



## Feature Selection Optimization using Hybrid Relief-f with Self-adaptive Differential Evolution

**M.N.Shah Zainudin<sup>1,2,\*</sup>, Md Nasir Sulaiman<sup>1</sup>, Norwati Mustapha<sup>1</sup>, Thinagaran Perumal<sup>1</sup>, Azree Shahrel Ahmad Nazri<sup>1</sup>, Raihani Mohamed<sup>1</sup>, Syaifulnizam Abd Manaf<sup>1</sup>**

<sup>1</sup>*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*

<sup>2</sup>*Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.*

\* Corresponding author's Email: noorazlan@utem.edu.my

---

**Abstract:** In various classification areas, the curse of dimensionality becomes a major challenge among the researchers. Thus, feature selection plays an important role in overcoming dimensionality problem. Relief-f is one of the filter methods to rank the most significant features based on their relevance. Although relief-f proved to be a powerful technique in filter strategy, but this method only rank the features based on their significant level. Hence, feature selection is embedded to select the most meaningful features based on their rank. Differential evolution (DE) is one of the evolutionary algorithms that are widely used in various classification domains. Simple and powerful in implementation, we combined relief-f with DE in our proposed feature selection method to solving the optimization problem. In this work, population size and generation size were adaptively determined from the number of features from relief-f. The performance of proposed method is compared with several feature selection techniques in order to prove their superiority using ten datasets obtained from UCI machine learning repository.

**Keywords:** Relief-f, Differential evolution, Evolutionary algorithm, Optimization.

---

### 1. Introduction

The aim of the classification is to classify a set of data into a number of categories. In the most classification problem, there is consists important information that represents the pattern of the class categories [1]. These parameters play the important rules in determining which category the instances will belong. These parameters are also known as attributes or features. When dealing with the large number of features, there is one challenge that might appear in various classification problems, the curse of data dimensionality. Storage, computational complexity and time are the major concern when the number of features processed is large. Thus, feature selection become a solution to help in understanding data, reducing computation requirement and storage, and at the same time improving the predictor performance [2], [3]. In another way around, the aim of the feature selection is to a select the most

relevance subset features by eliminating the irrelevant features with the good prediction result. However, it is difficult to transform original feature space into a new set of features due to vary of data representations. Hence, the further analysis needs to take into account in order to transform the features in more meaningful ways [4], [5]. In addition, to find the best features, feature selection must interact with machine learning (ML) techniques [6]. In any classification problem, feature selection method must be validating using ML classifier model. This process to ensure the selected features could optimize the classification performance [7]. In theoretically, feature selection can be grouped into two main categories; filter methods and wrapper methods [2], [8]–[10]. In filter methods, features relevance will be assess by looking at the intrinsic properties of the data. All the features will be ranked based on certain criteria. The features with higher ranking are selected and the low ranking features are removed. One of the search strategy method based

on filter approach that are widely used is using relief-f [11]. In this approach, the algorithm will rank the features based on their weighted using statistical criteria using distance measurement function. Despite the simplicity, speed, and computational efficiency, they are unable to detect indirect relationships between the features and the classifiers. Furthermore, the selection of the threshold value also needs to determine to exclude the noise. The second methods which involve optimizing a predictor of the accuracy obtained from the classifier. This method evaluates the feature subsets are obtained by training and testing on specific classifier model. However, they usually curse by the computationally expensive searches on the feature space [6] due to their computational complexity and the feature subsets grows exponentially with the number of features to search for an optimal subset.

On top of that, evolutionary algorithm also plays a part in the feature selection process. Genetic algorithm (GA) and differential evolution (DE) are the example of evolutionary algorithms that are widely used. DE is one of the simplest and most effective evolutionary algorithm used to solve high-dimensional optimization problems [12], [13]. By utilizing the same procedure like GA, DE utilizes the distance and direction information from current population in searching process. However, GA has higher potential to getting stuck at local maxima and time taken for convergence significantly high to obtain a good result. Simple but yet powerful, DE was utilized due to their advantage overtake GA. Hence, this study proposed a hybrid feature selection method using relief-f feature ranking strategy with simple and powerful DE algorithm to optimize the selection of the features. There are several contributions from our work. First, we utilized two different approaches using filter search strategy (relief-f) with an evolutionary search algorithm (differential evolution) namely RfDE with two parameters adaptively determined based on features from first stage. Second, we evaluated the performance of our proposed method using several datasets from different domain areas and proved that our proposed method outperforms other feature selection methods. This paper is organized as follows. Section 2 described the overall feature selection algorithm. Section 3 discussed the proposed feature selection method. Section 4 described the analysis and discussion. Section 5 discussed the conclusion and future direction of our work.

## 2. Feature selection algorithms

Several search methods have been reported to overcome the feature dimension issue with a different search strategy. Zou et al. [14] have proposed feature ranking strategy using Max-Relevance-Max-Distance(MRMD) to predict high-dimensionality datasets. They evaluate their proposed method using two different domain area; image and protein interaction datasets. Due to time and space complexity, Santos et al. [15] have come out with ensemble feature ranking strategy to improve the efficiency of the classification performance. They compared the performance of the proposed method with support vector machine (SVM), bagging, random forest (RF) and Naive Bayes (NB) using breast cancer dataset. Challita et al. [16] proposed a new feature selection technique using elastic-net with relief-f to recognize the signal from rotation machine. Those proposed techniques were evaluated using k-nearest neighbor (kNN) and SVM. They claimed that their proposed method presents pleasant classification error in their experiment. Capela et al. [17] presented the comparative study on feature selection method using relief-f, correlation-based feature selection (CFS) and fast correlation-based filter (FCBF) to differentiate the activity for three different groups of user; able-bodied, elderly and stroke patients. Canedo et al. [18] proposed ensemble feature selection filter using CFS, consistency-based filter, INTERACT, information gain and relief-f to reduce the data dimensionality. They concluded that; ensemble filter with C4.5 suitable when classifying classical datasets (with more samples than features) and SVM more robust when dealing with microarray dataset (with more features than samples). Khushaba et al. [10] proposed a combination of an artificial ant colony (ACO) with DE to enhance the capabilities for searching procedure. Two bio signal-driven applications were used in their study and the result shows the significance improvement in term of classification accuracy.

In any machine learning algorithm, feature selection aims to understand the data, reducing the computer requirement by minimizing the potential of being lagging and improving the performance of predictor [2]. Ghosh et al. [19] proposed a self-adaptive differential evolution (SADE) to classify hyperspectral image data. They evaluate the performance of subsets generation using fuzzy k-nearest neighbor. Apolloni et al. [6] tested microarray data to evaluate their proposed evolutionary algorithm called as binary differential

evolution (BDE) combination with rank based filter method. Four ML algorithms NB, SVM, C4.5 and kNN were utilized to evaluate their proposed method. Al-Ani et al. [20] have utilized DE to identify relevant feature subsets using two different versions. In the first version, the user needs to determine the desired number of features and the second version user only needs to set an upper limit to the feature subset size. Sikdar et al. [21] have come out with DE with a two-stages evolutionary approach for named entity recognition. They utilized conditional random field (CRF) and SVM as a classifier to evaluate the proposed method for three different Indian languages. Bharathi and Subashini [22] have presented their work on the classification of river ice types. They combined DE with extreme learning machine (ELM) to select the best feature subsets from the original feature set. Fish et al. [23] classify 14 activities from 14 tri-axial accelerometer sensors using decision tree. They utilized the combination of filter and wrapper approach based on mutual information to select the most relevant features. Idris et al. [24] investigated the significant between feature selection methods using particle swarm optimization (PSO), principle component analysis (PCA), Fisher's ratio, F-score and minimum redundancy maximum relevance (mRMR) to predict the customer's churn. They reported that PSO with mRMR using random forest (RF) performs well in predicting the customer churners.

### 3. Proposed feature selection method

This section describes the main contribution of this entire work. The subsection discusses the steps included in the proposed hybrid feature selection method. In general, by reducing the number of features, it will reduce the training time to build and minimize the complexity of the classifier model. Feature selection is to select the most relevant features before it adapted as an input to the classifier to obtained optimize solution. As a result, the less relevant or less significant features will be eliminated or removed from the list. Hybrid features selection method using feature ranking strategy namely relief-f with powerful evolutionary algorithm DE algorithm is proposed. The detail explanation of our proposed method is described in the next following subsection.

#### 3.1 Data pre-processing

In order to test the compatibility of the proposed method, 10 dataset from UCI machine learning repository database have been evaluated. The details of the datasets were explained in the analysis and

discussion section. When dealing with unprocessed data, filtered process need to undergo due to some of the values are missing. Hence, the instances that consist of missing values have been removed from the list. To validate our performance, 10 fold cross validation performance strategy was utilized during experiments. In this strategy, 10 equal sizes subsets are created and 10 runs of experiment are conducted. In every run, 1 fold reserve for testing and the rest 9 folds will be used for training. There is no feature extraction conducted in this work since the datasets obtained have been extracted.

#### 3.2 Relief-f filter search strategy approach

In 1994, Kononenko [25] has proposed an extension of relief algorithm. There are several limitations of relief algorithm which this algorithm only supports the binary class problem and unable to handle noisy or incomplete data. Thus, the extension of relief algorithm has proposed and called as relief-f. This algorithm is not limited to two class problem (binary class) but also have the advantages to cater the noisy and incomplete data [11]. This algorithm is widely used due to their simplicity and compatible to the various types of domain. Theoretically, this algorithm will rank the features and the highly ranked features will be ordered and selected as an input for predictor. This filter method that ranks features by weighting them based on quality (relevance) [17], [26]. The first instances will be selected randomly and it will find the nearest hit (data point from the same class) and nearest misses (data point from the different class). This process is calculated by weighting them based on their relevance as shown in the equation 1.

$$w_i = \sum_{j=1}^N (x_i^j - \text{nearmiss}(x^i)_j)^2 - (x_i^j - \text{nearhit}(x^i)_j)^2 \quad (1)$$

where  $w$  is the weight of the  $i^{\text{th}}$  feature,  $x_i^j$  is the value of the  $i^{\text{th}}$  feature for point  $x^j$  and  $N$  is the total number of data points. Nearhit  $x^j$  and nearmiss  $x^j$  are the nearest data point to  $x^j$  in the same and different class respectively. Hence, it will rearrange the features based on their weight and the significant level. The weight obtained of the features will be considered and the features below the certain threshold will be removed from the list. In this work, 0.02 threshold values were used and the features below this threshold value will be eliminated. Hence, only selected features (features with the highest relevance) will be used as an input to the next process. The relevant of this process is to minimize the complexity of the DE to evaluate the subset

performance by utilizing only significance features does not redundant in their input.

### 3.3 Differential evolution feature selection

Differential evolution (DE) is one of the methods from evolutionary algorithm where the features are search based on ant colony. A simple but yet powerful, DE provide the advantages generally demands like other optimization methods [12], [13]. There are several compensations from DE such; 1) ability to handle non-differentiable, nonlinear and multimodal cost functions, 2) parallelizability to cope with computation intensive cost functions, 3) ease of use, 4) good convergence properties. Like GA, DE utilize same parameters consists of mutation, crossover and selection. The performance of DE depends on the manipulation of target vector and difference vector in order to obtain a trial vector in searching process. Each D-dimensional real-value parameter, a population of NP members is generated. NP is the population size and D is the number of parameter to be optimized. In order to generate a trial vector, the weight different vector between two populations members  $x_{r2}$  and  $x_{r3}$  are added to a third member  $x_{r1}$ . This process called as a mutation. For each target vector  $x_{(l,G)}, i=1,2,3, \dots, N$ , a mutant vector is generate according to:

$$v_{i,G+1} = x_{r1,G} + F(x_{r2,G} - x_{r3,G}) \quad (2)$$

where  $r_1, r_2, r_3 \in \{1, 2, \dots, NP\}$  are randomly chosen integers, must be different from each other and also different from the running index  $i$ . Scaling factor  $F(0,1)$  use the control the rate which the population evolves. In order to increase the diversity of the perturbed parameter vectors, crossover is introduced. The trial vector:

$$u_{i,G+1} = (u_{1,i,G+1}, u_{2,i,G+1}, \dots, u_{D,i,G+1}) \quad (3)$$

is form where;

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} & \text{if } \text{rand}(0,1) \leq c_r \\ x_{ji,G+1} & \text{otherwise} \end{cases} \quad (4)$$

where  $u_{(ji,G)}$  is the  $j^{\text{th}}$  dimension from  $i^{\text{th}}$  trial vector along the current population  $G$ . The crossover probability  $c_r(0,1)$  is a user defined value that controls the fraction of the parameter values that are copied from the mutant. Selection is the step to choose the vector between the target vector and the trial vector with the aim of creating an individual for the next generation. If the newly generated vector

results in a lower objective function value (better fitness) than the predetermined population member, then the resulting vector replaces the vector with which it was compared [27]. However, several parameters from DE are automatically adaptive without required user to determine by trial and error method. In this work, population size and generation size are adaptively determine based on a number of features remained from relief-f. Hence, the user does not need to initialize those parameter values manually. All the features selected from this stage will be evaluated using ensemble classifier; random forest [28].

### 3.4 Proposed RfDE algorithm

This work is extended the work that have been done in [29] where they proposed a hybrid RfDE to recognize the human activity recognition from accelerometer sensor. Above 97% accuracy obtained using their proposed feature selection method in comparison with other state of the art feature selection methods. In this subsection, detail algorithm for proposed hybrid feature selection model RfDE was presented.

```

set all weights  $W[A] = 0$ 
for  $l = 0$  to  $m$  do
    randomly select an instance  $R_i$ 
    find nearest hit  $H$  and nearest misses  $M$ 
    for all  $A$  do
        calculate the equation (1)
    end for
end for
remove the features weightage below 0.02
list selected ranked features in ordered ( $M \times N$  matrices)
initialized ranked features as an input for DE
initialized all parameter (population size, no features (10 to 15), no generation, fitness function)
population size =  $N$  and no generation =  $N$ 
for  $i = 1$  to  $N(\text{population size})$  do
    select  $r_1, r_2, r_3 \in N$  randomly
    for  $j = 1$  to  $D(\text{dimension})$  do
        select  $J_{rand} \in D$ 
        if ( $\text{rand}() < CR$  of  $j = J_{rand}$ )
             $U_{ji,g+1} = x_{r1,g} + F * (x_{r2,g} - x_{r3,g})$  from equation (2)
        end if
        if ( $f(U_{ji,g+1}) < f(X_{ji,g})$ ) then from equation (4)
             $X_{ji,g+1} = U_{ji,g+1}$ 
        end if
    end for
end for

```

## 4. Analysis and discussion

Aforementioned in the previous section, the purpose of feature selection is to eliminate the

features that are not contributed useful information to the class category. In order to solve this issue, we come out with the new hybrid features selection method using relief-f with DE called as RfDE. In order to evaluate the performance and compatibility of our proposed method, several datasets from different domain areas were tested. All the datasets have been employed from UCI machine learning repository website. Ten datasets are used such as ionosphere, QSAR biodegradation, connectionist bench, kidney disease, diagnostic breast cancer, chess, dermatology, hepatitis, Landsat satellite and lung cancer data. The datasets obtained have been extracted and there is no feature extraction process has been performed. The details of the dataset used are listed in Table 1.

Table 1. List of datasets

Dataset	No. of Instances	No. of Features	No. of Classes
Ionosphere	351	34	2
QSAR Biodegradation	1055	41	2
Connectionist Bench	208	60	3
Kidney Diseases	400	25	2
Breast Cancer Diagnostic	369	32	2
Chess	3196	36	2
Dermatology	366	34	6
Hepatitis	155	19	2
Landsat Satellite	6435	36	6
Lung Cancer	32	56	3

In most classification areas, several performance indicators are required to evaluate the performance of any new methods. In this experiment, we utilized the overall accuracy as our performance measurement. In the first stage, the features have been ranked using relief-f filter ranking method. Since this method used distance function to measure the feature relevance, the number of neighbor  $k$  need to identify. Hence, several numbers of  $k$  (nearest neighbor) in this stage were choose to determine the optimize value of  $k$ . A different number of  $k$  will produce a different number of subsets. Three different value of  $k$  is 3, 5 and 10 were tested. As a result, relief-f will produce the ranked features based on their weight in ordered list. In order to eliminate the less significance features from the list, threshold parameter need to determine. Those features below this threshold will be removed from the list. The threshold value 0.02 was choosing as our benchmark. This criterion was choosing after a few experiments have been done and when the features from this below threshold included in the input list, it is not provide any improvement to the classification performance. Table 2 shows the number of features remain using relief-f ranking search strategy with different number of  $k$ .

Table 2. Number of feature remain using Relief-f

Dataset	No. of Original Features	K=3	K=5	K=10
Ionosphere	34	33	33	33
QSAR Biodegradation	41	15	16	12
Connectionist Bench	60	43	41	28
Kidney Diseases	25	18	16	16
Breast Cancer Diagnostic	32	24	21	21
Chess	36	14	15	15
Dermatology	34	34	34	34
Hepatitis	19	14	16	16
Landsat Satellite	36	36	36	36
Lung Cancer	56	22	19	19

From Table 2, less than half features remained using relief-f for QSAR, chess and lung cancer dataset. Fewest number of features remained recorded by QSAR datasets (12 features) when the value of  $k$  is 10 used. In comparison, there are no features eliminated for dermatology and Landsat dataset. All the features from both of these datasets remained and might be relevance and contribute the meaningful information to determine the class category. Only one feature has been eliminated for ionosphere data when three different values of  $k$  were used. For connectionist bench datasets, more than 50% of features removed when the  $k$  is 10 were used. However, there are no huge different number of features retained when  $k$  is 3 and 5. The range between 16 to 18 features was selected for kidney dataset. About 65% features remained in breast cancer dataset when the  $k$  is 5 and 10 were choose. For hepatitis dataset, only 3 to 5 features have been removed using this method. After the first stage has been completed, the experiment will be conducted in the second stage. All the features remained from different values of  $k$  have been used in the next process. In order to assess the performance of selected features from Table 2, random forest classifier [15], [24] was utilized in our experiment. Table 3 shows the classification performance from different values of  $k$ .

Table 3. Classification accuracy for various values of  $k$ 

Dataset	K=3	K=5	K=10
Ionosphere	0.929	0.932	0.932
QSAR Biodegradation	0.858	0.854	0.841
Connectionist Bench	0.832	0.851	0.842
Kidney Diseases	0.995	0.995	1.000
Breast Cancer Diagnostic	0.962	0.957	0.957
Chess	0.971	0.969	0.969
Dermatology	0.972	0.972	0.972
Hepatitis	0.824	0.833	0.833
Landsat Satellite	0.912	0.912	0.912
Lung Cancer	0.722	0.722	0.722
Average	0.898	0.900	0.898

From Table 3, it can clearly be seen that the different value  $k$  produced different accuracy performance and shows the acceptable performance. For QSAR datasets, the accuracy was slightly

decreased about 0.4% and 1.3% when the k is 5 and 10 respectively. K is 3 obtained the highest accuracy with 85.8%. Kidney and breast cancer datasets also recorded a small decrease about 0.5% when the value of k is 3 and 5 were used. 0.2% declined in the chess datasets when the value of k chooses is 5 and 10. When the value k is 3, highest accuracy recorded for chess and breast cancer datasets, 99.5% and 96.2% respectively. For ionosphere, about 0.3% increased when k is 5 or 10. The highest accuracy for bench datasets is 85.1% (k=3) followed by 84.2% (k=10) and 83.2% (k=3). There is no improvement in accuracy for dermatology, Landsat and lung cancer datasets when a different number of k was used. In average, the value of k is 5 considered is the optimize number of neighbor and achieved the highest accuracy obtained for all datasets. In order standardize the selection the value of neighbor; k is 5 was choose in our next experiments.

Next, after the features have been selected from the first stage, we precede the experiment to the second stage. The selected features from relief-f have been used as an input to the DE. In DE, there are several parameters need to be determined. Number of features required, number of populations, number of generation and fitness function are the parameters need to be identified. For the number of desired features, the range of the value between 10 to 15 features was used. These criteria chosen since different data will produce a different number of features with different accuracy performance. To determine the number of population and the number of generation, those numbers adaptively selected from the number of original features obtained from relief-f method. Hence, every dataset will have a different number of population and generation. KNN was used as fitness function for errors subsets estimation. Table 4 shows the classification performance using RfDE using random forest classifier. Default parameters value for classifier has been used in our work.

Table 4. Classification accuracy using RfDE

Dataset	No. of Features	Subset Error	Accuracy
Ionosphere	13	5.6604	0.927
QSAR Biodegradation	13	12.6183	0.840
Connectionist Bench	15	1.5625	0.826
Kidney Diseases	13	3.3333	0.993
Breast Cancer Diagnostic	13	5.8140	0.957
Chess	13	4.8958	0.971
Dermatology	15	0.8929	0.976
Hepatitis	10	6.3830	0.852
Landsat Satellite	15	8.9500	0.905
Lung Cancer	15	11.1111	0.667
Average	13.5	6.1221	0.891

From Table 4, it can be clearly seen that the error obtained from all the datasets recorded from 0.8 to 11.1. The lowest error recorded from dermatology datasets with 0.8929. Lung cancer reported the highest error rate obtained about 11.1111. In term of a number of features remained, 5 datasets obtained 13 features, 4 datasets obtained 15 features and only 1 datasets recorded the lowest number of features obtained (10 features). In average, the optimal number of features selected is around 13 to 15 features. Average errors obtained from all datasets were 6.1221. In term of average accuracy performance, 89.1% accuracy recorded for all datasets. It can be proved that our proposed method achieved a good accuracy performance almost 90%. In order to validate our proposed method, several benchmark feature selection methods have been applied. A comparative study between our proposed method with several well-known states of the art feature selection methods namely Tabu search (TS) [30], genetic algorithm (GA) [31] and particle swarm optimization (PSO) [24] were compared. In this experiment, we did a comparison with the work have been done by [13]. They utilized their proposed feature selection method using DE with wavelet packet transform to select the significant features and compare their performance with other methods such PSO and GA. They reported that the performance of their method show remarkable results for both accuracy and number of features for EEG signal used. Hence, we modify the original DE with our proposed method called RfDE with several other datasets. The comparisons of the classification performance are shown in Table 5. The value in the bracket represents the number of selected features.

Table 5. Comparison of Classification Accuracy using RfDE with Various Feature Selection Algorithms

Dataset	TS	GA	PSO	RfDE
Ionosphere	0.922(19)	0.910(27)	0.914(26)	<b>0.927(13)</b>
QSAR Biodegradation	0.862(23)	0.859(37)	<b>0.863(33)</b>	0.840(13)
Connectionist Bench	0.757(19)	0.764(42)	0.785(27)	<b>0.826(15)</b>
Kidney Diseases	0.989(10)	<b>0.996(13)</b>	<b>0.996(13)</b>	0.993(13)
Breast Cancer	<b>0.960(14)</b>	0.957(19)	<b>0.960(14)</b>	0.957(13)
Chess	0.939(9)	0.939(14)	0.939(9)	<b>0.971(13)</b>
Dermatology	0.972(19)	0.972(28)	0.972(28)	<b>0.976(15)</b>
Hepatitis	0.833(12)	0.843(13)	0.843(13)	<b>0.852(10)</b>
Landsat Satellite	0.911(27)	<b>0.914(30)</b>	0.911(26)	0.905(15)
Lung Cancer	<b>0.778(9)</b>	0.722(13)	0.722(10)	0.667(15)
Average	0.892	0.888	0.891	0.891

From Table 5, it is clearly shown that by using our proposed method, it will significantly improve

the accuracy of classification performance with minimal number of features required. It can clearly be seen that half of the datasets achieved the acceptable accuracy performance with a minimum number of features and considered as a good performance. 92.7% accuracy for ionosphere and 97.1% for chess dataset with 13 numbers of features obtained using RfDE. Bench data recorded 82.6% and dermatology recorded 97.6% with 15 numbers of selected features for both of them. 10 features considered the most relevant features for hepatitis dataset and obtained the highest accuracy 85.2% compared with other algorithms. All the five datasets outperform the others feature selection algorithms such as TS, GA and PSO in term of accuracy performance. Even though PSO recorded the highest accuracy for biodegradation datasets but, the total number of features required were 33 which considered large compared with our proposed method only 13 features. 99.6% accuracy obtained with 13 features using GA and PSO for kidney datasets. A small decreases about 0.3% when the same number of features applied in RfDE. 14 features were selected using TS and PSO for breast cancer datasets to produce 96.0% accuracy. Landsat satellite datasets obtained accuracy 91.4% using GA with 30 features followed by 91.1% using TS and PSO with 27 features respectively. In comparison with GA, our method performed better in minimization of a number of features. 15 features were sufficient to achieve above 90% accuracy for Landsat dataset. In TS, only 9 features were adequate to recognize the types of lung cancer and obtained the highest accuracy about 77.8% compared with other algorithms. In average, the highest accuracy obtained was from TS algorithm 89.2%. Both of the PSO and RfDE achieved 89.1% accuracy respectively followed by GA 88.8%. In summarizing, we concluded that our proposed method compatible with various types of domain area with acceptable accuracy performance. In addition, a minimum number of features range from 10 to 15 are sufficient to recognize the class of the instances. Lung cancer reported the lowest classification accuracy since the number of instances from this dataset was very low compared to others; only 32 instances. It might difficult to any classifier model to classify the pattern of the data with very a minimal number of instances. We also noticed that our proposed method not really fitting for data with a minimum number of features. It might minimize the chances of the selection of the most significant features used in the classification process. Previous works show that most of the features ranking search strategy only rearrange the features based on their

ordered without eliminate those features considered does not give any meaningful information to the class category. Threshold value 0.02 considered a significantly produce a good performance in our experiments. Hence, to optimize the selected features, we utilized the method propose by [20] using DE. To sum up, we proved that our proposed method significantly shows the good performance as reported by authors [10], [13], [20].

## 5. Conclusion

One of the challenges claimed among the researchers in the classification areas is the curse of data dimensionality. When the dimension of the data too big, it tends to increase the time, cost and complexity of classifier model and might significantly reduce the performance of the classifier model due to some of the features might not meaningful to determine the class of the instance. Specifically, this work presented the hybrid feature selection using relief-f with differential evolution to optimize the performance of the selected features. Relief-f used to rank the features based on their significance based on selected threshold before it is fed into DE. DE will perform the mutation to generate the mutant vector and generate the trial vector based on crossover process. The trial vector will be compared with target vector in order to select the most significance features from the list. Random forest classifier is used for evaluating the subsets from DE output. Our work proved that by combining the feature ranking strategy with the exploration searching based on ant's populations give advantages in the complexity of the searching space. In addition, the advantages of relief-f to handle the incomplete and noisy data also may lead a good performance of the result. The proposed methods have been compared with several well-known feature selection techniques such as Tabu search, genetic algorithm and particle swarm optimization for ten datasets from different domain areas obtained from UCI repository. Experiments on a number of datasets with different sizes proved that the proposed method can achieve remarkably good results when compared with some of the powerful feature selection methods. Our work shows acceptable result average 89.1% accuracy with minimal number of features required (10 to 15 features). It is clearly shows that using our proposed method will produce good performance with less number of features by minimizing the complexity of the classifier model. Furthermore, half of the datasets produced good accuracy performance in comparison with other three feature selection

methods mentioned above. There are several outcomes in this work; we shows that by using proposed feature selection method, minimal number of features sufficient to produce a good performance for medium sizes number of features. Those proposed method also able to classify for both integer and real number of attributes values. However, there are plenty of spaces for researchers to enhance their research in this area. Imbalance data distribution considered as a challenge that needs to take into account since recent work shows that imbalance data considered as the factors contributed to the classification performance. In projection, we will evaluate our proposed method using other classification methods such as deep learning strategy to cater the problem of imbalanced data distribution.

### Acknowledgments

This work was supported by the Universiti Putra Malaysia funded by Ministry of Higher Education Malaysia.

### References

- [1] R. A. Rahim, N. Othman, M. N. S. Zainudin, N. A. Ali, and M. M. Ismail, "Iris Recognition using Histogram Analysis via LPQ and RI-LPQ Method," no. 4, pp. 66–70, 2012.
- [2] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.
- [3] J. Tang, S. Alelyani, and H. Liu, "Feature Selection for Classification: A Review," *Data Classif. Algorithms Appl.*, pp. 37–64, 2014.
- [4] M. N. S. Zainudin, M. Mohd Said, and M. . Ismail, "Feature Extraction on Medical Image Using 2D Gabor Filter," *Appl. Mech. Mater.*, vol. 52–54, pp. 2128–2132, 2011.
- [5] M. N. S. Zainudin, H. . Radi, and S. . Abdullah, "Face Recognition using Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA)," *Ijens.Org*, no. 5, 2012.
- [6] J. Apolloni, G. Leguizamón, and E. Alba, "Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments," *Appl. Soft Comput. J.*, vol. 38, pp. 922–932, 2016.
- [7] M. N. S. Zainudin, N. Sulaiman, N. Mustapha, and T. Perumal, "Activity Recognition based on Accelerometer Sensor using Combinational Classifiers," pp. 68–73, 2015.
- [8] V. Bolon-Canedo, N. Sanchez-Marono, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, 2013.
- [9] J. A. Olvera-Lopez, J. A. Carrasco-Ochoa, J. F. Martinez-Trinidad, and J. Kittler, "A review of instance selection methods," *Artif. Intell. Rev.*, vol. 34, no. 2, pp. 133–143, 2010.
- [10] R. N. Khushaba, A. Al-Ani, A. AlSukker, and A. Al-Jumaily, "A Combined Ant Colony and Differential Evolution Feature Selection Algorithm," *Ant Colony Optim. Swarm Intell.*, vol. 5217, pp. 1–12, 2008.
- [11] M. Robnik-Siknja and I. Kononeko, "Theoretical and empirical analysis of Relief and RRelief," *Mach Learn*, vol. 53, pp. 23–69, 2003.
- [12] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, pp. 341–359, 1997.
- [13] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Feature subset selection using differential evolution and a statistical repair mechanism," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11515–11526, 2011.
- [14] Q. Zou, J. Zeng, L. Cao, and R. Ji, "A novel features ranking metric with application to scalable visual and bioinformatics data classification," *Neurocomputing*, vol. 173, pp. 346–354, 2016.
- [15] V. Santos, N. Datia, and M. P. M. Pato, "Ensemble Feature Ranking Applied to Medical Data," *Procedia Technol.*, vol. 17, pp. 223–230, 2014.
- [16] N. Challita, M. Khalil, and P. Beausery, "New technique for feature selection: combination between Elastic Net and Relief," 2015, pp. 3–8.
- [17] N. A. Capela, E. D. Lemaire, and N. Baddour, "Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients," *PLoS One*, vol. 10, no. 4, pp. 1–18, 2015.
- [18] V. Bolon-Canedo, N. Sanchez-Marono, and A. Alonso-Betanzos, "Data classification using an ensemble of filters," *Neurocomputing*, vol. 135, pp. 13–20, 2014.
- [19] A. Ghosh, A. Datta, and S. Ghosh, "Self-adaptive differential evolution for feature selection in hyperspectral image data," *Appl. Soft Comput. J.*, vol. 13, no. 4, pp. 1969–1977, 2013.
- [20] A. Al-Ani, A. AlSukker, and R. N. Khushaba, "Feature subset selection using differential evolution and a wheel based search strategy," *Swarm Evol. Comput.*, vol. 9, pp. 15–26, 2013.
- [21] U. K. Sikdar, A. Ekbal, and S. Saha, "Differential Evolution based Feature Selection

- and Classifier Ensemble for Named Entity Recognition,” *Proc. COLING 2012*, vol. 11, no. i, p. 11218, 2012.
- [22] P. T. Bharathi and P. Subashini, “Optimal Feature Subset Selection Using Differential Evolution and Extreme Learning Machine,” *Int. J. Sci. Res.*, vol. 3, no. 7, pp. 1898–1905, 2014.
- [23] B. Fish and a Khan, “Feature selection based on mutual information for human activity recognition,” ... , *Speech Signal ...*, pp. 1729–1732, 2012.
- [24] A. Idris, M. Rizwan, and A. Khan, “Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies,” *Comput. Electr. Eng.*, vol. 38, no. 6, pp. 1808–1819, 2012.
- [25] I. Kononenko, “Estimating attributes: Analysis and extensions of RELIEF,” *Mach. Learn. ECML-94*, vol. 784, pp. 171–182, 1994.
- [26] R. Akhavian and A. H. Behzadan, “Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers,” *Adv. Eng. Informatics*, vol. 29, no. 4, pp. 867–877, 2015.
- [27] A. K. Palit and D. Popovic, *Computational Intelligence in Time Series Forecasting*. 2005.
- [28] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [29] M. N. S. Zainudin, M. N. Sulaiman, N. Mustapha, T. Perumal, and A. S. Ahmad Nazri, “Hybrid Relief-f Differential Evolution Feature Selection for Accelerometer Actions,” *Adv. Sci. Lett.*, vol. 4, no. 2, pp. 400–407, 2016.
- [30] A. M.a, K. A.a, and A. S.I.b, “Classification of Physical Activities Using Wearable Sensors,” *Intell. Autom. Soft Comput.*, no. September 2016, pp. 1–10, 2015.
- [31] S. Chernbumroong, S. Cang, and H. Yu, “Genetic Algorithm-Based Classifiers Fusion for Multisensor Activity Recognition of Elderly People,” vol. 19, no. 1, pp. 282–289, 2015.