Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 1, pp. 167–176, DOI 10.12694/scpe.v25i1.2314

A HYBRID MODEL: RANDOM CLASSIFICATION AND FEATURE SELECTION APPROACH FOR DIAGNOSIS OF THE PARKINSON SYNDROME

SUMAN BHAKAR[*], MANVENDRA SHEKHAWAT[†], NIDHI KUNDU[‡], AND VIJAY SHANKAR SHARMA [§]

Abstract. Nowadays Parkinson's disease has been discovered that approximately 94% of people suffer from voice disorder problems. A neurodegenerative can identify PD patients through examination and multiple scanning tests. So, it usually takes more time to diagnose the disease at the early stage. Current work has identified that speech disorders can be a significant signal for Parkinson's disease. Therefore, this work proposed a fusion model to identify the speech disorder at the starting stage of the disease. In this process, the author has tested a model with a different pattern of feature selection method as well as classification mode and created a system with the best pattern. For the creation of pattern, three types of feature selection methods namely Chi-square, genetic algorithm and Embedded random forest method and four classifier models such as KNN, Naïve Bayes, SVM, Decision tree and Random Forest have been utilized. To analyze the performance of the system speech public dataset from the UCI repository,the authors applied the combination of the Embedded random feature selection method and random forest classification algorithm provides 97.89% of accuracy. However, this outcome is better than the recent work. The SMOTE is utilized for the balancing of the dataset.

Key words: Machine learning, Parkinson's Disease, Feature Selection, SVM Detecion

1. Introduction. Parkinson's disease is a neurodegenerative syndrome. It is the second disease after Alzheimer's that is slowly developing the neurodegenerative disorder [1]. This disease has affected 7 to 8 million people worldwide. The main source of PD is the damage to the nerve cells i.e., the portion of the brain known as substantia nigra. Nerve cells also produce a liquid called dopamine. Dopamine is the courier between the brain and nervous system. It helps to regulate the body part and body movement. Unfortunately, if nerve cells are injured due to some reason then the ratio of dopamine is minimized in the brain system. So, gradually brain movement control cannot work properly like in a normal person [3]. Both women and men are affected by Parkinson's disease. it cannot cure but can early diagnosed at the early stage, which helps the person for proper treatment and avoids the dangerous situation. Additionally, the ratio of 3:2 men and women are being affected by Parkinson's disease. This disease generally occurs at the age of sixty but it can also occur at the age of fifty [5, 35].

The starting stage of this disease is extremely gentle and unnoticeable. But these gentle symptoms go through a severe stage if predication does not occur at an early stage [6, 4]. The symptoms of PD fluctuate from patient to patient. The evaluation of the disease can be processed by the motor and nonmotor symptoms. The motor symptoms consist of rigidity, tremor, and bradykinesia and non-motor symptoms consist of sensory impairment, change in handwriting, and vocal disturbance. Many researchers have concluded that 90% of patients have shown vocal disturbance at the initial stage of Parkinson's disease.

There are several reasons for the prediction of Parkinson's disease in the initial phase. Especially the neurologists and experts can detect the Parkinson's disease after complete study and frequent scans of the patients. But both techniques are very time-consuming and difficult for patients which are aged above fifty [7, 21]. A physician with a specific degree in this Parkinson's background can diagnose the disease in the early phase through some symptoms. There is no expert doctor at the mountain area. So, there is need to develop

[*]Department of Computer and Communication Engineering, Manipal University Jaipur, India (Suman.bhakar@jaipur.manipal.edu)
[†]Department of Computer and Communication Engineering, Manipal University Jaipur, India
[‡]Department of Computer Science, SKNAU, Jobner, India
[§]Department of Computer and Communication Engineering, Manipal University Jaipur, India (Corresponding author: vijayshankar.sharma@jaipur.manipal.edu)

a model to detect the disease with high accuracy.

Many researchers have utilized many ways i.e. EMG signals, SPECT images, handwritten pictures, gait signals, and MRI images to detect Parkinson's disease [8]. The fluctuations in the speech are known as dysphonia. It is also a symptom of the early detection of Parkinson's disease. The early symptoms of Parkinson's disease can be diagnosed in the initial phase. So most researchers utilized the voice dataset to identify the disease. because it is a very low cost and simple method. Researchers have used many machine learning approaches to identify the disease at an early stage. The machine learning approach uses many pre-processing, and feature extraction approaches to identify the useful features. It also utilizes the classification model [9] to classify the disease and different validation process that is used to help the validation process. Additionally, the pattern of the disease can easily be identified by medical datasets. The pre-processing methods acquired for the balancing and normalization of the datasets. The feature engineering process involves the feature extraction and selection method. The feature selection approachs are used to enhance the accuracy and reduce the computation cost of the system [10, 2].

The major contribution of this articles are :

- The feature selection has been done by three methods such as Chi-Squared, Random forest and Genatic algorithm. The first approach and second approach select 11 features and the third approach select the 5 features out of 23 are selected. The selected feature divides into the training as well as testing datasets. Then these datasets are passed through the different classification methods to find the best detection accuracy result.
- Second, the speech dataset is imbalanced due to 130 out of 180 samples of the details of Parkinson's patients. The SMOTE is utilized to handle the imbalance datasets issues.
- At the last phase performance parameters of the classifier viz Naive Bayes, k-nearest neighbors, and andom forest are analyzed on the complete feature and reduced subsets of features. Finally, it shows the Embedded random forest .algorithm gives the better accuracy i.e. 97 results compare then the existing techniques.

The paper is prepared as tracks. Section 2 defines the overview of the literature review on Parkinson's disease. Section 3 defines the proposed methodology and datasets. The experimental results and their discussion are shown in section 4. The conclusion and future scope are given in section 6.

2. Literature Review. Many Scientists have developed many techniques to diagnose the PD through various speech signals. Little [11] developed a medical specialist model to detect Parkinson's disease.

The authors [12] developed a system to diagnose Parkinson's disease through a clinical specialist scheme. In this system, the half affected patients with Parkinson's disease voice has recorded with vocalization "a" sound for five sec. the dataset contains the recording of three different sounds. The authors utilized the 240*44 i:e row and column ratio dataset for this system. The waveform algorithm is utilized for the five feature extraction processes. The classification has been processed by the Bayesian approach. The system achieves 75.2% accuracy after the process of validation. The authors proposed a solution for dealing with small files, the concept can be used in implementing database with Hadoop and can provide a simple way to store health data [28, 30, 29].

The authors [13] proposed a decision and KNN classification method to detect Parkinson's disease. The useful features have been extracted from the cuttlefish algorithm to optimize the system. The authors utilized the HandPD, voice, and speech datasets. The 92.19% accuracy was achieved by the proposed system.

The authors [14] developed a system to predict the severity rate of Parkinson's disease by using a deep learning model. The rating scale has done by the UPDRS (Unified Parkinson's disease rating score). The performance has been measured through the motors as well as total motor signals. The authors have proved that the motors score the 81.66% of accuracy which is more than the total UPDRS signals.

The authors [15] have utilized the voice signal to detect Parkinson's disease. The feature selection has been proceeded by the eight different ranking parameters. The SVM classification method is utilized for the detection of healthy and Parkinson's disease patients. The Wilcox statistic is processed to get useful information. The system achieved the 92.21% of accuracy through the SVM method.

The authors [8] developed a model to predict PD patients through speech signals. the signals have been recorded by smartphones as well as acoustic cardioids. The Pre-processing has been performed for the identification of the voice and unvoiced signal through unique software. The feature extraction technique is also applied to get 144 features. The classification is processed by the SVM, MLP, and KNN method to predict the PD patients. The experimental results also proved that the acoustic cardioids i.e. audio signals achieved 3% higher accuracy than smartphone signals.

The authors [16] implemented an antlion optimization algorithm to predict Parkinson's disease. The feature selection has been proceeded to get the optimized result. The filtered features are trained with the decision tree and random classification model. The maximum accuracy was achieved at 95.91%.

Researcher [17] in the article developed a model to diagnose Parkinson's disease. In this model, the data utilization is done through speech and Hand PD datasets. To enhance the optimization of the system grey wolf-based optimization feature selection method is utilized. The authors also used the three classification methods such as decision tree, KNN, and random forest. The experimental result has proven 93.87% accuracy in the speech dataset.

The authors [18] developed an adaptive-based optimization algorithm. The authors used speech signals for the detection of the disease. Authors utilized the sparse encounter method for the reduction of dimension. The classification has been processed by eight classifiers.

The authors [19] developed a model based on a spectrogram by using the feature extraction method. To implement this model PC-GITA datasets are utilized. The authors developed three methods to diagnose Parkinson's disease. In this first approach, the signal based on speech was transformed into a spectrogram and applied the ALEXNET for feature selection method to train the CNN model. In the second approach, feature extraction was done by the CNN model. In the last step, spectral and acoustic signals are passed through classifiers and achieved 99.3% accuracy by using the multilayer classification model.

The authors [20] has developed a novel method to diagnose Parkinson's disease and gender redeployment. The singular value decomposition (SVD) is applied for the features extraction process. And after the features extraction method, the feature selection process i.e. neighborhood component analysis (NCA) is applied for the selection of the features. The authors utilized the six-type based model to diagnosis the PD disease. The developed model achieved the 98.41% and 99.21% of accuracy for the detection of Parkinson's disease and gender. however, only one feature selection method is used to extract the features. and there is no criteria to select the feature for optimization of the model. Also, author applied only one machine learning approach for detection of the disease.

The authors [19] The feature selection is processed by the MRMR method to identify the useful features. The authors also applied the 8-classification method to classify the features. The experimental results proved that the hybrid of RFE and XGboost achieved 95.39% of accuracy. Still, they failed to discuss the feature selection criteria such as, at what parameter and how many features are selected for the detection model, and also did not elaborate the degree of severity for speech dataset.

The authors [23] implemented two CNN-based models to detect the disease. In this article UCI based data i.e. speech dataset was utilized. The author used the two frameworks such as feature and model level for the implementation of the model. The model level-based framework achieved 86.9% of accuracy. However, authors did not use feature extraction and processing approach for the optimisation process and also detection of disease detected by two classification models based on CNN. Also did not compare with another existing machine learning model.

The authors [20] utilised speech signals dataset . The SMOTE is utilized for the pre-processing steps. The authors applied the random forest classification. Although system achieved 89 percentage of accuracy but there is a need to develop a method to handle the unbalanced dataset.so, there is need a method to handle the imbalance dataset.

The authors [22] illustrated the work for the diagnosis of the disease. In this process , the authors used different vowels to analyze the disease. The MAMa tree is utilized for the pre-processing and the singular value and relief method is utilized for the feature selection method. the authors also applied the five-classification methods. The system achieved 92.4% of accuracy by using the KNN classifier. However, author didn't address the number of feature selected in the feature selection method and did not differentiate the experimental result optimization with and without feature selection process.

The authors [37] developed a Parkinson's diagnosis system to detect the disease. The authors used the vowels for the experiments. The relief method is utilized for the selection of Acoustic features. The authors

Table 2.1: Pros and cons of various reviewed models

DL model	Pro	Con
KNN	Simple and Intuitive. Adaptable to Changes. Effective with small datasets.	Computationally Expensive. High Memory Usage.
Naive Bayes	Efficiency with High-Dimensional Data. Works Well with Categorical Data.	Sensitive to Irrelevant Features. Limited Modeling Capacity.
Decision Tree	Handles non-Linearity and mixed data.	Overfitting. High variance
SVM	Robust to overfitting. Effective in High-Dimensional Space.	Difficulty with Large Datasets. Problem with multiclasses.
Random Forest	High Accuracy. Reduced overfitting.	Bias in Feature Selection.

applied the KNN and SVM methods for the classification. The SVM classifier achieved 91.25% of accuracy. Although ,the proposed method achieved good accuracy but still there is possibility to enhance the detection accuracy.

The authors [24] propose the three feature selection models such as Chi-Squared, Random forest and Genetic algorithm. The model is customized in such a way that it can select the 11 features from the information gain and genetic approach and 5 features from the genetic algorithm. The first approach and second approach select 11 features and the third approach selects 5 features out of 23. The table 2.1 elaborate various pros and cons of the models.

The major contribution of these articles are :

- The public dataset is collected from the UCI repository. The speech-based signal datasets comprise of 8 healthy patient's information as well as 23 Parkinson's disease patient's information.
- Three feature selection methods such as Chi-Squared, Random forest and Genatic algorithm to extract the features.
- The selected features are divided into training as well as testing datasets.
- The SMOTE is utilized to handle the imbalance datasets issues
- In the last phase Classifier's performance are analysed on the complete feature and reduced subsets of features.
- To deliver a method which provides, high accuracy, sensitivity and specificity compare than existing method.

3. Methodology. Firstly, the feature selection has been processed [24] by three methods such Chi-Squared, Random Forest and Genetic algorithm. The first approach and second approach select 11 features and the third approach selects 5 features out of 23. The selected feature is divided into training as well as testing datasets. Then these datasets are passed through different classification methods to find the best detection accuracy result.

Secondly, the speech dataset is imbalanced due to 130 out of 180 samples of the details of Parkinson's patients. The SMOTE is utilized to handle the imbalance datasets issues.

Thirdly in the last phase the performance of the classifier i.e naive Bayes, k-nearest neighbors, and random forest are analyzed on the complete feature and reduced subsets of features. Finally, it shows the Embedded Random forest algorithm gives the better accuracy i.e 97.44% of results compared to the existing techniques.

3.1. Dataset. This section describes the datasets.In this, the public dataset is collected from the UCI repository. The speech-based signal datasets comprise of 8 healthy patient's information as well as 23 Parkinson's disease patient's information. The Max Little Oxford University created the dataset. It divides into 195:31 (row: column) wherein the row defines the voice signals and the column defines the voice features. The PD patients are 147 out of 195 and leftover healthy patients' voices are available. The column also has two values 0 and 1. The value 0 defines the healthy and 1 defines the Parkinson's patients.

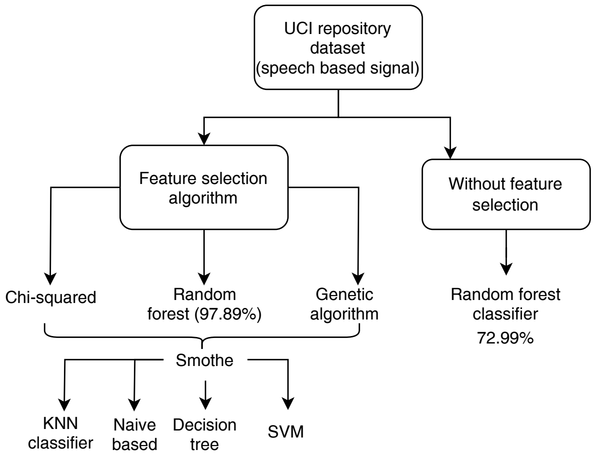

Fig. 3.1: Flowchart

3.2. Feature Selection Method. The feature selection method is the pre-processing step to recognize the vital features from the raw dataset. There are mainly two advantages to the feature selection method. The first advantage is diminishing the dimension of the datasets, and the second is optimizing the performance of the classifier model. In this proposed system, commonly three features are utilized i.e.5 as well as 11 features from 23 features.

3.3. Flowchart. The overall flow of the architecture is as shown in Figure 3.1. The authors using the UCI repository dataset applied classifier with and without feature selection method. The feature selection helps in achieving the high accuracy of 97.89%.

3.3.1. Features selection method using Chi-Squared method. It is the method to examine the freedom of the two incidents [25]. It is based on the statistics to identify the observed value as well as the expected value. It also helps to differentiate the expected value count (E) from the observed value count (O). In our dataset, there are two types of datas i.e. dependent and independent data. The Chi-square method utilizes for the retrieval of the features that are highly dependent on the response estimate as shown Equation 3.1.

$$x_c^2 = \sum sin(\frac{(O_i - E_i)^2}{E_i} \tag{3.1}$$

Here c is the degree of incidents, and 0_i, E_i is the observed as well as the expected value.

3.3.2. Features selection method using genetic algorithm . The Algorithm is also used for the feature selection method. It is the procedure to identify the predicted signal and noise. It is the process that represents the natural selection criteria. It helps to solve the complex problem that takes more time in the process. This approach is based upon the probabilistic nature to get the optimized result.

3.3.3. Features selection method using embedded random forest algorithm. It merges both methods such as the filter method and wrapper method [26]. The random forest method contains approximately 4 to 100 decision trees. It extracts the random observation from the database and extracts the different random features from the datasets.

3.4. Classification model.

3.4.1. KNN classifier. K-NN algorithm [27] is the supervised machine learning approach. It finds the relationship between new data and existing data and places the new data in the most relatable place. It is utilized for the classification and regression process. This approach is non-parametric based algorithm because it does not contain the assumption data. Calculate the Euclidean distance between two points by the Equation 3.2.

$$Distance = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{3.2}$$

Here (X_2, X_1) and (Y_2, Y_1) are the coordinates of two-point.

3.4.2. Naïve bayes algorithm. The naïve classification algorithm [31] is based upon the Bayes algorithm. When the different features are independent then this algorithm is the best approach for the classification. The naïve Bayes algorithm provides the best result when the dataset is unbiased and dependent on features based on functionality. This approach's accuracy is not completely dependent on the scale of dependent features but also depends on the classification of the features as well as common information between features.

3.4.3. Decision Tree. Decision tree is supervised algorithm. It is utilized for both regression and classification methods [32]. In this approach, nodes represent the features, and branches/edges represent the rule of features. This process uses two types of nodes i.e., the decision node, and the leaf node. It is also known as the attribute selection method. The unnecessary nodes are deleted to find the optimal result of the classification method.

3.4.4. Support Vector Method. SVM method [33]is used for the classification process. This approach uses the best line used to help with segregation in classes. So that new data can be put into incorrect categories. The hyperplane is the best decision edge in the decision tree. The vector machine determines the vectors which help to create the hyper line. These tremendous points are known as support vectors, so this algorithm is defined as a support vector machine. In SVM multiple lines are used to help the isolate n dimension. Although it require various lines to isolate the class in the different dimension areas.

4. Performance Parameter. The classification of the Parkinson's disease datasets is measured using evaluation metrics such as F1 score, TPR, FPR, accuracy, and Kappa score [34] in Equations 4.1-4.7. These accuracy helps in the assessment of the models. Accuracy is the measure of models performance for all the classes. The Kappa score, also known as Cohen's Kappa coefficient, is a statistic that measures the inter-rater agreement between two or more raters who are assessing the same categorical items. It is particularly useful when dealing with situations where there is a need to assess agreement beyond what might be expected by chance alone as shown in equation 4.1. The True Positive Rate, also known as sensitivity or recall, measures the proportion of actual positive cases that are correctly identified as positive by a classification model as mention in equation 4.4. Similarly, the False Positive Rate measures the proportion of actual negative cases that are incorrectly identified as positive by a classification model. Mathematically, it is calculated as mentioned in 4.5.

$$K = \frac{p_0 - p_e}{1 - p_e} \tag{4.1}$$

$$P_0 = \frac{\sum_{c=1}^{c} TP_C}{\sum_{c=1}^{c} (TP_C + FN_c)} \tag{4.2}$$

$$P_e = \frac{\sum_{c=1}^{c} TP_C * (Tp_c + FN_c)}{N^2} \tag{4.3}$$

$$TPR = \frac{TP}{TP + FN} \tag{4.4}$$

$$FPR = \frac{FP}{FP + TN} \tag{4.5}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4.6}$$

$$AVG = \frac{1}{k + F1 + AUC} \tag{4.7}$$

Table 5.1: KNN classifier

Feature Selection Algorithm	Accuracy(%)	F1-Score(%)	Sensitivity(%)	Specificity(%)	Precision(%)	AUC
All Features	72.11	70.11	72.11	72.11	72.11	0.66
Chi-Squared Method	79.48	85.18	88.46	61.53	82.14	0.75
Genetic Algorithm	87.17	92.06	90.62	71.42	93.54	0.81
Embedded Random Forest Method	89.74	92.59	96.15	76.92	89.28	0.86

Table 5.2: Naive Bayes' Classifier

Feature Selection Algorithm	Accuracy(%)	F1-Score(%)	Sensitivity(%)	Specificity(%)	Precision(%)	AUC
All Features	66.66	73.46	62.06	71.99	71	0.66
Chi-Squared Method	71.79	75.55	89.11	84.61	89.47	0.75
Genetic Algorithm	81.88	77.77	75.91	85.71	95.45	0.75
Embedded Random Forest Method	84.61	88.88	92.30	89.11	85.71	0.80

Table 5.3: SVM Classifier

Feature Selection Algorithm	Accuracy(%)	F1-Score(%)	Sensitivity(%)	Specificity(%)	Precision(%)	AUC
All Features	75	72	70	40	67.99	0.6
Chi-Squared Method	76.92	85.24	80.11	72	74.28	0.65
Genetic Algorithm	94.87	96.96	81.23	71.42	94.11	0.85
Embedded Random Forest Method	84.61	89.28	96.15	77	83.33	0.78

Table 5.4: Decision Tree Classifier

Feature Selection Algorithm	Accuracy(%)	F1-Score(%)	Sensitivity(%)	Specificity(%)	Precision(%)	AUC
All Features	70	72	77	70	71	0.67
Chi-Squared Method	82.05	87.27	92.30	61.53	82.75	0.76
Genetic Algorithm	87.17	92.06	90.62	71.42	93.54	0.81
Embedded Random Forest Method	87.17	90.19	88.46	84.61	92	0.86

5. Results and Discussions. This part of the section provides a compressive study of the various classifier along with the feature extraction method [36]. in this article, we applied three types of feature selection methods such as genetic, chi-square method, and decision tree. Additionally, detection of the disease is processed by the naïve Bayes algorithm, SVM, Random Forest, and decision tree. For the validation of the result, 10-k cross-validation is applied. The system performance is measured by the five parameters such as F1.

The important features are measured by the chi-square [37], genetic, and embedded random forest algorithm. The top eleven features are measured by the chi square method. In the Genetic algorithm, each feature information gain is calculated. The high values of information gain are utilized for the classification method. The top eleven features are selected by the genetic algorithm. Additionally, the last five features are calculated through a random forest algorithm [38]-[39].

The performance parameters i.e. accuracy, sensitivity and specificity are measured by classifier. In this method first features are extracted through feature selection method and then compare the optimization performance of the classifier include all the features and exclude selected features.

The KNN classifier achieves 72.11% of accuracy without feature selection method. further the features selection method named as random forest method is used to enhance the classifier performance. Then classifer optimised the 89.74% of accuracy with feature selection approach as shown in Table 5.1.

The Naïve Bayes classifier achieves 66.66667% of accuracy without feature selection method. further the features selection method named as random forest method is used to enhance the classifier performance. Then classifier optimised the 89.61% of accuracy with feature selection approach as shown in Table 5.2.

The SVM classifier achieves 75% of accuracy without feature selection method. further the features selection method named as random forest method is used to enhance the classifier performance. Then classifier achieved

Table 5.5: Random Forest Classifier

Feature Selection Algorithm	**Accuracy(%)**	**F1-Score(%)**	**Sensitivity(%)**	**Specificity(%)**	**Precision(%)**	**AUC**
All Features	72.99	74.88	34	70	67.99	0.70
Chi-Squared Method	84.61	88.88	92.30	69.23	85.71	0.80
Genetic Algorithm	93.55	95.23	93.75	85.71	96.77	0.89
Embedded Random Forest Method	97.89	96.69	92.30	92.30	96	0.92

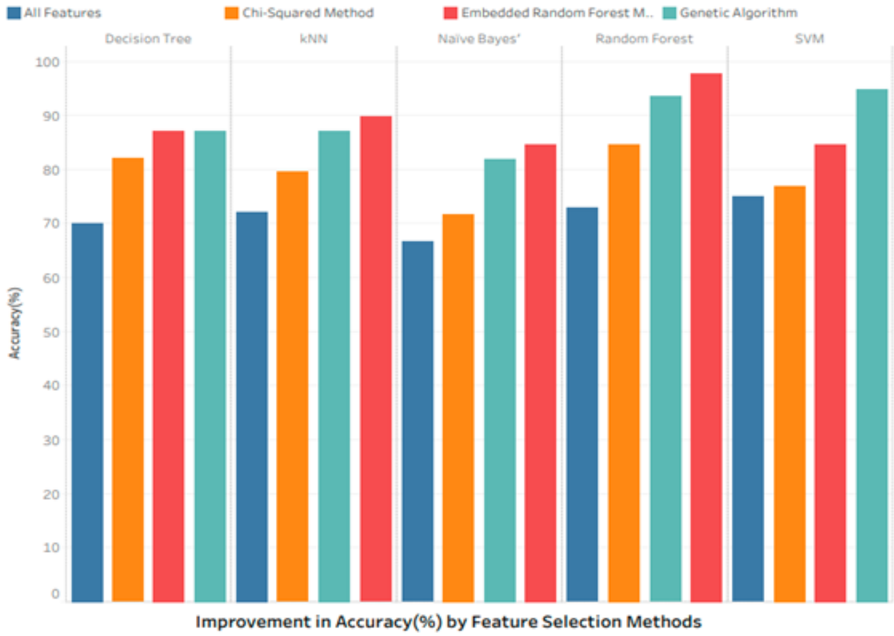

Fig. 5.1: Accuracy improvement by feature selection methods

the 84.61% of accuracy with feature selection approach as shown in Table 5.3.

The Decision classifier achieves 70% of accuracy without feature selection method. further the features selection method named as random forest method is used to enhance the classifier performance. Then classifier achieved the 87.17% of accuracy with feature selection approach as shown in Table 5.4.

The Random forest classifier achieves 72% of accuracy without feature selection method. further the features selection method named as random forest method is used to enhance the classifier performance. Then classifier achieved the 97.89% of accuracy with feature selection approach as shown in Table 5.5.

The performance improvement of the proposed methodology with the existing method is defined in Fig. 5.1. The detection accuracy of the classifier is optimized through the feature selection method. The Embedded Random Forest method provides the best result compared to the existing feature selection method. The proposed detection model of Parkinson's disease gives a better result compared to the existing model. The proposed model accuracy of 97.89% in comparison to the existing models is defined in Fig. 5.2.

6. Conclusion. Parkinson's Disease is mainly voice disorder due to neurodegenerative. Therefore, there is only one way to improve the patent's health i.e., through early detection of the disease. The proper diet plan and medication can improve the symptoms of patients with Parkinson's disease. This experimental result shows the detection of the disease at early phase. The author has utilised different combinations of feature selection as well as the classifier methods. Finally, the proposed model provides 97.89% of accuracy which is better than the existing recent work of literature.

In the future, author will plan to work on other datasets namely voice and work on slowness in handwriting

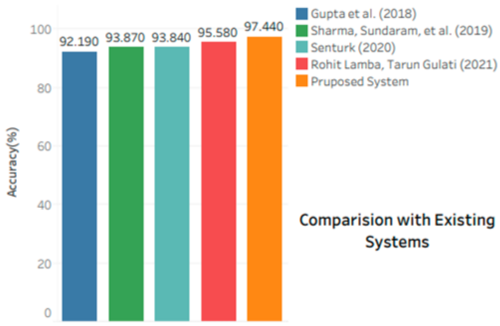

Fig. 5.2: Comparison with existing systems

skill symptoms. Developing personalized diagnostic models that consider individual variations and response to treatments is a growing area of interest. Random classification and feature selection techniques can aid in tailoring diagnoses to each patient's unique characteristics. Explaining the rationale behind a diagnosis is crucial for building trust in medical AI systems. Future research could focus on incorporating interpretability techniques alongside random classification and feature selection methods.

REFERENCES

[1] J. S. Almeida, P. P. Rebouças Filho, T. Carneiro, W. Wei, R. Damaševičius, R. Maskeliūnas, and V. H. C. de Albuquerque, *Detecting parkinson's disease with sustained phonation and speech signals using machine learning techniques*, Pattern Recognition Letters, 125 (2019), pp. 55–62.

[2] S. Bhattacharya, P. K. Reddy Maddikunta, Q.-V. Pham, T. R. Gadekallu, S. R. Krishnan S, C. L. Chowdhary, M. Alazab, and M. Jalil Piran, *Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey*, Sustainable Cities and Society, 65 (2021), p. 102589.

[3] G. Chandrashekar and F. Sahin, *A survey on feature selection methods*, Computers & Electrical Engineering, 40 (2014), pp. 16–28.

[4] C. L. Chowdhary and D. Acharjya, *Segmentation and feature extraction in medical imaging: A systematic review*, Procedia Computer Science, 167 (2020), pp. 26–36. International Conference on Computational Intelligence and Data Science.

[5] F. N. Emamzadeh and A. Surguchov, *Parkinson's disease: biomarkers, treatment, and risk factors*, Frontiers in neuroscience, 12 (2018), p. 612.

[6] S. Grover, S. Bhartia, A. Yadav, K. Seeja, et al., *Predicting severity of parkinson's disease using deep learning*, Procedia computer science, 132 (2018), pp. 1788–1794.

[7] H. Gunduz, *Deep learning-based parkinson's disease classification using vocal feature sets*, IEEE Access, 7 (2019), pp. 115540–115551.

[8] S. Lahmiri and A. Shmuel, *Detection of parkinson's disease based on voice patterns ranking and optimized support vector machine*, Biomedical Signal Processing and Control, 49 (2019), pp. 427–433.

[9] R. Lamba, T. Gulati, K. A. Al-Dhlan, and A. Jain, *A systematic approach to diagnose parkinson's disease through kinematic features extracted from handwritten drawings*, Journal of Reliable Intelligent Environments, (2021), pp. 1–10.

[10] R. Lamba, T. Gulati, and A. Jain, *Comparative analysis of parkinson's disease diagnosis system*, Adv Math Sci J, 9 (2020), pp. 3399–3406.

[11] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, *Suitability of dysphonia measurements for telemonitoring of parkinson's disease*, Nature Precedings, (2008), pp. 1–1.

[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 classification methods for improving the diagnosis of parkinson's disease*, Cognitive Systems Research, 54 (2019), pp. 90–99.

[13] Y. S. Murthy and S. G. Koolagudi, *Classification of vocal and non-vocal segments in audio clips using genetic algorithm based feature selection (gafs)*, Expert Systems with Applications, 106 (2018), pp. 77–91.

[14] L. Naranjo, C. J. Perez, Y. Campos-Roca, and J. Martin, *Addressing voice recording replications for parkinson's disease detection*, Expert Systems with Applications, 46 (2016), pp. 286–292.

[15] I. Nissar, D. R. Rizvi, S. Masood, and A. N. Mir, *Voice-based detection of parkinson's disease through ensemble machine learning approach: A performance study*, EAI Endorsed Transactions on Pervasive Health and Technology, 5 (2019), pp. e2–e2.

[16] R. Olivares, R. Munoz, R. Soto, B. Crawford, D. Cárdenas, A. Ponce, and C. Taramasco, *An optimized brain-based algorithm for classifying parkinson's disease*, Applied Sciences, 10 (2020), p. 1827.

[17] K. Polat, *A hybrid approach to parkinson disease classification using speech signal: the combination of smote and random forests*, in 2019 scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT), Ieee, 2019, pp. 1–3.

[18] P. Rani, R. Kumar, N. Ahmed, and A. Jain, *A decision support system for heart disease prediction based upon machine learning. j reliab intell environ*, 2021.

[19] P. Rani, R. Kumar, and A. Jain, *Multistage model for accurate prediction of missing values using imputation methods in heart disease dataset*, in Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2020, Springer, 2021, pp. 637–653.

[20] P. Rani, R. Kumar, A. Jain, and R. Lamba, *Taxonomy of machine learning algorithms and its applications*, Journal of Computational and Theoretical Nanoscience, 17 (2020), pp. 2508–2513.

[21] G. T. Reddy, S. Bhattacharya, S. Siva Ramakrishnan, C. L. Chowdhary, S. Hakak, R. Kaluri, and M. Praveen Kumar Reddy, *An ensemble based machine learning model for diabetic retinopathy classification*, in 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1–6.

[22] S. G. Reich and J. M. Savitt, *Parkinson disease*, Medical Clinics of North America, (2018).

[23] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, *A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform*, Applied Soft Computing, 74 (2019), pp. 255–263.

[24] I. H. Sarker, A. Kayes, and P. Watters, *Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage*, Journal of Big Data, 6 (2019), pp. 1–28.

[25] Z. K. Senturk, *Early diagnosis of parkinson's disease using machine learning algorithms*, Medical hypotheses, 138 (2020), p. 109603.

[26] P. Sharma, R. Jain, M. Sharma, and D. Gupta, *Parkinson's diagnosis using ant-lion optimisation algorithm*, International Journal of Innovative Computing and Applications, 10 (2019), pp. 138–146.

[27] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, *Diagnosis of parkinson's disease using modified grey wolf optimization*, Cognitive Systems Research, 54 (2019), pp. 100–115.

[28] V. S. Sharma, A. Afthanorhan, N. C. Barwar, S. Singh, and H. Malik, *A dynamic repository approach for small file management with fast access time on hadoop cluster: Hash based extended hadoop archive*, IEEE Access, 10 (2022), pp. 36856–36867.

[29] V. S. Sharma and N. Barwar, *An efficient approach to enhance the scalability of the hdfs: Extended hadoop archive (ehar)*, in 2021 Emerging Trends in Industry 4.0 (ETI 4.0), 2021, pp. 1–6.

[30] V. S. Sharma and N. C. Barwar, *Performance evaluation of merging techniques for handling small size files in hdfs*, in Data Analytics and Management, A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, eds., Singapore, 2021, Springer Singapore, pp. 137–150.

[31] G. Solana-Lavalle, J.-C. Galán-Hernández, and R. Rosas-Romero, *Automatic parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features*, Biocybernetics and Biomedical Engineering, 40 (2020), pp. 505–516.

[32] T. Tuncer and S. Dogan, *A novel octopus based parkinson's disease and gender recognition method using vowels*, Applied Acoustics, 155 (2019), pp. 75–83.

[33] T. Tuncer, S. Dogan, and U. R. Acharya, *Automated detection of parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels*, Biocybernetics and Biomedical Engineering, 40 (2020), pp. 211–220.

[34] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, *Comparing different supervised machine learning algorithms for disease prediction*, BMC medical informatics and decision making, 19 (2019), pp. 1–16.

[35] S. P. Vijay Kumar Gurani, Chiranji Lal Chowdhary, *Exploring breast cancer classification of histopathology images from computer vision and image processing algorithms to deep learning*, International Journal of Advanced Science and Technology, 29 (2020), pp. 43 – 48.

[36] Y. Xiong and Y. Lu, *Deep feature extraction from the vocal vectors using sparse autoencoders for parkinson's classification*, IEEE Access, 8 (2020), pp. 27821–27830.

[37] O. Yaman, F. Ertam, and T. Tuncer, *Automated parkinson's disease recognition based on statistical pooling method using acoustic features*, Medical hypotheses, 135 (2020), p. 109483.

[38] L. Zahid, M. Maqsood, M. Y. Durrani, M. Bakhtyar, J. Baber, H. Jamal, I. Mehmood, and O.-Y. Song, *A spectrogram-based deep feature assisted computer-aided diagnostic system for parkinson's disease*, IEEE Access, 8 (2020), pp. 35482–35495.

[39] T. A. Zesiewicz, Y. Bezchlibnyk, N. Dohse, and S. D. Ghanekar, *Management of early parkinson disease*, Clinics in geriatric medicine, 36 (2020), pp. 35–41.

Edited by: Chiranji Lal Chowdhary
Special issue on: Scalable Machine Learning for Health Care: Innovations and Applications
Received: May 25, 2023
Accepted: Aug 24, 2023