

# **A MODIFIED MAXIMUM RELEVANCE MINIMUM REDUNDANCY FEATURE SELECTION METHOD BASED ON TABU SEARCH FOR PARKINSON'S DISEASE MINING**

Waheeda Almayyan

Computer Information Department, Collage of Business Studies, PAAET, Kuwait

## **ABSTRACT**

*Parkinson's disease is a complex chronic neurodegenerative disorder of the central nervous system. One of the common symptoms for the Parkinson's disease subjects, is vocal performance degradation. Patients usually advised to follow personalized rehabilitative treatment sessions with speech experts. Recent research trends aim to investigate the potential of using sustained vowel phonations for replicating the speech experts' assessments of Parkinson's disease subjects' voices. With the purpose of improving the accuracy and efficiency of Parkinson's disease treatment, this article proposes a two-stage diagnosis model to evaluate an LSVT dataset. Firstly, we propose a modified minimum Redundancy-Maximum Relevance (mRMR) feature selection approach, based on Cuckoo Search and Tabu Search to reduce the features numbers. Secondly, we apply simple random sampling technique to dataset to increase the samples of the minority class. Promisingly, the developed approach obtained a classification Accuracy rate of 95% with 24 features by 10-fold CV method.*

## **KEYWORDS**

*Parkinson's disease; Medical data mining; maximum Relevance Minimum Redundancy, Tabu Search & Simple Random Sampling.*

## **1. INTRODUCTION**

Parkinson's disease (PD) is a complex progressive, degenerative universal disorder of the central nervous system. PD is the second most common neurodegenerative disorder in elderly. According to the World Health Organization (WHO), it was estimated that there are about 7-10 million PD patients in the world [1]. Two percent of the population above the age of 60 is affected by the PD. It ultimately leads to several motor and non-motor manifestations. The cause of PD is still unknown and early signs may be mild and unnoticeable, the diagnoses of PD disease mainly relies on clinical criteria [2]. Research has shown that approximately 90 percent of PD patients face some form of vocal deficiency [3].

Classification systems play a major role in machine learning and knowledge-based systems. In the last 30 years Computer-Aided Diagnosis (CAD) has become one of the vital research topics in medical diagnostic tasks. Medical practitioners depend upon the experience beside the existing information, whereas, CAD usually apply intelligent machine learning techniques to help physicians in making decisions. A number of published articles suggested several strategies to process the physician's interpretation and decisions [4]. During classification process, accuracy

of clinically diagnosed cases is particularly important issue to be considered. But due to the voluminous and heterogeneous nature of medical data: classification techniques are required to make accurate decisions. The size of most medical datasets is normally large-scaled; accordingly, this matter effects the complexity of the data mining process [5]. Therefore, dimension reduction techniques play a major role in excluding irrelevant features from medical datasets [6]. Therefore, instead of using all existing features, dimension reduction procedure aims to reduce computational complexity with the possible advantages of enhancing the overall classification performance and eventually reduce the computational cost.

This article proposes a two-stage feature selection algorithm to assess the PD habitation speech treatment based on Tabu and minimum redundancy - maximum relevance (mRMR) search techniques. Firstly, we select the first set of features by applying Tabu Search to construct a first reductive feature set. Secondly, we apply a modified mRMR feature selection approach, based on Cuckoo Search to further reduce the dimension and to discover the important relationship between features. Finally, we apply the random sampling technique to handle the imbalanced distribution of the medical data. This paper attempts to extend the work of [7] to deal with the limited volume of data and improve the accuracy of the prediction event.

The reminder of this article is organized in the following manner; introduction and related research are briefly described in sections 1 and 2. Section 3 explains the theoretical approach of feature selection methods and the proposed technique. The dataset, evaluation procedure and the experimental results are presented in sections 4 and 5. Finally, conclusion is summarized in Section 6.

## 2. RELATED WORK

Several studies have been carried out by numerous researchers on the early diagnosis of PD based on machine learning methods [8-20]. Little and his co-researchers conducted an important study to detect PD by introducing a new measure of dysphonia [8]. This investigation has presented a new dysphonia measure, known as the pitch period entropy. They employed Support Vector Machine (SVM) classifier with Gaussian radial basis kernel functions to predict PD and performed feature selection to select the optimal subset of features. Their best overall accuracy rate was 91.4%. In 2010, Das compare Neural Networks (ANN), DMneural, Regression and Decision Tree to predict PD. The results have shown that the ANN classifier approach outperformed the other two models with an overall accuracy of 92.9% [9]. Authors in [10] introduced a nonlinear model based on Dirichlet mixtures. The experimental results showed that the suggested model outperformed the multinomial logit models, decision trees and SVM with an overall accuracy of 87.7%.

Sakar and Kursun (2010) selected a minimal subset of dysphonia features according to the Maximum Relevance Minimum Redundancy (mRMR) criterion. An accuracy of 92.75% with the SVM classifier was obtained [11]. Psorakis et al. (2010) introduced sample selection strategies, novel convergence measures and model improvements for multiclass multi-kernel relevance vector machines (mRVMs), and finally, the improved mRVMs achieved the classification accuracy rate of 89.47% [12]. Researchers in [13] built a classification model based on genetic programming and the expectation maximization algorithm to detect PD. According to the experimental results the best classification accuracy obtained during the experiments was 93.1%. Luukka (2011) suggested a feature selection model based on fuzzy entropy measures together with the similarity classifier to predict PD and an additional three medical datasets. The mean classification accuracy of predicting PD was 85.03% with only two features from original 22 features [14].

Authors in [15] achieved an accuracy of 93.47% with a model based on non-linear fuzzy transformation method in combination with the principal component analysis together with SVM classifier. Ozcift and Gulten (2011) started with reducing feature set with the correlation-based feature selection (CFS) algorithm. Then, they applied the classification of the rotation forest ensemble classifiers that composed of 30 machine-learning algorithms to achieve an average accuracy of 87.13% [16]. For the prediction of PD, AStröm and Koker (2011) proposed 9 parallel feed-forward NN structure. The output of each neural network was evaluated using a rule-based system to make final decision. Moreover, in the proposed approach unlearned data of each neural network during the training process is collected and used in the training set of the next neural network. The highest reported classification accuracy was 91.20% [17]. Researchers in [18] applied evolutionary-based techniques in combination with the Optimum-Path Forest classifier to reduce the number of features before detecting PD. The best classification accuracy of 84.01% was reported using Gravitational Search Algorithm technique with eight features. Sriram et al. (2015) deployed Sieve multigram data and Survey graph to obtain the statistical analysis on the voice data. Then they performed clustering using the KStar and NNge classifiers [19]. Recently, with the help of data mining and bioinformatics analysis, several potential therapeutic drugs that may be used for prevention and treatment of Parkinson's disease were discovered [20]. This certainly open new perspectives for drugs development using AI techniques.

From this quick review, we can notice that most of the well-known classifiers models from machine learning have been utilized to improve the technological assessments of PD. So, choosing an excellent classifier is of significant importance to the tackle the PD classification problem. Also, we have noticed that much work is needed to tackle the imbalanced nature of medical datasets. Therefore, we will study the impact of the data pre-processing strategy that uses data sampling after applying several attribute selection techniques, on the classification models constructed. Aiming at improving the efficiency and effectiveness of the classification accuracy for PD diagnosis with minimum number of features, in this article, a classification system based on two-level of feature extraction is introduced. In this work we will focus on choosing the appropriate features to describe the feature set based on Random Forest as a base classifier.

### 3. THE PROPOSED APPROACH

Experts address the rapidly rising number of deaths from PD. New study predicted that by the year 2030 there will be around 1.2 million people living with PD in the United States [21]. A practically collected LSVT dataset [8] is used in this article to demonstrate the proposed procedure. More details on the dataset will be introduced in the next section. The proposed method for the purpose of diagnosis of PD applied in this study is illustrated in Figure 1. The main objective of the proposed approach is to explore the performance of PD diagnosis using a two-stage hybrid modelling procedure through integrating mRMR with Tabu Search technique. Firstly, the proposed method adopts mRMR to construct the discriminative feature space through Cuckoo search algorithm, then the Tabu technique is applied to help in improving the discrimination ability of classifiers. In the last step the resulted feature space is processed by the random resampling technique and ready to be evaluated with different types of activation functions to perform the classification.

Eventually, five machine-learning classification methods, which are considered very robust in solving non-linear problems, are chosen to estimate the PD possibility. These methods include Random Forest (RF), Fuzzy Unordered Rule Induction Algorithm (FURIA), AdaBoost, Rough sets and C4.5. In order to evaluate the prediction performance of the suggested model, we used six performance metrics, Sensitivity, Specificity, Accuracy, Precision,  $f$ -measure, and MCC, to test the classification performance.

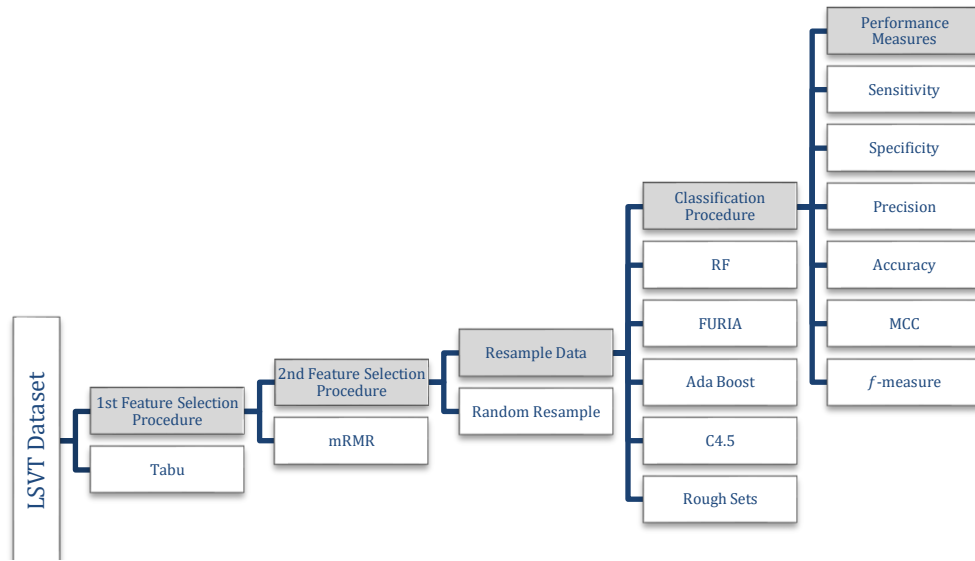


Figure.1. The proposed methodology

### 3.1. mRMR feature selection method based on Cuckoo search

The mRMR method helps in extracting crucial features and this can minimize the classification error [22]. The mRMR is a heuristic technique approach proposed by Peng et al. to measure the relevancy and redundancy of features and determine the informative features [23]. Several experimental results denoted that mRMR is an effective method to improve the performance of feature selection [24,25]. The mRMR feature selection technique is meant to identify a set of crucial features that have maximum relevancy for target classes and minimum redundancy with other features in the dataset at the same time.

The basic concept of mRMR method is to use two mutual information MI operations: one between classes and each feature in order to measure the relevancy, while the second mutual information between every two features to evaluate the redundancy.  $S$  denotes the selected features and  $Rl$  measures the relevancy of a group of selected features  $S$  that can be defined as follows

$$Rl = \frac{1}{|S|} \sum_{G_x \in S} I(G_x, C) \quad 1$$

where  $I(G_x, C)$  denoted the value of mutual information between an individual feature  $G_x$  that belongs to  $S$  and the target class  $C$ . When the selected features have the maximum relevance  $Rl$  value, it is possible to have high redundancy between these features. Hence, the redundancy  $Rd$  of a group of selected features  $S$  is defined as

$$Rd = \frac{1}{|S|^2} \sum_{G_x, G_y \in S} I(G_x, G_y) \quad 2$$

Where  $|S|$  is the number of feature in  $S$  and  $I(G_x, G_y)$  is the mutual information between the  $x$ th and  $y$ th features that measures the mutual dependency of these two features.

*The main purpose of applying the mRMR feature selection method is to find a subset of features from  $S$  with  $m$  features,  $\{x_i\}$ , that either jointly have the largest dependency on the target class  $C$  or have the minimal redundancy on the selected features subset  $S$ . Thus, Peng et al. [23] recommend searching for balanced solutions through the composite objective. This criterion combines the two criteria, which are maximal relevance criterion and minimal redundancy criterion, as follows*

$$\max (Rl, Rd) = Rl - Rd \quad 3$$

Our goal is to increase the prediction accuracy and reduce the number of selected features. Thereafter, features are selected one by one by applying Cuckoo search to maximize the objective function, which is a function of relevance and redundancy. Hence, we applied the mRMR method to filter irrelevant and noisy features and eventually reduces the computational load.

### 3.2. Tabu Search

Tabu Search is a memory-based metaheuristic algorithm proposed by Glover in 1986 to solve combinatorial optimization problems [26,27]. Since then, Tabu Search has been successfully applied in other feature selection problems [28]. This search technique is a local neighborhood search algorithm that simulates the optimal characteristics of human memory functions. Tabu Search involves a local search combined with a tabu mechanism. It starts with an initial feasible solution  $X' \in \Omega$  among the neighborhood solutions, where  $\Omega$  is the set of feasible solutions, and at each iteration, the algorithm searches the neighborhood of the best solution  $N(X) \subseteq \Omega$  to obtain a new one with an improved functional value. A solution  $X' \in N(X)$  can be reached from  $X$  in two cases,  $X'$  is not included in the tabu list; and  $X'$  is included in the tabu list, but it satisfies the aspiration criterion [29]. Surely, if the new solution  $X'$  is superior to  $X_{best}$ , the value of  $X_{best}$  is overridden. To avoid cycling, solutions that were previously visited are declared forbidden or tabu for a certain number of iterations and this surely improve the performance of the local search. Then, the neighborhood search is resumed based on the new feasible solution  $X'$ . This whole procedure is iteratively executed until the stopping criteria is met. After the iterative process has terminated, the current best solution so far  $X_{best}$  is considered the final optimal solution provided by the Tabu Search method [30].

---

```

Begin
    Start with an initial feasible solution  $X \in \Omega$ 
    Initialize tabu list and aspiration level
    For fixed number of iterations Do
        Generate neighbour solutions
        Find best  $X_{best}$ 
        If move from  $X$  to  $X'$  is not in tabu list Then
            Accept move and update solution
            Update tabu list and aspiration level
            Increment iteration number
        Else
            If  $\text{Cost}(X') < \text{aspiration level}$ 
                Accept move and update best solution
                Update tabu list and aspiration level
                Increment iteration number
            End If
        End If
    End For
End

```

---

Pseudocode 1: Pseudocode of of Tabu Search (TS) (Sait and Youssef, 1999[31]).

### **3.3. Simple Random Sampling**

Class imbalance problems occurs when one class is represented by a significantly larger number of instances than other classes. Consequently, classification algorithms tend to ignore the minority classes. Simple random sampling has been recommended as a method to increase the sensitivity of the classification algorithm to the minority class by scaling the class distribution [32,33]. Changing the class distribution can be conducted via different resampling strategies. However, the simple random sampling technique has gained extra attention. The advantage of a such technique is that it is external and therefore, easily transportable as well as very simple to implement [34]. Moreover, over-sampling the minority class data avoids unnecessary information loss [35].

Weiss and Provost conducted an empirical study where the authors used twenty datasets from UCI repository has showed quantitatively that classifier accuracy might be increased with a progressive sampling algorithm [32]. They used decision trees to evaluate classification performances with the use of a sampling strategy. Another important study used simple random sampling to scale the class distribution of biomedical datasets [33]. The authors measure the effect of the suggested sampling strategy by the use of nearest neighbour and decision tree classifiers. In simple random sampling, a sample is randomly selected from the population so that the obtained sample is representative of the population. Therefore, this technique provides an unbiased sample from the original data.

Regarding simple random sampling there are two approaches while making random selection, in the first approach the samples are selected with replacement where the sample can be selected more than once repeatedly with an equal selection chance. In the other approach the selection of samples is done without replacement where the sample can be selected only once, so that each sample in the dataset has an equal chance of being selected and once selected it cannot be chosen again [36].

## **4. DATASET AND EVALUATION PROCEDURE**

### **4.1. LSVT Dataset**

Tracking the early Parkinson's disease symptom progression using the speech disorders has shown a great sign in the advancement of Parkinson Disease detection. About 90% of people with Parkinson's disease present some kind of vocal deterioration. In this article, we use the dataset originally collected by Tsanas et al. [7] to analyze the impact of LSVT (Lee Silverman Voice Treatment) in the voice rehabilitation of 14 patients with PD. LSVT is an intensive speech therapy program designed to improve respiratory, laryngeal, and articulatory functions during speech in patients with Parkinson's disease [37]. Tsanas et al. measured 309 dysphonia features to evaluate whether a sustained phonation is "acceptable" or "unacceptable" according to the clinical criteria of six experts. Among the 126 phonations, 42 is labeled as "acceptable" while the remaining 84 is labeled as "unacceptable". The authors reported a classification score of 90% considering a voting scheme. RF and SVM.

Table 1 describes the class distribution, which clearly shows that the dataset is imbalanced. Out of 84phonations, 66.67% classified as unacceptable. A common problem with the imbalanced data is that the minority class contributes very little to the standard algorithm accuracy.

Table 1. Class distribution of the LSVT dataset

Index	Class name	Class size	Class distribution (%)
1	acceptable	42	33.33
2	unacceptable	84	66.67

## 4.2. Performance Analysis

When learning of imbalanced data, the measures such as Accuracy and Error Rate usually favors the majority class [38]. By using the 10-fold cross validation procedure, the accuracy of the selected classifiers was calculated using sensitivity, specificity, accuracy, precision,  $f$ -measure and MCC which are more appropriate measures for imbalanced datasets [39]. The main formulations are defined in Equations 4-10, according to the confusion matrix, which is shown in Table 2. In the confusion matrix of a two-class problem, TP is the number of true positives that represent in our case the cases with Parkinson's disease that was classified correctly. FN is the number of false negatives that represents the cases with Parkinson's disease that was classified incorrectly as healthy. TN is the number of true negatives, which represents the healthy cases that was classified as healthy. FP is the number of false positives that represents the healthy cases that was classified as Parkinson's disease cases.

Table 2. The confusion matrix

Hypothesis	Predicted patient state	
	Classified case with Parkinson's disease	Classified healthy case
<i>Hypothesis positive Parkinson's disease case</i>	True Positive TP	False Negative FN
<i>Hypothesis negative Healthy case</i>	False Positive FP	True Negative TN

The ROC (Receiver Operating Characteristic) curve is a graphical representation of the sensitivity versus the specificity index for a classifier varying the discrimination threshold value. The ROC curve is a standard tool for summarizing classifier performance over a range of trade-offs between TP and FP error rates [40]. ROC usually takes values between 0.5 for random drawing and 1.0 for perfect classifier performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad 4$$

The  $f$ -measure is therefore used as it measures the harmonic mean of the classifier's precision and Recall as

$$f\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad 5$$

Accordingly, we can define Precision and Recall as

$$\text{Precision} = \frac{TN}{FP + TN} \quad 6$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad 7$$

Considering class-imbalanced datasets, the Matthews Correlation Coefficient (MCC) is an appropriate measure that considered balanced. It can be used even if the classes are of very different in sizes, as it is a correlation coefficient between the observed and predicted classification decisions. The MCC measure falls within the range of  $[-1,1]$ . The larger the MCC coefficient indicates better classifier prediction. The MCC measure can be calculated directly from the confusion matrix using the following formula

$$MCC = \frac{TP.TN - FN.FP}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}} \quad 8$$

Sensitivity and specificity measures can be also used to improve interpretability as follows

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad 9$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad 10$$

## 5. RESULTS AND DISCUSSION

One of the interesting aspects in biomedical data mining is to build computational models with abilities to extract hidden knowledge using data mining schemes. The number of trees for RF algorithm and decision trees for Adaboost was arbitrarily set to 100, since it has been shown that the optimal number of trees is usually 64–128, while further increasing the number of trees does not necessarily improve the model's performance [41]. The MINNO (minimum total weight of the instances in a rule) of the FURIA algorithm has been set to 2.0. The performance results are the average after 10 runs of 10 folds-validation on each classifier to obtain its prediction measures. The suggested model for the purpose of evaluating LSVT data applied in this study is carried out in two major phases. In order to verify the effectiveness of the proposed model, firstly we compare with the performance of the selected classifier on the original feature space. Table 3, shows the classification results on the original feature space. It can be observed that the classification Accuracy fluctuates between 75% and 84% with the whole feature set. RF has achieved the results of 94%, 64.3%, 84%, 88.8% and 66.8% in terms of Sensitivity, Specificity, Precision,  $f$ -measure and MCC.

Table 3. Performance measures on the original features of LSVT dataset

Classifier	Sensitivity	Specificity	Accuracy	Precision	f-measure	MCC
<b>RF</b>	0.940	0.643	0.841	0.840	0.888	0.668
<b>FURIA</b>	0.940	0.619	0.833	0.832	0.883	0.651
<b>RoughSet</b>	0.917	0.714	0.849	0.865	0.890	0.672
<b>AdaBoost</b>	0.833	0.595	0.754	0.805	0.819	0.445
<b>C4.5</b>	0.845	0.571	0.754	0.798	0.821	0.444



Table 4. Selected features of LSVT dataset

FS technique	Number of selected features	Selected features labels
<b>1. Tabu</b>	37	Jitter->F0_TKEO_prc75 Jitter->pitch_TKEO_std Shimmer- >Ampl_TKEO_mean Shimmer- >Ampl_TKEO_prc5 Shimmer- >Ampl_TKEO_prc25 Shimmer- >Ampl_TKEO_prc95 OQ->prc5_95 DFA GNE->NSR_TKEO GNE->NSR_SEO VFER->mean IMF->SNR_entropy IMF->NSR_SEO IMF->NSR_entropy MFCC_2nd coef MFCC_3rd coef MFCC_6th coef MFCC_7th coef MFCC_8th coef MFCC_10th coef 1st delta 2nd delta 5th delta-delta Ea Ed_8_coef Ed_9_coef Ed2_9_coef entropy_shannon4_9_coef entropy_log4_6_coef entropy_log4_7_coef entropy_log4_8_coef det_TKEO_mean4_10_coef det_TKEO_std4_3_coef det_TKEO_std4_4_coef det_TKEO_std4_5_coef det_TKEO_std4_6_coef det_TKEO_std4_9_coef
<b>2. Tabu-mRMR</b>	8	Jitter->F0_TKEO_prc75 Shimmer- >Ampl_TKEO_prc25 DFA IMF->NSR_SEO MFCC_2nd coef MFCC_3rd coef Ea Ed_8_coef

Next, this reduced dataset is presented to the proposed approach that further optimizes the dimensions of the data and finds an optimal set of features. At the end of this step a subset of features is chosen for the next round. The optimal features by the Tabu and mRMR techniques are listed in Table 4. It is worth noting that the number of features has remarkably reduced, compared with mRMR technique, therefore less storage space is required for the execution of the classification algorithms. In this phase we reduced the size of LSVT features from 309 to only 8 features. Table 5, shows the classification results of the first step. It can be observed that the classification Accuracy using Tabu technique varies between 76% and 82% with the selected 37

feature set. In this step RF has achieved the results of 92.9%, 61.9%, 83%, 87.6% and 62.7%, in terms of Sensitivity, Specificity, Precision,  $f$ -measure and MCC.

Table 5. Performance measures after first RF

Classifier	Sensitivity	Specificity	Accuracy	Precision	$f$ -measure	MCC
<b>RF</b>	0.929	0.619	0.825	0.830	0.876	0.627
<b>FURIA</b>	0.893	0.619	0.802	0.824	0.857	0.561
<b>RoughSet</b>	0.821	0.643	0.762	0.821	0.821	0.464
<b>AdaBoost</b>	0.857	0.714	0.810	0.857	0.857	0.571
<b>C4.5</b>	0.881	0.632	0.803	0.841	0.860	0.542

In the next step, mRMR technique is applied on the selected features and the resulted feature set is used as the inputs to the classifiers. In Table 6, we depict the performance of the classifiers after applying the second step of reduction. From this table it is noticed that the highest Accuracy rate is associated with RF classifier was 84.1% with 8 features. In Table 7, we can see the comparative results of the classification performance of the second phase that deploy mRMR algorithm to detect the most significant features. Clearly, we can observe that the mRMR helped in reducing the dimension of features. Yet, this step did not improve the overall classification performance.

Table 6. Performance measures of second stage RF

Classifier	Sensitivity	Specificity	Accuracy	Precision	$f$ -measure	MCC
<b>RF</b>	0.929	0.667	0.841	0.848	0.886	0.662
<b>FURIA</b>	0.857	0.714	0.810	0.857	0.857	0.571
<b>RoughSet</b>	0.833	0.595	0.754	0.805	0.819	0.445
<b>AdaBoost</b>	0.833	0.643	0.770	0.824	0.828	0.482
<b>C4.5</b>	0.893	0.667	0.817	0.843	0.867	0.596

Tables 7 presents the last classification results of the suggested technique for mining the LSVT dataset. The random resampling technique is applied on the second reduced dataset to increase the samples of the minority class without replacement. The training set was resized at an over-sampling rate of 100%, in order to balance the number of instances in the two classes. This step should lead to make a more distinct and balanced dataset. We can observe that the highest Accuracy rate is associated with RF classifier is 95.2% with 8 features. In this step RF has achieved the results of 96.4%, 92.9%, 96.4%, 96.4% and 89.3%, in terms of Sensitivity, Specificity, Precision,  $f$ -measure and MCC.

The results demonstrated that the reduced features are fairly sufficient to represent the dataset's class information. In terms of performance measures our proposed technique succeeded in significantly improving the classification accuracy of the minority while the classification accuracy of major class remains high. The outcomes from the suggested two-level attribute selection techniques show better results compared to datasets which are not pre-processed and also when these attribute selection techniques are used independently. It is clear now that our developed method obtained promising classification results compared to the previously published results in [7] which scored an average Accuracy of 90% using the original dataset.

Table 7. RF versus other machine learning techniques in predicting LSVT with resampling pre-processing step

Classifier	Sensitivity	Specificity	Accuracy	Precision	f-measure	MCC
<b>RF</b>	0.964	0.929	0.952	0.964	0.964	0.893
<b>FURIA</b>	0.905	0.905	0.905	0.950	0.927	0.774
<b>RoughSet</b>	0.917	0.905	0.913	0.951	0.933	0.794
<b>AdaBoost</b>	0.881	0.738	0.833	0.871	0.876	0.627
<b>C4.5</b>	0.917	0.881	0.905	0.939	0.928	0.779

According to figure 2, the hybrid mRMR/Tabu technique has helped in investigating the LSVT dataset. Starting from the raw dataset, we applied mRMR feature reduction technique to reduce the features numbers. And in the second stage we applied Tabu Search technique to identify the significant features. This helped in reducing the dataset from 309 to 8 features. Then these features we fed into five classification algorithms. The RF algorithm performed scores were notably higher in the two phases.

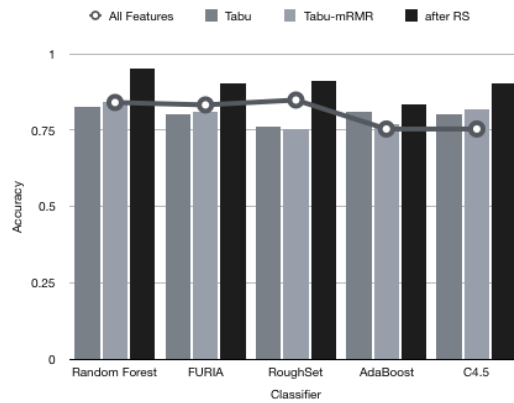


Figure 2. Comparison of proposed model based on accuracy

## 6. CONCLUSION

In this article, we have investigated a two-phase technique to improve the assessment of LSVT dataset. We concluded that the proposed algorithm could improve the accuracy performance and achieve promising results with fewer features that in the original dataset. The experiments have shown that the Tabu and mRMR based RF classification strategy helped in reducing the feature space. Whilst applying simple random sampling technique helped in adjusting the region area of the minority class in favour of handling the existing imbalanced data property.

## REFERENCES

- [ 1 ] Aarli, J.A.; Dua, T.; Janca, A.; Muscetta, A. Neurological Disorders: Public Health Challenges. World Health Organization: Geneva, Switzerland, 2006.
- [ 2 ] De Lau, L.M.; Breteler, M.M. Epidemiology of parkinson's disease. Lancet Neurol. 2006, 5, 525–535.
- [ 3 ] Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. Behavioural Neurology, 11, 131–138.

- [ 4 ] Ciosa, K.J. and Mooree, G.W. (2002) Uniqueness of medical data mining, *Artif. Intell. Med.* 26 1–24.
- [ 5 ] Huang, S.H. Wulsin, L.R. Li, H. and Guo, J. (2009) Dimensionality reduction for knowledge discovery in medical claims database: application to antidepressant medication utilization study, *Comput. Methods Programs Biomed.* 93, 115–123.
- [ 6 ] Luukka, P. (2011) Feature selection using fuzzy entropy measures with similarity classifier, *Expert Systems with Applications*, Volume 38, Issue 4, 4600-4607, ISSN 0957-4174, <http://dx.doi.org/10.1016/j.eswa.2010.09.133>.
- [ 7 ] A. Tsanas, M.A. Little, C. Fox, L.O. Ramig: Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 22, pp. 181-190, January 2014.
- [ 8 ] 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 Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015-1022, 2009.
- [ 9 ] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, 37, 1568–1572.
- [ 10 ] Shahbaba, B., & Neal, R. (2009). Nonlinear models using Dirichlet process mixtures. *The Journal of Machine Learning Research*, 10, 1829–1850.
- [ 11 ] Sakar, C. Okan, and Olcay Kursun. "Tediagnosis of Parkinson's disease using measurements of dysphonia." *Journal of medical systems* 34.4 (2010): 591-599.
- [ 12 ] Psorakis, I., Damoulas, T., & Girolami, M. A. (2010). Multiclass relevance vector machines: sparsity and accuracy. *Neural Networks, IEEE Transactions on*, 21, 1588–1598.
- [ 13 ] Guo, P. F., Bhattacharya, P., & Kharna, N. (2010). Advances in detecting Parkinson's disease. *Medical Biometrics*, 306–314.
- [ 14 ] Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, 38, 4600–4607.
- [ 15 ] Li, D. C., Liu, C. W., & Hu, S. C. (2011). A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artificial Intelligence in Medicine*, 52, 45–52.
- [ 16 ] Ozcift, A., & Gulden, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput Methods Programs Biomed*, 104, 443–451.
- [ 17 ] AStröm, F., & Koker, R. (2011). A parallel neural network approach to prediction of Parkinson's Disease. *Expert Systems with Applications*, 38, 12470–12474.
- [ 18 ] Spadoto, A. A., Guido, R. C., Carnevali, F. L., Pagnin, A. F., Falcao, A. X., & Papa, J. P. (2011). Improving Parkinson's disease identification through evolutionarybased feature selection. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (pp. 7857–7860).
- [ 19 ] Sriram, Tarigoppula V.S., Rao, M. Venkateswara, Narayana, G.V. Satya & Kaladhar, D.S.V.G.K. (2015). Diagnosis of Parkinson Disease Using Machine Learning and Data Mining Systems from Voice Dataset. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*. (pp. 151-157).

- [ 20 ] Chuan Xu, Jiajun Chen, Xia Xu, Yingyu Zhang, and Jia Li, "Potential Therapeutic Drugs for Parkinson's Disease Based on Data Mining and Bioinformatics Analysis," *Parkinson's Disease*, vol. 2018, Article ID 3464578, 8 pages, 2018. <https://doi.org/10.1155/2018/3464578>.
- [ 21 ] Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, Abbott RD, Savica R, Van Den Eeden SK, Willis AW, Tanner CM; Parkinson's Foundation P4 Group. Prevalence of Parkinson's disease across North America. *NPJ Parkinsons Dis*. 2018 Jul 10;4:21. doi: 10.1038/s41531-018-0058-0. PMID: 30003140; PMCID: PMC6039505.
- [ 22 ] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.
- [ 23 ] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [ 24 ] El A, Aouatif A, El A, Driss O (2011) A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. *Knowl Inf Syst* 26:487–500
- [ 25 ] Li A, Hu L, Niu S, Cai Y, Chou K (2012) Predict and analyze Snitrosylation modification sites with the mRMR and IFS approaches. *J Proteom* 75:1654–1665
- [ 26 ] F. Glover, "Tabu search—part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.
- [ 27 ] F. Glover, "Tabu search—part II," *ORSA Journal on Computing*, vol. 2, no. 1, pp. 4–32, 1990.
- [ 28 ] Tahir, M.A., Bouridane, A., Kurugollu, F., Amira, A., 2004a. Feature Selection using Tabu Search for Improving the Classification Rate of Prostate Needle Biopsies. In: *Proc. 17th International Conference on Pattern Recognition (ICPR 2004)*, Cambridge, UK.
- [ 29 ] Tahir, M.A., Bouridane, A., Kurugollu, F., 2004b. Simultaneous Feature Selection and Weighting for Nearest Neighbor Using Tabu Search. In: *Lecture Notes in Computer Science (LNCS 3177)*, 5th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2004), Exeter, UK.
- [ 30 ] Korycinski, D., Crawford, M., Barnes, J.W, and Ghosh, J., 2003. Adaptive feature selection for hyperspectral data analysis using a binary hierarchical classifier and Tabu Search. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS*.
- [ 31 ] Sait, S.M., Youssef, H., 1999. General iterative algorithms for combinatorial optimization. *IEEE Computer Society*.
- [ 32 ] G. Weiss and F. Provost, "Learning when Training Data are Costly: The Effect of Class Distribution on Tree Induction," *J. Artificial Intelligence Research*, vol.19,315-354,2003.
- [ 33 ] Park, B.-H., Ostrouchov, G., Samatova, N.F., Geist, A.: Reservoir-based random sampling with replacement from data stream. In: *SDM 2004*, 492-496, (2004)
- [ 34 ] Estabrooks, A., Jo, T. & Japkowicz, N., 2004. A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Computational Intelligence*, 20(1), pp.18– 36.
- [ 35 ] Yang, Y.Y. et al., 2011. Adaptive neural-fuzzy inference system for classification of rail quality data with bootstrapping-based over-sampling. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, pp.2205–2212.
- [ 36 ] Mitra SK and Pathak PK. The nature of simple random sampling. *Ann. Statist.*, 1984, 12:1536-1542.

- [ 37 ] Ramig L, Pawlas A, Countryman S. The Lee Silverman voice treatment (LSVT®): a practical guide to treating the voice and speech disorders in Parkinson disease. Iowa City, IA: National Center for Voice and Speech; 1995.
- [ 38 ] F. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing induction algorithms, Proceedings of the Fifteenth International Conference on Machine Learning, San Francisco, CA: Morgan Kaufmann, 445-453, 1998.
- [ 39 ] M.A. Maloof, Learning when data sets are Imbalanced and when costs are unequal and unknown, ICML-2003 Workshop on Learning from Imbalanced Data Sets II, 2003.
- [ 40 ] Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D: PROSO II-a new method for protein solubility prediction. FEBS J 2012, 279(12):2192-2200.
- [ 41 ] Oshiro TM, Perez PS, Baranauskas JA (2012) How many trees in a Random Forest? In: Machine learning and data mining in pattern recognition. Springer, Berlin, pp 154–168 .