3.6 Classification and Regression Trees (CART)

The CART is a predictive mechanism which is used in Machine Learning. It describes the prediction of target variable's values based on different values. It is a decision tree algorithm which works on target variable. It is basis for boosted decision trees, random forest and bagged decision trees. It describes the pseudo code of the CART algorithm. The classification model is constructed after reading training data by CART algorithm.

Algorithm: CART
Input: trained samples, no of input values
Output: Decision Tree Model
Step1: Reading trained samples
Step2: Taking total possible values of total variables
Step3: Evaluate for every possible value of total variables
Step4: Consider variable to get maximum Gini score
Step5: If X< s then transfer data to left node of tree otherwise transfer data to right node
Step6: Next iterate this process on two nodes of tree
Step7: Return Decision Tree Model

3.7 HOEFFDING TREE (HT)

Hoeffding Tree is an induction algorithm which is used to take decision and with self-learning capabilities using huge data streams. It exploits the features as optimal splitting attribute. It supports mathematical operations and quantifies needed observations to estimate statistics within fixed boundaries.

Algorithm: Hoeffding Tree
Inputs: P is a sequence of samples
R is a set of discrete attributes
F (.) is an evaluation function for splitting
δ is correct attribute with desired probability at any given node
Output: HT is a decision tree
Procedure HoeffdingTree (P, R, F, δ)
Let HT be a tree with a single leaf l_1 (the root).
Let R_1 = R ∪ {Rφ }.
Let F'(Rφ) be the F' obtained by predicting the most frequent class in P.
For each class y_k
For each value x_{ij} of each attribute $X_i \in X$
Let $n_{ijk}(l_1) = 0$.
For each example (x, y_k) in S
Sort (x, y) into a leaf l using HT
For each x_{ij} in x such that $X_i \in X_l$

Increment $n_{ijk}(l)$.
Label l with the majority class among the examples seen so far at l.
If the examples seen so far at l are not all of the same class, then
Compute $G'(X_i)$ for each attribute $X_i \in X_l - \{X\phi\}$ using the counts $n_{ijk}(l)$.
Let X_a be the attribute with highest G'.
Let X_b be the attribute with second-highest G'.
Compute € using Equation 1.
If $G'(X_a) - G'(X_b) >$ € and $X_a \neq X\phi$, then
Replace l by an internal node that splits on X_a.
For each branch of the split
Add a new leaf l_m, and let $X_m = X - \{ X_a \}$.
Let $G'_m(X\phi)$ be the G' obtained by predicting the most frequent class at l_m.
For each class y_k and each value x_{ij} of each attribute $X_i \in X_m - \{X\phi\}$
Let $n_{ijk}(l_m) = 0$.
Return HT

IV. EXPERIMENT AND RESULTS

Hadoop framework is used to perform experiments for storing, processing and analyzing data set. The data set is injected as input to HDFS system and Mahaout to analyze data features which is taken from UCI. The CART and HD are compared with existing methods for prediction of PD. In This paper, we processed data pre-processing and divided into training and test data in the ratio 3:1. Then, it is applied to PCA for selecting best features to predict the results accurately.

The performance of system is measured using performance measuring tools such as Recall, Accuracy, F-measure, Precision, and processing time. The results are shown in Table2.

Classification Models	AUC (%)	CA (%)	F1-Score (%)	Precision (%)	Recall (%)	Specificity (%)
CART	97.8	96.2	96.8	95.0	97.6	95.5
HD	96.9	94.9	96.4	94.9	93.9	95.1
KNN	95.3	89.2	88.8	89.0	89.2	75.4
DT	96.4	95.5	95.5	96.5	94.5	96.7
SVM	95.7	90.3	89.4	91.4	90.3	70.2
RF	96.9	94.9	97.4	94.9	93.9	95.1
NN	94.4	96.4	96.3	93.5	96.4	90.4

Table2. Results obtained with different methods and classifiers.

As shown in Table2, the proposed method was produced good results in PD prediction by different features.

Features	Accuracy (%)
14	97.18
13	*96.89*
12	96.31
11	96.24
10	98.67
9	94.46
8	93.45

Table3. Comparison of the proposed method with different features in the UCI database

Figure2 shows the comparison of different classifiers performance on best features. The proposed approach was shown good performance for classification accuracy than other machine learning techniques.

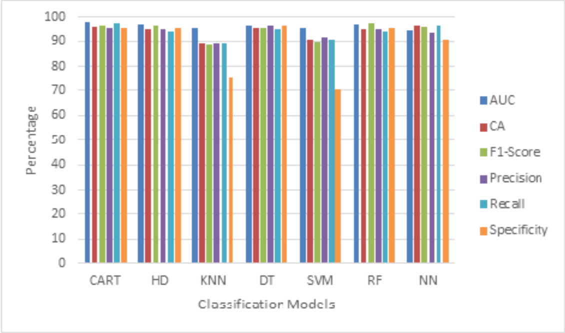

Figure2. Comparison of different classifiers performance

Figure3 shows the performance of accuracy with different features. The proposed approach was shown good performance of classification accuracy with different features.

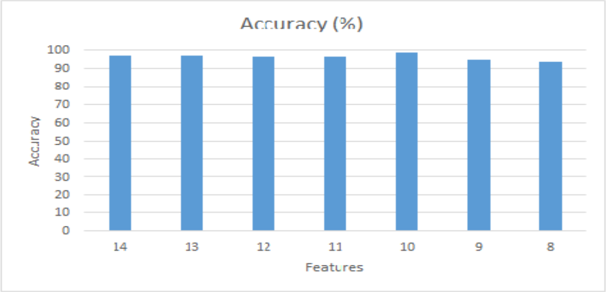

Figure3. Performance of classification accuracy with different features

Figure4 shows the performance of accuracy of different classifiers for different population. The diagnostic test probability for PD patients was identified with values of specificity and sensitivity. In Figure4, the high accuracy of classifier is shown with low specificity and highest sensitivity. All the classifiers were produced the result effectively for each population. For making better decision, the highest accuracy of classifiers can consider from total populations.

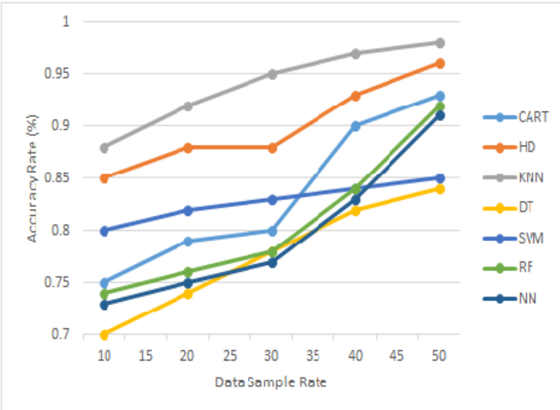

Figure4. Accuracy Rate as increasing samples

Figure5 shows the plotted for the proposed approach with ROC curve metric. The inverse linear relationship is depicted with ROC curve and sensitivity decreases as specificity increases [24]. The graph is described that the false negative rate decreases as true positive rate increases. It shows the predictive accuracy with high probability against PD and the accuracy and consistency are the main key factors.

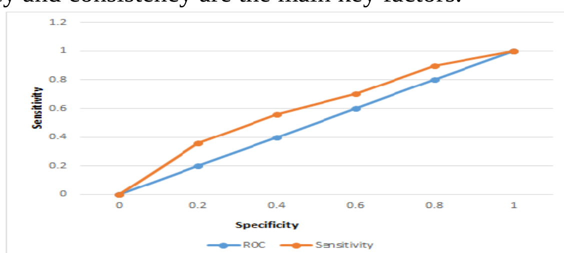

Figure5. ROC analysis

4.1 Performance Metrics

The performance and efficiency of proposed method is evaluated using the accuracy criterion.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

True Negative (TN): It has been correctly identified as state and whose true category is incorrect.

True Positive (TP): It has been correctly identified as state and whose true category is correct.

False Positive (FP): It has been misdiagnosed and that actual category is correct.

False Negative (FN): It has been correctly identified as state and whose true category is incorrect.

To evaluate the accuracy of the test dataset using proposed approach, the F-Measure criterion was used. The confusion matrix is used in this process that includes TP, FP, FN and TN. The F-Measure is computed using the following tool:

$$F - measure = \frac{2 \times Precision \times Recall}{Precison + Recall}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

V. CONCLUSION

In this research work, the feature selection and PCA techniques are used for diagnosing PD patients on voice signals data set. The PCA is used for reducing the dimensions of the feature. It improves accuracy and provides efficient features for diagnosing PD. The ten features are taken to each sample from 22 features by applying reduction in feature vector process. The main aim of this process is used to estimate the status of PD with different tests. The CART and HD classifiers are used to classify the data into two classes such as healthy people and PD affected people. The storage space is reduced by considering ten features for each sample in the proposed method. The performance of PD detection is increased by reducing the storage space using PCA. In the proposed method, the best accuracy is 97.11% on UCI data set. In future, this work can be extended for analyzing multi-variant data set for achieving 100% accuracy in the PD prediction.

VI. REFERENCES

[1] Gayathri Nagasubramanian and Muthuramalingam Sankayya, "Multi-variate vocal data analysis for detection of Parkinson disease using deep learning," Neural Computing and Applications, 2020.

[2] Indira Rustempasic And Mehmet Can, "Diagnosis of Parkinson's Disease using Principal Component Analysis and Boosting Committee Machines", Southeast Europe Journal of Soft Computing, Vol. 2(1), Pp. 102-110, 2015.

[3] Muntasir Hoq, Mohammed Nazim Uddin and Seung-Bo Park, "Vocal Feature Extraction-Based Artificial Intelligent Model for Parkinson's Disease Detection, MDPI (Multidisciplinary Digital Publishing Institute), vol. 11(6), pp. 2-22, 2021.

[4] Kemal Akyol, Safak Bayır, Baha Sen, "Importance of Attribute Selection for Parkinson Disease", Academic Platform Journal of Engineering and Science, vol. 8(1), pp. 175-180, 2020.

[5] S. Kanagaraj, M.S. Hema, M. Nageswara Gupta, "Machine learning techniques for prediction of Parkinson's disease using Big Data," International Journal of Innovative Technology and Exploring Engineering, vol. 8(10), pp. 3788- 3791, 2019.

[6] Neharika D Bala, Anusuya S, "Machine learning algorithms for detection of Parkinson's disease using motor symptoms: speech and tremor," International Journal of Recent Technology and Engineering, vol. 8(6), pp. 47-50, 2020.

[7] R. M. Sadek, S. A. Mohammed, A. R. Abunbehan, A. K. Ghattas, M. R. Badawi, M. N. Mortaja, B. S. Abu-Nasser, and S. S. Abu-Naser, ``Parkinson's disease prediction using artifcial neural network," Int. J. Acad. Health Med. Res., vol. 3(1), pp. 1-8, 2019.

[8] Neharika D Bala and Anusuya S, "Machine learning algorithms for detection of Parkinson's disease using motor symptoms: speech and tremor," International Journal of Recent Technology and Engineering," vol. 8(6), pp. 47-50, 2020.

[9] N. Shamli and B. Sathiyabhama, "Parkinson's brain disease prediction using big data analytics," International Journal of Information Technology and Computer Science, vol.6, pp.73-84, 2016.

[10] Deepak Gupta, Arnav Julka, Sanchit Jain, Tushar Aggarwal, Ashish Khanna, Arunkumar N and Victor Hugo C. de Albuquerque, "Optimized Cuttlefish Algorithm for Diagnosis of Parkinson's Disease," Cognitive Systems Research, vol. 52 , pp. 36-48, 2018.

[11] Prerna Sharma, Shirsh Sundaram, Moolchand Sharma, Arun Sharma and Deepak Gupta, "Diagnosis of Parkinson's disease using modified grey wolf optimization," Cognitive Systems Research, vol. 54, pp. 100-115, 2019.

[12] S. A. Mostafa, A. Mustapha, M. A. Mohammed, R. I. Hamed, N. Arunkumar, M. K. Abd Ghani, M. M. Jaber, and S. H. Khaleefah, "Examining multiple feature evaluation and classi_cation methods for improving the diagnosis of Parkinson's disease," Cognitive Systems Research., vol. 54, pp. 90-99, 2019.

[13] Sakar B. E.et al., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," IEEE J. Biomed. Health Inform., vol. 17, no. 4, pp. 828–834, Jul. 2013.

[14] Elmehdi Benmalek, Jamal Elmhamdi and Abdelilah Jilbab, "Multiclass classification of Parkinson's disease using different classifiers and LLBFS feature selection algorithm," Vol. 20(1), pp. 179-184, 2017.

[15] Ali L., Zhu C., Zhou M., and Liu Y., "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," Expert Syst. Appl., vol. 137, pp. 22–28, Dec. 2019.

[16] Hakan Gunduz, "Deep Learning-Based Parkinson's Disease Classification using Vocal Feature Sets," IEEE Access, Vol. 7, Pp. 115540–115551, 2019.

[17] Emamzadeh, F. N and Andrei Surguchov, "Parkinson's Disease: Biomarkers, Treatment, and Risk Factors," frontiers in Neuroscience, vol. 12 , pp. 1-14 , 2018.

[18] Salama A. Mostafa, Shihab Hamad Khaleefah, Mohd Sharifuddin Ahmad and Mazin Abed Mohammed, "Evaluating the Performance of Three Classification Methods in Diagnosis of Parkinson's Disease," pp.43-52, 2018.

[19] Zehra Karapinar Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," Medical Hypotheses, vol.138, pp. 1-5, 2020.

[20] G Uday Kiran and D Vasumathi, "Predicting Parkinson's Disease using Extreme Learning Measure and Principal Component Analysis based Mini SOM," Annals of R.S.C.B, vol. 25(4), pp. 16099-16111, 2021.

[21] Hassan Harb, Hussein Mroue, Ali Mansour, Abbass Nasser and Eduardo Motta Cruz, "A Hadoop-Based Platform for Patient Classification and Disease Diagnosis in Healthcare Applications," vol. 20, pp. 1-20, 2020.
Cheryl Ann Alexander and Lidong Wang, "Big Data Analytics in Heart Attack Prediction," Journal of Nursing and Care, vol. 6(2), pp. 1-9, 2017.