Journal of Theoretical and Applied Information Technology
30th November 2022. Vol.100. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PARKINSON'S DISEASE CLASSIFICATION USING ADAPTIVE RANDOM FOREST AND IMPROVED FEATURE SELECTION METHODS

REPUDI PITCHIAH[1] , DR T. SASI ROOBA[2] , DR K. UMA PAVAN KUMAR [3]

[1]Dept. of Computer Science Engineering, Annamalai University, Chidambaram, Tamil Nadu.
[2]Dept. of Computer Science; Engineering, Annamalai University, Chidambaram, Tamil Nadu.
[3]Dept of Computer Science and Engineering, MRIT, Hyderabad, Telangana.
Email: [1]pitchaiah99@gmail.com, [2]sasiruba@gmail.com, [3]dr.kethavarapu@gmail.com

ABSTRACT

One of the most prevalent neurological illnesses, Parkinson's disease (PD) mainly affects the motor system of the brain's core system. In actuality, PD is marked by speech impairment, tremors, muscular rigidity, and gait inaccuracies. Even though the primary symptoms of Parkinson's disease cannot be clearly distinguished from those of other disorders, a definitive diagnosis of the condition is often dependent on several neurological, psychiatric, and physical studies. As a result, several machine learning-based automatic diagnostic assistance systems have lately been used to aid in the assessment of PD patients. One of the most difficult medical issues at hand today is the automatic diagnosis of early Parkinson's disease using feature data sets. Such datasets contain several characteristics that are either worthless or plagued by issues like noise that hinder learning and add to required computing load. This article suggests a hybrid feature selection algorithm built on an enhanced correlation method with Bootstrap to increase the effectiveness of feature selection to ensure most accurate performance of the classier. By combining the finer elements of filters and wrappers, such an algorithm finds the ideal subset of features by removing a majority of noisy or unrelated information. Select optimal features from overall data features is also specific issue behind implementation of feature extraction In order to overcome such problems, we have proposed an adaptive random forest classifier method that uses ensemble feature selection technique for better information gain (IG), improved correlation (IC) and gain ratio (GR). Also,s it seeks to solve the class imbalance problem by applying bootstrap re-sampling for medical data. According to the evaluations in terms of accuracy, precision, recall, and F-Score, the developed model has been found to be more efficient than conventional methods. The analysis of experimental results indicates better accuracy of the proposed framework (88.3% accuracy) as compared to other techniques.

Keywords: *Parkinson's disease (PD), Multiclass Classification, Feature Selection, Random Forest, and High Dimensional Data***.**

1. INTRODUCTION

Parkinson's disease, which is known to affect a large number of people worldwide, primarily affects the central nervous system. The majority of PD sufferers are noted to be demanding both physically and mentally. They experience painful spasms, problems concentrating, depression, and other negative emotions. The clinical characteristics of PD include a wide spectrum, from motor to nonmotor symptoms. Resting tremor, stiffness, and hypo phonic speaking are a few of the motor signs. Hallucinations, melancholy, constipation, sleeping problems, cognitive decline, and issues with impulse control are examples of non-motor symptoms. More illness is evident through non-motor symptoms instead of motor [1,3]. In a majority of instances, clinicians have trouble

predicting if a particular patient is manifesting symptoms or will eventually acquire it [7]. To overcome such eventualities, a computational model that assesses and analyses the data of a specific patient and accurately forecasts the likelihood of PD manifestations. The majority of PD patients have vocal impairment symptoms, sometimes referred to as dysphonia. In order to examine patients at different phases, there are a number of dysphonia-related assessments, including voice-related issues [14]. Speech processing is a highly active area of study. A crucial requirement for an effective speech processing system, the choice of feature extraction approach plays a major impact in obtaining improved outcomes [20].Some proven feature extraction algorithms like Linear Predictive Coefficient (LPC), Mel Frequency Cepstral Coefficient (MFCC), Perceptual Linear Prediction (PLP), Relative Spectral Perceptual Linear Prediction (RASTAPLP), and Wavelet Transform (WT) were reported by various authors in earlier works [21]. The authors suggested a medical diagnostic support system (MDSS), based on ANN, AdaBoost, and DT machine learning algorithms, with the aim of predicting atherosclerosis. The strengths and shortcomings of these approaches are also illustrated in [22]. In order to remove the majority of the random or noisy characteristics and identify the ideal subset of features, the method blends the advantages of filters and wrappers. In order to overcome such problems, we have proposed an adaptive random forest classifier method that uses ensemble feature selection technique for better information gain (IG), improved correlation (IC) and gain ratio (GR). Also, it seeks to solve the class imbalance problem by applying bootstrap resampling for medical data. The result of the proposed method proved that adaptive RF (Random Forest) classifier offers better accuracy, precision, and F-score values than standard Random Forest and KNN classification algorithms. The overall performance of algorithms was tested over five real datasets. From the results, the proposed classifier method shows better performance acrossall real datasets as compared to standard methods.

2. REVIEW OF RELATED WORK

The research studies on detection of Parkinson's disease have frequently been reported by various researchers at different points of time. Here, the research works specifically done to identify Parkinson disease using subject voice samples is briefly reviewed here. In their research work, Max A. Little et al. [15] have proposed a unique method for dividing participants into Parkinson sick and control subjects by identifying dysphonia. Pitch period entropy (PPE), a novel, reliable measure of dysphonia, was developed in their research. The data, which included 195 sustained vowel phonations, were gathered from 31 individuals (23 PD patients and 8 healthy persons). Three steps made up their methodology: feature computation, preprocessing, feature selection, and classification. They employed a linear kernel support vector machine for classification (SVM). The suggested model attained a 91.4% accuracy rate.

Richa Mathur et al [16] developed a PD prediction method. On the provided dataset, they performed data preparation, classification, and outcome analysis using the weka tool to build the algorithms. They combined k-NN with Adaboost. Bagging, MLP, and M1. The finding was that KNN + Adaboost. The highest classification accuracy at 91.28% was obtained by M1.

A.Yasar et al [17] developed a mechanism to detect symptoms of Parkinson's disease through artificial neural networks. The UCI machine learning repository was used to obtain the dataset, wherein 45 attributes were taken as input values and one output for the classification while applying MATLAB programme. The suggested model seemed to have a 94.93% accuracy in separating the healthy patients from the PD patients.

Achraf Benba et al [18] carried out study to segregate PD patients and the control group. In their research, the data consisted of 34 sustained vowels that were recorded from 34 patients, 17

of whom had Parkinson's disease. Mel-frequency cepstral coefficients (MFCC) ranging from 1 to 20 were collected from each individual. For classification, SVM with several kernel types was employed. The cross-validation method employed was LOSO. On the basis of the top 12 MFCC coefficients, the linear kernel SVM showed the greatest accuracy of 91.17%

3. BASIC PRILIMINARIES

Designing a classification system for large-scale multiclass medical data is a difficult process which is discussed earlier. The most desirable step to complete the classification process is the data preprocessing which includes the selection of features and the ranking of features. The feature selection method used to find the optimal features removes the irrelevant features [5]. Three different of feature selection techniques are applied in this work which includes Information Gain, Gain Ratio and Correlation. From the feature selection methods, the desired features are retrieved based on the ranking scores. In case of information theory, the Entropy is commonly used to measure unpredictability of the system. The entropy of $\mathbf{y}$ is calculated using (1)

$$\mathbf{H(y) = -\sum_{y\in Y} p(y)\log_2(p(y))} \quad (1)$$

From the (1), $\mathbf{P(y)}$ denotes representation of the marginal probability density function of the variable $\mathbf{y}$ .Next, $\mathbf{y}$ values in the training dataset $\mathbf{S}$ are split based on the features of $\mathbf{x}$. When there is a comparison between entropy of $\mathbf{y}$ to partitions caused by $\mathbf{x}$ is lower than the entropy of $\mathbf{y}$ prior to partitioning, then there occurs a convergence between the features $\mathbf{y}$ and $\mathbf{x}$ .The entropy of $\mathbf{y}$ following observation of $\mathbf{x}$ is then:

$$\mathbf{H(y\mid x) = -\sum_{x\in X} p(x)\sum_{y\in Y} p(y\mid x)\log_2(p(y\mid x))} \quad (2)$$

From (2), $\mathbf{p(y\mid x)}$ is the conditional probability of $\mathbf{y}$ given $\mathbf{x}$.

3.1. Information gain (IG)

The impurity mechanism assigned to training set $\mathbf{S}$ is referred to as entropy, defined as quantifying the extra information found in $\mathbf{y}$ as given by $\mathbf{x}$ that is measured after the entropy $\mathbf{y}$ declines. The complete idea is known as Information Gain (IG) and is represented using (3)

$$\mathbf{IG=H(y) - H(y\mid x) = H(x) - H((x\mid y)} \quad (3)$$

From the above (3), it is observed that information retained on $\mathbf{y}$ while observing $\mathbf{x}$ is almost similar to the information retained on $\mathbf{x}$ after noting $\mathbf{y}$ .In case features have multiple values and are not more informative and biased as produced by the $\mathbf{IG}$ criterion. It becomes the weakness criteria for $\mathbf{IG}$ .Similarly, for each feature, if discretization cardinality of a function $\mathbf{f_m = F=\{f_1,..,f_M\}}$, the information gain can be evaluated by using (4),

$$\mathbf{IG_B\left(f_m\right) = \frac{I\left(C\mid f_m\right)}{\log_2 \Gamma}} \quad (4)$$

From (4), it may be noted that if C is the class variable and Γ is the discretization cardinality of $\mathbf{f_m}$, then, $\mathbf{I(C\mid f_m)}$ is represented as an information gain corresponding to the $\mathbf{f_m}$, and $\mathbf{\Gamma}$ is penalty for the information gain.

3.2. Gain Ratio (GR)

An enhanced feature selection method called gain ratio is used to solve the problem of bias generated from the information gain. Also, it complements the $\mathbf{IG}$ method. The measure of information gain related to the entropy of $\mathbf{F_{ei}}$ in case of the $\mathbf{GR}$ can be represented in (5) as described below:

$$\mathbf{GR(C,F_{ei}) = \left[\frac{H\left(C\right) - H\left(C\mid F_{ei}\right)}{H\left(F_{ei}\right)}\right]} \quad (5)$$

From (5), $\mathbf{H(C)}$ denotes the class entropy of $\mathbf{C}$, similarly $\mathbf{H(C\mid F_{ei})}$ denotes the class

entropy C related to the features F_{ei}, and finally, $H(F_{ei})$ is considered as F_{ei} entropy of measure

4. PROPOSED APPROACH

Let a P be the set of m patients $P=\{p_1,p_2,p_3,...,p_m\}$ and their n voice samples $V_i=\{v_1,v_2,v_3,...,v_n\}$.Each sample V_i, can be represented by a set of r features $F_i=\{f_1,f_2,f_3,...,f_r\}$. The task of diagnosing the health status of patient as healthy P_i or suffering from PD when the set of voice samples of patient P_i and the features of such samples are analyzed. Let us introduce the class $C=\{0,1\}$ to which patient P_i belongs. Class 0 indicates that the patient is healthy, while class 1 indicates PD. Also, let $c_{ij} \in C$ denote the class of a single sample j of patient i. Let us denote the decision that the sample V_{ij} to belongs to class $c_{ij} \in C$ as:

$$d_{ij} = \square(V_i) \quad (6)$$

where stands for classifier model, and V_{ij} the j^{th} sample, from the i^{th} patient. n samples of patient P_i are analyzed, and each sample is classified separately. The final decision d_i^{final} $i \in C$ C whether patient P_i has PD or is healthy is on the basis of majority voting:

$$d_i^{final} = \begin{cases} 0, \text{ if } d_{ij}^0 \geq d_{ij}^1 \\ 1, \text{ if } d_{ij}^1 > d_{ij}^0 \end{cases} \quad (7)$$

where $d_{ij}^1 = \sum_{j=0}^{n} d_{ij}$ is the decision that the sample is from a patient with Parkinson's disease, $d_{ij}^0 = n - d_{ij}^1$ represents decision that the sample is from a healthy person.

Other aspects of activities, including speaking, are also impacted by Parkinson's disease. A weak, monotonous, or nasal voice, delayed speaking, difficulties initiating a sentence, rhythm or accent irregularity, and stuttering are common issues. As a result, Parkinson's disease may be identified utilizing human speech analysis. The non-contact and non-invasive nature of this approach is a benefit. Without the need for specialized, frequently costly healthcare equipment, voice samples may be adequately prepared, pre-recorded, and analyzed at any time. The collection of samples does not need to be done in professional medical departments; it is possible to collect data at home and send them for testing, for example, by email. Online registration and diagnosis are also possible over the Internet. It is possible to extract the right voice features from the recorded sample and classify them. As a consequence, it is feasible to ascertain if the sample being studied is from a normal or ill person.

4.1. Feature Selection

4.1.1. Improved Correlation Method

The simplest method used to evaluate feature optimal subsets is when used with the criteria of the correlation-based mapping [6]. The selected optimal subsets are retrieved from the correlation method having features with a strong correlation, it is also identified that features are not correlatedwith each other. The selection of optimal features with improved correlation is shown in Algorihm1.From the algorithm, all the unimportant featuresare simply discarded that have lesser importance as compared to the threshold ε.

Algorithm :Improved Correlation Method

1. Input: 1. Training Set $T=\{(u_1,v_1),..,(u_N,v_N)\}$
   2. Number of Features $FS=\{f_1,..,f_M\}$
   3. D^* Size of the Feature Set FS^*
   4. ε - Threshold.

2. First initialize the feature set empty $FS^* \Leftarrow \{\ \}$

3. Set $FS_{ù} \Leftarrow FS$

4. forevery feature in the feature set $f_m \in FS_{ù}$ do

$$IG_B(f_m)=\frac{I(C|f_m)}{\log_2 \Gamma};$$

Set the Importance Value $Imp(f_m) \Leftarrow IG_B(f_m)$;

$FS_{ù} \Leftarrow FS_{ù} - \{f_m\}$ if $Imp(f_m) < \varepsilon$.

end for

5. for $i=1$ to D^*

do

Find Feature $f_i \in FS_{ù}$, whose importance value $Imp(f_i)$ is the maximum in $FS_{ù}$;

Set $FS^* \Leftarrow FS^* \cup \{f_i\}$ and $FS_{ù} \Leftarrow FS_{ù} - \{f_i\}$;

$\forall (f_m \in FS_{ù})$ update $Imp(f_m)$ by using

$Imp(f_m)=IG_B(f_m)\times(1-Corr(f_m,FS_{ù}))$

where $Corr(f_m,FS_{ù}) = Corr(f_m,FS_{ù})=Max(Corr_n(f_m,f_k)), \forall (f_k \in FS_{ù}$, the normalized correlation $Corr_n(f_m,f_n)$ can be defined as for any two features $(f_m,f_n) \in FS$

$$Corr_n(f_m,f_n)=\frac{Corr(f_m,f_n)}{Max(Corr(f_k,f_l))} \forall f_k \neq f_l \in FS_{ù}$$

If $Imp(f_m) < \varepsilon$ set $FS_{ù} \Leftarrow FS_{ù} - \{f_m\}$

end for

6. end

7. Output: The Selected Feature Subset FS^*

4.1.2. Random Sampling - Bootstrap

Random sampling extracts sub-samples from the original samples by using the bootstrap resampling method. The $k+1$ datasets are randomly extracted from the original dataset; and then a number of random records from the k sets are considered as training sets, while rest of the records in the set is called validation set. The complete idea of the bootstrap method is illustrated in Figure 1.

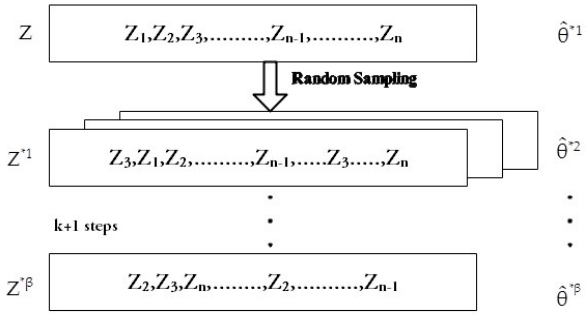

Fig. 1. Resampling with the Bootstrap Method

Algorithm: Random Sampling – Bootstrap

1. Select Z independent bootstrap samples $z_1^*,.....,z_\beta^*$, total n data values drawing from Z.

2. Then, for each bootstrap sample, we generated equal replication of bootstrap with the below equation

$$\hat{\theta}^*(b)=s(z_b^*);b=1,.....,Z$$

3. Finally , the standard error $SE_f(\hat{\theta})$ of the total Z replicatesis calculated as

$$SE_Z=\left[\frac{1}{Z-1}\sum_{b=1}^{Z}\left\{\hat{\theta}^*(b)-\hat{\theta}^*(.)\right\}^2\right] \text{ with}$$

$$\hat{\theta}^*(g)=\frac{1}{Z}\sum_{b=1}^{Z}\hat{\theta}^*(b)$$

In the proposed work, the results are compared with different classifiers using different feature selection techniques. The complete scheme of the proposed classification system for large-scale PD data is shown in Figure 2. We compared the following classification types: Random Forest and SVM (Support Vector Machine).

4.2. Classification

4.2.1. Adaptive Random Forest

A popular machine learning based algorithm that is used for classification task is called the Random Forest [14, 15]. The classifier is represented as a tree structure in case of the

random forest. For every tree existing in the forest, a unit vote is assigned and also every entry is represented as the class label. The advantage of this classification method is that it is a fast and robust method, which can easily identify nonlinear patterns from the given data. The method can work on both numeric and categorical data efficiently, which can also adjust with over adjustment once the forest adds a greater number of trees.3.4. Adaptive Random Forest. The proposed RFC approach identifies optimal features from multiple feature selection techniques for information gain, correlation, and gain ratio. The collected optimal features are applied to the training data of the classification algorithm. Later, Bootstrap process is applied in order to solve class imbalance problem. The functionality of bootstrap-resample helps to enhance the functionality of the RFC algorithm. However, if the generated data is uniform in nature, then it is not advisable to apply bootstrap. Next, random forest is applied to data set which is generated from the bootstrap result. Finally, the performance of the classification algorithm is measured with the given performance metrics that include classification accuracy, F metric, ROC, sensitivity, and specificity, respectively. The proposed RFC approach is represented with the Algorihm2, which is shown below.

Algorithm: Adaptive Random Forest

Input: $Z_T = \{z_1, z_2, z_3, ..., z_n\}$ // Training dataset

Output: Accuracy of the Classification algorithm Ψ.

1. Retrieved optimal subset using multiple feature selection methods, which is then applied to training data Z_T and obtains optimal feature subset δ_m.

2. Next, bootstrap approach is applied to δ_m of Z_T and the final training data corresponds to bootstrap result is Z_T.

3.Finally, RFC is applied to the Z_T and accuracy Ψ is calculated.

4. Return Accuracy Ψ.

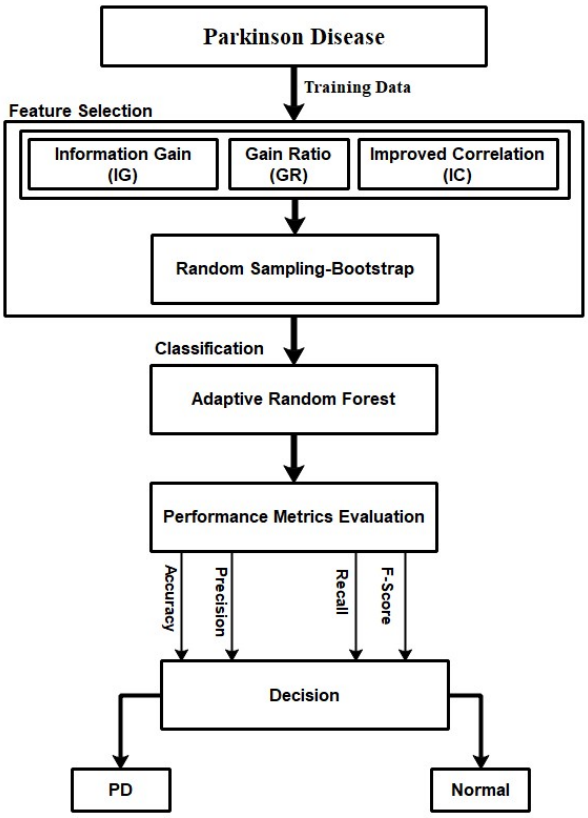

Fig. 2. Schematic process for the proposed classification of Parkinson Disease.

5. EXPERIMENTAL RESULTS

As shown in Table 1, some classifier algorithms like True positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were used in the study to assess their efficiencies. True Positive (TP)is a basic measure resulting in positive predictions, while false positive (FP) is representation of incorrect positive predictions and is opposite to TP. True negative (TN) is a measure of the correct negative predictions of the given instances and false negative (FN) is the inaccurate negative predictions of the instances. The performance measures and its formulas are shown in (6) and (7).

Table 1. The Confusion Matrix

Actual Values			
Predicted Values		Positive	Negative
	Positive	TP	FP
	Negative	FN	TN

$$Accuracy=\frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall=\frac{TP}{TP+FN} \quad (10)$$

$$F-Score = \frac{(2*TP)}{(2*TP+FP+FN)} \quad (11)$$

For binary classification issues, the AUC-ROC Curve is a probability curve and a performance indicator that is often utilized. This curve illustrates how well the classification model can discriminate between different classes. The genuine positive rate is represented by the y-axis in the ROC curve, while the false positive rate is shown by the x-axis. AUC has a value between 0 and 1. When the model's AUC is near to 1, it performs very well in terms of classification, whereas when it is close to 0, it performs poorly in terms of separability, and when it is close to 0.5, it is unable to separate data.

A) Results and Discussions

The simulation analysis used in this work was carried out using the Python platform. SVM, RF, and ARF were all used to classify the data. 75% of the data (147 observations) were used to train the model, while 25% (48 observations) were used to test each classifier. The ten-fold cross validation approach had been employed in detecting signs and symptoms of PD to evaluate the system's efficiency.

6.1. Dataset description

The UCI Repository has been cited as the source for the Parkinson data collection [5]. There are 756 samples in the database. The Class label attribute is one of its 754 characteristics. The 188 PD patients (107 men and 81 women) with ages ranging from 33 to 87 who were treated at the Department of Neurology at the Cerrahpaşa Faculty of Medicine, Istanbul University, provided the data for this study. The control group comprises of 64 healthy people aged between 41 and 82 (23 male and 41 female).The primary focus of this paper is to analyze and compare the impact of different feature ranking techniques over classification algorithm for this dataset. In the proposed section, we have described the 3 feature ranking techniques that have been used in this feature ranking experiment.

Table 2. Feature Ranking results for each ranking technique

Rank	Gain Ratio	Score	Information Gain	Score	Improved Correlation	Score
1	MDVP.FloHz	0.3941	PPE	18.1976	Spread1	0.1636
2	Spread1	0.2190	MDVP.FloHz	17.3215	PPE	0.1564
3	MDVP.APQ	0.2157	Spread1	16.2466	Spread2	0.1362
4	PPE	0.2108	Spread2	12.2796	DFA	0.1056
5	NHR	0.1977	MDVP.FhiHz	11.3707	RPDE	0.0991
6	Spread2	0.1952	MDVP.FloHz	10.8613	MDVP.FoHz	0.0964
7	MDVP.FhiHz	0.1915	MDVP.APQ	9.4973	MDVP.FloHz	0.0924

8	MDVP .RAP	0. 18 81	RPDE	8. 93 84	HNR	0. 08 89
9	Jitter.D DP	0. 18 82	MDVP .Shim mer	8. 69 74	Shimm er.APQ 3	0. 08 09
1 0	MDVP .Shim mer	0. 18 79	MDVP .JitterA bs	8. 59 95	Shimm er. DDA	0. 08 07
1 1	Shimm er.APQ 5	0. 18 29	Shimm er.APQ 5	8. 30 34	MDVP .Shim mer	0. 07 81
1 2	MDVP .Shim merdB	0. 17 54	Shimm er.APQ 3	8. 28 13	Shimm er.APQ 5	0. 07 44
1 3	MDVP .FoHz	0. 16 76	HNR	8. 26 52	MDVP .PPQ	0. 07 03
1 4	Shimm er.APQ 3	0. 16 19	MDVP .RAP	8. 08 34	MDVP .JitterA bs	0. 06 71
1 5	Shimm er.DD A	0. 16 07	Shimm er.DD A	8. 04 19	MDVP .RAP	0. 06 19
1 6	MDVP .JitterA bs	0. 15 95	Jitter.D DP	8. 00 06	Jitter.D DP	0. 06 19
1 7	MDVP .PPQ	0. 15 66	DFA	7. 79 38	MDVP .Shim merdB	0. 06 16
1 8	MDVP .Jitter	0. 14 85	MDVP .Shim merdB	7. 72 32	MDVP .Jitter	0. 06 09
1 9	HNR	0. 10 99	D2	7. 29 63	MDVP .APQ	0. 05 43
2 0	RPDE	0. 08 45	MDVP .PPQ	6. 96 08	MDVP .FhiHz	0. 04 48
2 1	D2	0. 07 84	MDVP .Jitter	6. 80 61	D2	0. 04 05
2 2	DFA	0. 07 24	NHR	6. 52 59	NHR	0. 02 65

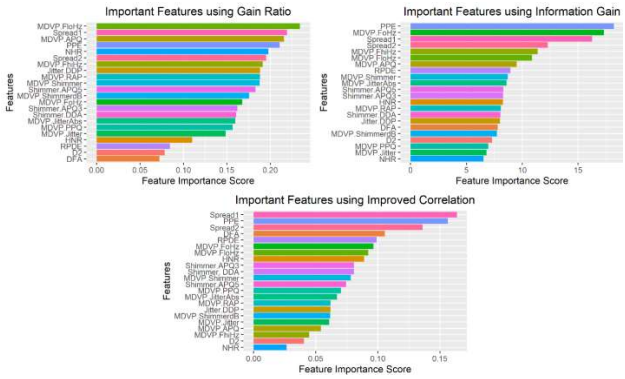

Fig.3. Feature Ranking results for each ranking technique

For each classifier, a 10-fold cross validation was used during the experimentation.

A classifier's effectiveness was evaluated using its accuracy, precision, recall, and F-score. Table.3 displays the experimental findings for several machine learning techniques with and without feature selection techniques.

Table.3. Performance of Classifiers without Features Selection

Performance Classifier	Accuracy	Precision	Recall	F-Score
Support Vector Machine	0.700	0.701	0.700	0.700
Random Forest	0.738	0.738	0.738	0.737
Adaptive Random Forest	**0.804**	0.807	0.804	0.804

Table.4. Performance of Classifiers with Features Selection based on Gain Ratio

Performance Classifier	Accuracy	Precision	Recall	F-Score
Support Vector Machine	0.717	0.718	0.717	0.716
Random Forest	0.800	0.802	0.801	0.809
Adaptive Random Forest	**0.829**	0.830	0.829	0.829

Table.5. Performance of Classifiers with Features Selection based on Information Gain

Performance Classifier	Accuracy	Precision	Recall	F-Score
Support Vector Machine	0.779	0.781	0.779	0.779
Random Forest	0.817	0.820	0.817	0.816
Adaptive Random Forest	**0.838**	0.839	0.838	0.837

Table.6. Performance of Classifiers with Features Selection based on Improved Correlation

Performance Classifier	Accuracy	Precision	Recall	F-Score
Support Vector Machine	0.804	0.806	0.804	0.804
Random Forest	0.846	0.848	0.846	0.846
Adaptive Random Forest	**0.883**	0.884	0.883	0.883

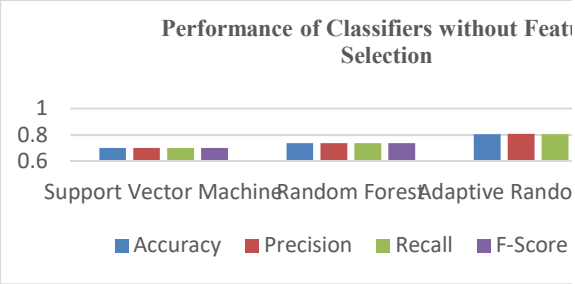

Fig.4. Performance of Classifiers without Features Selection

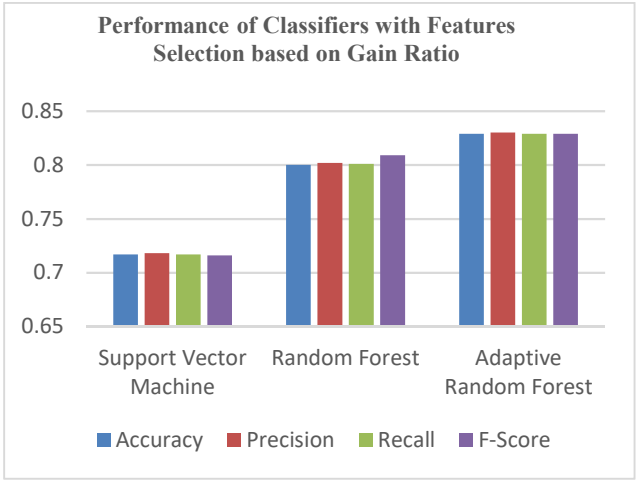

Fig.5. Performance of Classifiers with Features Selection based on Gain Ratio

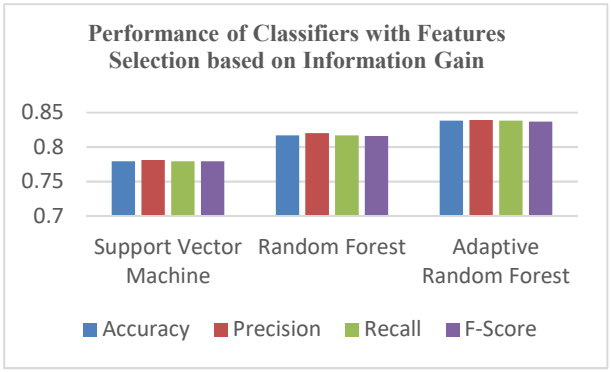

Fig.6. Performance of Classifiers with Features Selection based on Information Gain

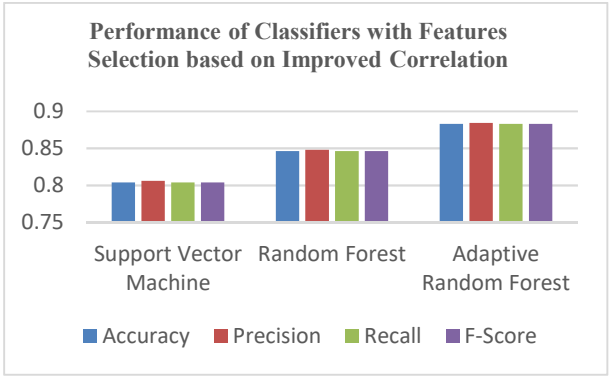

Fig.7. Performance of Classifiers with Features Selection based on Improved Correlation

6.2. Confusion Matrix

The confusion matrix and other classification performance metrics for several classification

methods on the Parkinson's dataset are shown in the following table, accordingly. Implementing the stated modules allowed the support, recall, F1-Score, and precision to be derived from the matrix of true positives, true negatives, along with false positives and false negatives. These parameters were used to compute the final accuracy.

Table.8. Confusion Matrix for different algorithms

S. No	Algorithm	Confusion Matrix			
1	SVM	**Predicated Values**		Positive	Negative
		Actual Values	Positive	108	08
			Negative	12	24
2	RF	**Predicated Values**		Positive	Negative
		Actual Values	Positive	93	22
			Negative	12	25
3	ARF	**Predicated Values**		Positive	Negative
		Actual Values	Positive	107	03
			Negative	30	12

Table.7.Performace Metrics for the three Methods using Ensemble Method

S. No	Algorithm	Metrics				
			Precision	Recall	F1_Score	Accuracy
1	SVM	Abnormal	0.76	0.68	0.72	78.34%
		Healthy	0.91	0.94	0.92	
		Avg	0.84	0.8	0.82	
2	RF	Abnormal	0.54	0.69	0.61	82.37%
		Healthy	0.89	0.82	0.86	
		Avg	0.82	0.79	0.79	
3	ARF	Abnormal	0.81	0.32	0.43	84.54%
		Healthy	0.79	0.98	0.88	
		Avg	0.81	0.78	0.74	

The conclusion drawn from the results is as described as: the Random Forest algorithm provides the best accuracy, at 82.37%, closely followed by the Adaptive Random Forest method at 84.54%. Finally, these algorithms can aid in determining whether or not a person will develop PD

6.3. ROC AnalysisParkinson's Disease Data

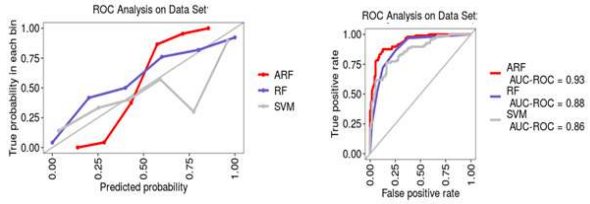

Fig.8. ROC Analysis of Adaptive Random Forest Classifier on Parkinson's Disease Data

From Figure 8, it is observed that the proposed classification method using the suggested feature selection method improved AUC-ROC values in case of data sets i.e., Parkinson's Disease Data 93 % compared to RF with 88% and SVM with 86%.

6. CONCLUSION

Medical data classification continues to be applied in case of complex and challenging tasks in the field of medical informatics. Specifically, the focus was to propose an ensemble feature ranking based methodology using a features election mechanism and later, train and build the model relying on just the features having high ranks. The performance of 10-fold cross validation is then used to generate a subset of the top-ranked features based on the ranking. Later, to avoid class imbalance problem, resampling is done using the bootstrap approach. Finally, the SVM, Random Forest, and Adaptive Random Forest algorithm is applied on the optimal subset on various real medical data sets. Finally, the proposed and standard classification models with ensemble feature selection models were compared. The experiments were done using various benchmarks which include accuracy, F-measure, and sensitivity. The proposed ensemble feature ranking models have been shown enhancing the performance of the suggested classification methods over the standard model. Further improvement of research is to predict Parkinson disease based on optimal/relevant features which are explored using present approach.

REFERENCES

[1] C. O. Sakar and O. Kursun, "Telediagnosis of Parkinson's Disease Using Measurements of Dysphonia," J. Med. Syst., vol. 34, no. 4, pp. 591–599, Aug. 2010, https://doi.org/10.1007/s10916-009-9272-y

[2] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "Classification and Prediction of atherosclerosis diseases using machine learning algorithms," in 2019 5th International Conference on Optimization and Applications (ICOA), Kenitra, Morocco, Apr. 2019, pp. 1–5. https://doi.org/10.1109/icoa.2019.8727688

[3] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "Atherosclerosis disease prediction using Supervised Machine Learning Techniques," in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, Apr. 2020, pp. 1–5. https://doi.org/10.1109/iraset48871.2020.9092082

[4] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "A fuzzy medical diagnostic support system for cardiovascular diseases diagnosis using risk factors," in 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Dec. 2018, pp. 1–6. https://doi.org/10.1109/icecocs.2018.8610649

[5] S. Laghmati, A. Tmiri, and B. Cherradi, "Machine Learning based System for Prediction of Breast Cancer Severity," in 2019 International Conference on Wireless Networks and Mobile Communications (WINCOM), Fez, Morocco, Oct. 2019, pp. 1–5. https://doi.org/10.1109/wincom47513.2019.8942575

[6] S. Hamida, O. E. Gannour, B. Cherradi, H. Ouajji, and A. Raihani, "Optimization of Machine Learning Algorithms Hyper-Parameters for Improving the Prediction of Patients Infected with COVID-19," in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, Dec. 2020, pp. 1–6. https://doi.org/10.1109/icecocs50124.2020.9314373

[7] O. El Gannour, S. Hamida, B. Cherradi, A. Raihani, and H. Moujahid, "Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID19 on X-Ray Images," in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, Dec. 2020, pp. 1–6. https://doi.org/10.1109/icecocs50124.2020.9314458

[8] O. Daanouni, B. Cherradi, and A. Tmiri, "Self-Attention Mechanism for Diabetic Retinopathy Detection," in Emerging Trends in ICT for Sustainable Development, M. Ben Ahmed, S. Mellouli, L. Braganca, B. Anouar Abdelhakim, and K. A. Bernadetta, Eds. Cham: Springer International Publishing, 2021, pp. 79–88. https://doi.org/10.1007/978-3- 030-53440-0_10

[9] V. Kate and P. Shukla, "Breast Cancer Image Multi-Classification Using Random Patch Aggregation and Depth-Wise Convolution based Deep-Net Model," Int. J. Online Biomed. Eng. IJOE, vol. 17, no. 01, p. 83, Jan. 2021, https://doi.org/10.3991/ijoe.v17i01.18513

[10] Y. Chang and F. Abu-Amara, "An Efficient Hybrid Classifier for Cancer Detection," Int. J. Online Biomed. Eng. IJOE, vol. 17, no. 03, p. 76, Mar. 2021, https://doi.org/10.3991/ ijoe.v17i03.19683.

[11] N. A. Ali, A. E. abbassi, and B. Cherradi, "The performances of iterative type-2 fuzzy Cmean on GPU for image segmentation," J. Supercomput., Jun. 2021, https://doi.org/10. 1007/s11227-021-03928-9

[12] N. A. Ali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, "New parallel hybrid implementation of bias correction fuzzy C-means algorithm," in 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, May 2017, pp. 1–6. https://doi.org/10.1109/atsip.2017.8075519

[13] N. A. Ali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, "Parallel Implementation and Performance Evaluation of some Supervised Clustering Algorithms for MRI Images Segmentation," in Proceedings of the 4th International Conference on Big Data and Internet of Things, Rabat Morocco, Oct. 2019, pp. 1–7. https://doi.org/10.1145/ 3372938.3373007.

[14] N. Aitali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, "GPU based implementation of spatial fuzzy c-means algorithm for image segmentation," in 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), Tangier, Morocco, Oct. 2016, pp. 460–464. https://doi.org/10.1109/cist.2016.7805092

[15]M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman,and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans.Biomed. Eng., vol. 56, no. 4, pp. 1010–1022, 2009.

[16] Mathur, R., Pathak, V., & Bandil, D. (2019). ParkinsonDisease Prediction Using Machine Learning Algorithm. In Emerging Trends in Expert Applications and Security (pp.357-363). Springer, Singapore.

[17] Yasar, A., Saritas, I., Sahman, M. A., & Cinar, A. C.(2019, November). Classification of Parkinson disease datawith artificial neural networks. In IOP Conference Series:Materials Science and Engineering (Vol. 675, No. 1, p.012031). IOP Publishing.

[18] Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S.(2015, March). Voiceprints analysis using MFCC andSVM for detecting patients with Parkinson's disease. In2015 International conference on electrical andinformation technologies (ICEIT) (pp. 300-304). IEEE.

[19] I. Kamran, S. Naz, I. Razzak, and M. Imran, "Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease," Future Gener. Comput. Syst., vol. 117, pp. 234–244, Apr. 2021, https://doi.org/10.1016/j.future.2020.11.020

[20] D. Purbohadi, S. Afriani, N. Rachmanio, and A. Dewi, "Developing Medical Virtual Teaching Assistant Based on Speech Recognition Technology," Int. J. Online Biomed. Eng. IJOE, vol. 17, no. 04, Art. no. 04, Apr. 2021. https://doi.org/10.3991/ijoe.v17i04. 21343

[21] M. A. Mazumder and R. A. Salam, "Feature Extraction Techniques for Speech Processing: A Review," Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 1.3, pp. 285–292, Aug. 2019, https://doi.org/10.30534/ijatcse/2019/5481.32 019

[22] O. Terrada, S. Hamida, B. Cherradi, A. Raihani, and O. Bouattane, "Supervised Machine Learning Based Medical Diagnosis Support System for Prediction of Patients with Heart Disease," Adv. Sci. Technol. Eng. Syst. J., vol. 5, no. 5, pp. 269–277, 2020, https://doi.org/ 10.25046/aj050533

[23] B. Sakar, G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease," PLOS ONE, vol. 12, p. e0182428, Aug. 2017, https://doi.org/10.1371/journal.pone.0182428

[24] H. Karimi Rouzbahani and M. R. Daliri, "Diagnosis of Parkinson's Disease in Human Using Voice Signals," Basic Clin. Neurosci., vol. 2, no. 3, pp. 12–20, Apr. 2011.