# Feature Ranking based Dimensional Reduction Algorithm in High Dimensional Data Classification

Thinzar Saw
*University of Computer Studies, Mandalay*
Mandalay, Myanmar
thinzarsaw@ucsm.edu.mm

Si Si Mar Win
*University of Computer Studies, Mandalay*
Mandalay, Myanmar
sisimarwin@ucsm.edu.mm

***Abstract*—Features are the most important things and huge influence on machine learning models. The removing of noisy and irrelevant features causes to increase the efficiency and accuracy of the model. The recent increase in feature dimension presents a significant challenge to many current methods of feature selection in terms of learning accuracy. In this work, combined correlation with distance measures approach is presented for the selection of informative features in high dimensional classification problem. Firstly, Pearson Correlation Coefficient is used to measure the relevance between attributes and class. And then, to calculate the redundancy of two feature vectors, Euclidean Distance is applied. To evaluate the presented approach, various classifiers with ten-fold cross validation are utilized based on five well-known datasets from UCI machine learning repository. The results show that the attribute subsets of all five well-known datasets generated by the proposed algorithm are more likely to be truly representative of classifier performance compared with original attribute set.**

***Keywords—Feature Selection, High Dimensional Data Classification, Feature Ranking, Distance Measures, Pearson Correlation Coefficient.***

## I. INTRODUCTION

In recent times, huge volumes of data have become widely available in data mining and information systems has appealed a great deal of interest from researchers in converting such data into useful knowledge. This indicates the presence of low value, inaccurate, repetitive and unrelated data which have a negative impact on the information observation process and useful strategy. As a result, researchers need to use feature selection methods to extract relevant data from huge facts. Feature selection is the most recognizable method to identify related attributes, and remove unnecessary attributes. How can the most powerful features in the original dataset be picked from the feature vector?

There are essentially two types of attribute selection: the wrapper method and the filter method [1]. The subset evaluator is used, in the wrapper process, to build all possible sub-sets from the feature vector. Formerly, a learning algorithm such as Naïve Bayes, kNN and C4.5 is used to render classifiers from the features of each subset. And eventually, it takes into account the feature subset for which the classifier performs the finest. To find a subset, the evaluator uses a search technique, such as random, breadth first and hybrid search. The possible subsets assessed from the feature vector are then based on these techniques of searching. The previous work [2] has formerly explored this method for healthcare data with small attributes and instances using particle swarm optimization (PSO) search through various classifiers.

An attribute evaluator and a ranker in the filter method rank all features in the dataset. It's always possible to specify the number of features selected from the function vector. After having the rank features, the features with lower ranks are omitted and the highest rank can be kept for the predictive accuracy of the learning algorithm. One problem with the filter approach is that the weights placed by the ranker algorithms by the classification algorithm are different from those weights.

Additionally, dependence measures are one of the measures used for feature selection. Many approaches based on dependence have been suggested. The main criterion is correlation based method. In order to find the association between the continuous features and the class function, Pearson's Correlation approach is used. At [3], measures of symmetric uncertainty are used to find the relation between the class function and the discrete function. It is often used to delete unnecessary features to find the function correlation. The method of correlation is used to find the relationship between the features. The correlation coefficient may not produce the output of the outcome. Thus, the use of the correlation coefficient is statistically important to analyze whether the relationship between the characteristics.

The selection algorithm of several feature subsets introduces mechanisms to carry out the task of defining and eliminating irrelevant and redundant features. Some function subset selection algorithms, however, can effectively delete the irrelevant features, but when there is high data dimensionality, they struggle to handle redundant features [4]. The relevance of the feature is poor. This constraint is a key issue in the choice of feature subsets from a large number of datasets. To solve the issue of redundant features in high-dimensional data, this paper describes the ranking based feature selection approach using correlation and distance measures for high dimensional data classification.

The proposed method operates in two steps; the first step focuses primarily on obtaining the attributes related to the target class by eliminating irrelevant attributes, and the second step focuses on removing redundant features from relative attributes by selecting representatives from different attribute subsets. The attribute subset was finally created using maximum accuracy rate with various classifiers.

The rest of the paper is as follows, Section 2 describes the literature survey done by researchers on the field of high dimensional classification. Section 3 explains the operational work of proposed approach, and Section 4 compares the proposed work with standard existing algorithms. Section 5 comprises the results and discussions. The conclusion and future work are then continued by Section 6.

## II. Related Works

In this section, literature surveys on attribute selection for high dimensional classification are appraised. Many researchers use different attribute selection methods to select the most important attributes from the original dataset to increase the accuracy rate [5, 6].

In 2015, the author [7] investigated on a comparison of filter methods to predict if patients who feel hepatitis diseases are more alive in future when this disease they were felt to rather than other diseases. In particular, this article revealed that the performance of learning algorithms is proficiently enhanced by using filter based feature selection methods. Nevertheless, no single filter-based approach is the greatest for selecting features in the real-world application. This thorough article describes the methods used in the study are able to enhancing learning algorithms efficiency to predict the threats of hepatitis disease. The author also details the deep comparison results of standard filter based feature selection methods for machine learning trainings. The author found that Naïve Bayes and Decision Table classifiers among four learning algorithms are more accurate on the hepatitis dataset through analyzing the six filter feature selection methods than the others.

The fast-ensembles of minimum redundancy feature selection with two variants is proposed by M. Savic and authors [8]. The first is Inner Ensemble algorithm that uses the linear correlation for each feature without computation of the mean values of each feature beforehand. The second is Fast Ensemble produces an ensemble of the entire selection process, instead of stabilizing each single correlation representing the strongest correlations between features to decrease high dimension. The method was experimented on nine datasets with high dimension. Two new algorithms are more robust than the existing maximum-relevancy-minimum-redundancy or correlation feature selection approach. Also they are much faster than a standard ensemble of these approaches.

Song et al. [9] have shown that by using hybrid feature selection, the accuracy of detecting a suspicious behavior pattern with unimportant features can be improved. Two stages feature selection using both filter (chi-square statistics) and wrapper manner with RF algorithm proposed for intrusion detection. The idea is to detect the best attribute subsets that are attribute importance by numerical values from original attribute set. The experiments were lead to make comparison with three sophisticated detection models and six classifiers on KDD Cup 99 dataset. The best detection accuracy was acquired by mixing the proposed filter with a wrapper approach, up to 12.38 % improved, following by an wrapper to retain the optimized set of features.

Using the strengths of univariate filtering methods, correlation matrix (CM) and principal component Analysis (PCA) techniques, Yasset, et al [10] developed the attribute selection workflow for high-dimensional omics data. The number of features was reduced by 80 percent without losing the accuracy of the proposed model when univariate filtering was used for well-known learning issues: classification and regression. The processor time of the learning model was reduced by eliminating redundant strongly relevant variables. To evaluate the proposed workflow, seven different proteomics and gene expression datasets were used in this work. The performance of proposed workflow compared with the results obtained by Li et al [11]. The authors noted that the proposed feature selection workflow performed better than maximum relevance-maximum distance for all datasets by selecting less than 10% of the features in all cases, with a decrease in compute time of more than 80%.

The authors described the feature selection based approach for high dimensional Genomic Microarray Data. This research aimed to classify genes with relevant biological relationships to the classification problem. Markov blanket filtering was utilized for deciding on particular subsets of features for each cardinal subset. Evaluation of performance measure is done depend on three classification algorithms such as Gaussian, logistic regression and nearest neighbor classifier in this paper. This study found that specific selection of features yields better performing classifiers than regularization methods and performed substantially better in the selected feature set than in the full feature set [12].

In [13], a comparison of six feature ranking methods implemented on two real datasets was proposed by the authors. Their presented approach can handle two categories: statistical based like Relief-F, One-R, Chi-Squared and entropy-based, such as Information Gain, Gain Ratio and Symmetrical Uncertainty. To build models, four learning algorithms namely, nearest neighbor classifier, Naive Bayes, C4.5 decision tree and the radial basis function network are adopted. This paper showed that for classification accuracy, the selection of ranking methods could be important. It is found that for different datasets and different classifiers, there is no better ranking index. As a function of the number of features used, precision curves may differ significantly. Classification accuracy is affected by the choice of rankings scores.

Ding (2005) proposed a minimum redundancy and maximum relevance feature selection framework based on mutual information as a measure of relevance of genes [14]. The proposed method is predicted on various classifiers like Naïve Bayesian (NB), Logistic Regression (LR), Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) using leave-one-out cross validation accuracy. The experiment results showed that this method with NB for multi-class dataset is better than SVM, LDA and LR and clearly and consistently demonstrated that only on maximum relevance, the baseline attribute sets is outperformed by the proposed approach.

In [15], Sigmis based on correlation approach focused on the handling of continuous features and missing data is proposed. This algorithm uses t-test as relevancy measure for choosing the most important features from the original data set. Finite Differences measure is used to find out and remove the missing values. For experimental analysis, three datasets from UCI machine learning repository are used by evaluating the performance with three algorithms namely CFS, FCBF and Consistency evaluator. There are also some excellent filter methods incorporating the class label information into learning methods.

Moreover, some variable assessment method, such as ReliefF (1992), Chi squared (1995), Correlation Feature Selection (CFS, 1999) and Principal Component Analysis (PCA, 2004), have also been provided on the Weka software framework to support successful data classification (2009).

These methods employ various search techniques for feature ranking, such as BestFirst, Exhaustive, Greedy- Stepwise, Random, and Ranker Search [16].

## III. Procedure for Feature Ranking Approach

In this portion, the process of ranking approach for high dimensional data classification is discussed. This approach contains two aspects: the first is the correlation between attribute and target class and the second is the redundancy between attribute and attribute. In this study, Pearson's coefficient is utilized to calculate the correlation and Euclidean distance is exploited to measure the redundancy.

### A. Pearson's Correlation Coefficient (PCC)

PCC is used to calculate linear dependency between two continuous X and Y variables as an indicator. The value differs between -1 and +1. Pearson's correlation between two variables (attribute and class) is given as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where, $\bar{x}$ and $\bar{y}$ are the mean of the x values and y values. The formulas return a value from -1 to 1. One indicates a strong positive relationship. Minus one indicates a strong negative relationship and zero indicates no relationship. The larger the value is, the stronger the relationship. The greater the PCC value, the higher the significance between attribute A and target class C.

### B. Euclidean Distance (ED)

ED is calculated to remove the redundant features. It is the most common use of distance and also known as simply distance. This measure is the best proximity measure when data is dense or continuous. The Euclidean distance between two data points in a plane is easily calculated by using Equation (2).

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2)$$

where, n is the number of instances. $x_i$ and $y_i$ are the two feature vectors of point x and y. The greater the distance, the more independence there is. The redundancy of the attribute subset becomes lower with the increase in the distance of the attributes.

In this study, the rank-based dimensional reduction approach using PCC and ED is presented. In the rank-based approach, to choice the best selected attribute subset from the original dataset, the learning model is applied. Algorithm 1 describes the procedure of the ranking-based approach for feature selection.

The selected attribute subset by presented approach has strong relevance with target class and minimal redundancy. Besides, it makes time consuming to task without preprocess such as discretization [17]. Although the input dataset is large, too much time and low accuracy prospects are costly for the FR approach. By a large scale, it can cut dimension but not decrease considerable accuracy rate.

**Algorithm 1: Feature Ranking (FR)**

***Input***: Dataset $D = \{A, c\}$ with a set of attributes A = $\{a_1, a_2 \ldots a_n\}$ and the target class $c$.

***Output***: Best selected attributes set $A_{selectedF}$

1. Calculate PCC values for each attribute $a_i$ and class label $c$ using $r = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(c_i - \bar{c})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2 \sum_{i=1}^{n}(c_i - \bar{c})^2}}$.
2. Calculate ED values for every attribute $a_i, a_j \in D, i \neq j$ using $d(a_i, a_j) = \sqrt{\sum_{i=1,j=1}^{n}(a_i - a_j)^2}$.
3. Add the PCC values and ED values for each attribute $a_i$ as ranking values.
4. Calculate the maximum value and then each ranking value is divided by this maximum value to scale the data.
5. Sort the normalized scores in descending order.
6. For each attribute $a_i$, do the following:
   a) Randomly split the training and testing data using ten-fold cross validation.
   b) Evaluate the accuracy rate of the ten-fold cross validation with various classifiers for each fold.
7. Determine the best attribute subset with the highest accuracy rate.
8. To construct $A_{selectedF}$, keep above selected attributes set with ranking scores.

## IV. Analysis and Experiments

This section discusses the primary portion of the presented attribute selection approach for high dimensional classification. It consists of four steps: benchmark datasets description, classifiers, performance measures, results and discussion. The details of the steps will be presented in the following subsections:

### A. Benchmark Datasets

In experimental study, five popular benchmark datasets in different area, namely Musk2, Madelon, Movement, Arrhythmia and Isolet5 are employed from UCI machine learning repository [18]. Different datasets from this repository with a fusion of attributes such as similar and different area, small and large number of categories, etc., are utilized to perform a dissimilar and rational evaluation. The datasets were pre-processed to fill missing values and converted nominal to numeric. The briefly description of datasets are defined in Table I. Java programming is used for algorithm in this experiment. In experiments, Intel Core i5 is the CPU of the machine, clocked at 2.50GHz, 4 GB memory and 64-bit Window 10 operating system.

TABLE I. Briefly Description for Benchmark Datasets

| No | Dataset | No. of Samples | No. of Attributes | Class | Area |
|---|---|---|---|---|---|
| 1 | Musk2 | 6,598 | 168 | 2 | Physical |
| 2 | Madelon | 2600 | 500 | 2 | Medical |
| 3 | Movement | 360 | 91 | 15 | Image |
| 4 | Arrhythmia | 452 | 279 | 16 | Medical |
| 5 | Isolet5 | 1559 | 617 | 26 | Letter Recognition |

### B. Classifiers

In this paper, four popularly classifiers including Naïve Bayesian (NB), k-Nearest Neighbors (kNN), C4.5 Decision Tree and Random Forest (RF) classifiers provided by the WEKA software [19] is used to prevent any unfairness to a specific attribute selection process implemented by a classifier.

***Naive Bayes*** is one of the most efficient learning algorithms for data mining. It is based on the assumption of conditional independence between classes by the Bayes rule. Using the joint probabilities of sample observations and classes, the algorithm tries to estimate the conditional probabilities of classes given an observation.

***k- Nearest Neighbors*** is one of the simplest classifiers that are based on finding distances between vectors. It classifies the samples in the training data based on their nearest k-neighbors. The category of the test vector is then decided using a majority voting. Euclidean distance measure is used to give good results for classification. In this paper, kNN classifier with k = 1 to verify the classification accuracy of the five datasets.

***Decision Tree***: The C4.5 classification method was selected because it is easy to interpret due to the white box method. This method's performance is a decision tree that is easily understood by the expert who validates the results obtained.

***Random Forest*** (RF, numTrees = 10) starts with constructing multiple trees. By choosing and using a random number of attributes, the diversity of the trees is increased to create the nodes and leaves of a random tree classifier. Using a different bootstrap sample, each tree is built. In this study, the random trees in building the RF classifier are used with no pruning from a computational perspective [20].

To evaluate the classification performance, ten-fold cross validation was adopted to obtain an unbiased experimental result in the experiments. Each dataset has been stratified into ten-fold. Nine folds are used as training sample and the remaining one fold is used as a test sample.

### C. Performance Measures

Accuracy measure is mainly employed as evaluation index to analyze the results. To carry out the experiments, various learning algorithms with ten-fold cross validation are performed to estimate the accuracy of selected attribute set for five different datasets. To demonstrate the efficiency of the proposed approach for these datasets, accuracy is evaluated as follow:

$$Accuracy\ (\%) = \frac{TP+TN}{TP+FP+FN+TN} \qquad (3)$$

where True Positive (TP) is present outcome correctly identified as present. True Negative (TN) is absent outcome correctly identified as absent. False Positive (FP) is a present outcome identified as absent. False Negative (FN) is an absent outcome identified as present.

### D. Results and Discussion

The experiment is mainly divided into three portions: the first compares the number of attributes selected by different learning algorithms, the second and third compares the classification results and running time on the effect of the selected attributes under different datasets. The comparisons on the presented approach with various classifiers are shown in the following Table II, III, IV, Figures 1 and 2 respectively, in terms of number of selected attributes, classification accuracy, running time on number of selected attributes and total attributes. Note that "All" represents the all attributes of the datasets. The presented FR approach select the attributes with the highest accuracy rate as the best attribute subset in order to ensure the comparability of the number of attributes with various classifiers.

*a) Comparison on number of selected attributes:* Table II shows the comparison of the number of selected attributes with various classifiers based on presented algorithm FR. For Madelon dataset, number of selected attributes evaluated by all classifiers are the same. Generally, FR outperforms random forest classifier not only in the accuracy rate but also its number of selected attributes is reduced by other classifiers on Arrhythmia and Isolet5 datasets. Moreover, it can also be seen that the selected attributes by random forest classifier are the fewest on these datasets. Compared with the C4.5, NB, kNN classifiers, FR with random forest classifier has the number of selected attributes with obvious advantage of accuracy rate on two datasets. By comparison, it can be concluded that the FR approach generating the attribute subset with various classifiers will not only make the selected attribute set growth the classification accuracy but also search the best attributes set with more rapidly convergence speed.

TABLE II. COMPARISON OF NO. OF SELECTED ATTRIBUTES WITH VARIOUS CLASSIFIERS ON DIFFERENT DATASETS

| Datasets | All | NB | C4.5 | kNN | RF |
|---|---|---|---|---|---|
| Madelon | 500 | 13 | 13 | 13 | 13 |
| Musk2 | 168 | **139** | 162 | 145 | 156 |
| Movement | 91 | 88 | 87 | **67** | 85 |
| Arrhythmia | 279 | 70 | 164 | 201 | **44** |
| Isolet5 | 617 | 590 | 427 | 523 | **223** |

*b) Comparison on classification results under the effect of selected and total attributes:* The classification accuracy of five benchmark datasets by four learning algorithms based on presented approach under the effect of selected attributes and total attributes is shown in Figure 1 and 2.

In Table III, the comparison of accuracies on all attributes and selected attributes before and after using proposed FR algorithm evaluated by various classifiers are described. The accuracy with kNN and RF classifiers are typically increased for all datasets. The accuracies with NB on Isolet5 dataset and C4.5 on Movement dataset are slightly decreased.

*c) Comparison on running time under the effect of selected and total attributes:* The running time comparison with four classifiers based on FR approach under the effect of selected attributes and total attributes for all datasets is illustrated in Table IV.

TABLE III. ACCURACIES OF VARIOUS CLASSIFIERS BEFORE AND AFTER USING FR ALGORITHM FOR DIFFERENT DATASETS

| Datasets | NB | | C4.5 | | kNN | | RF | |
|---|---|---|---|---|---|---|---|---|
| | Before | After | Before | After | Before | After | Before | After |
| Madelon | 59.54 | 87.77 | 69.04 | 81.69 | 54.27 | 87.77 | 64.27 | 88.27 |
| Musk2 | 83.86 | 84.62 | 96.88 | 97.35 | 95.80 | 95.98 | 90.76 | 91.18 |
| Movement | 62.78 | 63.33 | 69.72 | 69.44 | 85.83 | 87.78 | 83.61 | 85.00 |
| Arrhythmia | 61.73 | 63.72 | 64.60 | 67.04 | 53.54 | 55.53 | 67.04 | 71.46 |
| Isolet5 | 83.77 | 83.58 | 78.00 | 78.58 | 85.57 | 86.47 | 92.30 | 93.20 |

TABLE IV. COMPARISION WITH VARIBOUS CLASSIFIERS ON DIFFERENT DATASETS BASED ON RUNNING TIME IN SECONDS

| Datasets | NB | | C4.5 | | kNN | | RF | |
|---|---|---|---|---|---|---|---|---|
| | All | Selected | All | Selected | All | Selected | All | Selected |
| Madelon | 0.42 | 0.02 | 2.75 | 0.42 | 0.00 | 0.01 | 3.93 | 1.38 |
| Musk2 | 0.17 | 0.14 | 2.09 | 2.11 | 0.00 | 0.00 | 0.25 | 0.13 |
| Movement | 0.02 | 0.01 | 0.08 | 0.05 | 0.00 | 0.00 | 0.38 | 0.18 |
| Arrhythmia | 0.08 | 0.05 | 0.56 | 0.14 | 0.02 | 0.00 | 1.03 | 0.27 |
| Isolet5 | 0.45 | 0.17 | 2.88 | 1.70 | 0.00 | 0.00 | 3.69 | 2.11 |

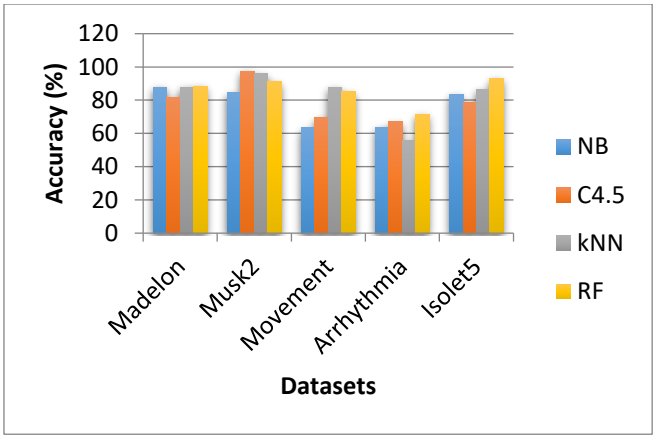


Fig. 1. Performance comparison of various classifiers on selected attributes for different datasets

Overall, for Madelon, Arrhythmia and Isolet5 dataset, the accuracy rates of FR approach with random forest classifier is located at the top tier and is superior to all other classifiers on selected attributes for datasets. On all attributes, both arrhythmia and Isolet5 datasets are high accuracy over other datasets. It can be seen that all the attributes are directly used to train classifiers leads to the performance of the classifier is very bad.

Additionally, the following implications can be drawn.

- Compared with using all attributes, the classification accuracy is considerably increased for all datasets. In addition, the computational overhead can be greatly condensed due to the attribute dimension is reduced.
- From the point of classifier performance view, the presented FR approach has a significant upgrading in classification accuracy compared with various classifiers on selected attributes set for all different datasets.
- Though the various ranking method in literature does not target at the specific classifier when selecting the best attribute subset, the FR approach consider the accuracy rate with cross validation. Therefore, the stability of the feature selection is ensured for high dimensional datasets.

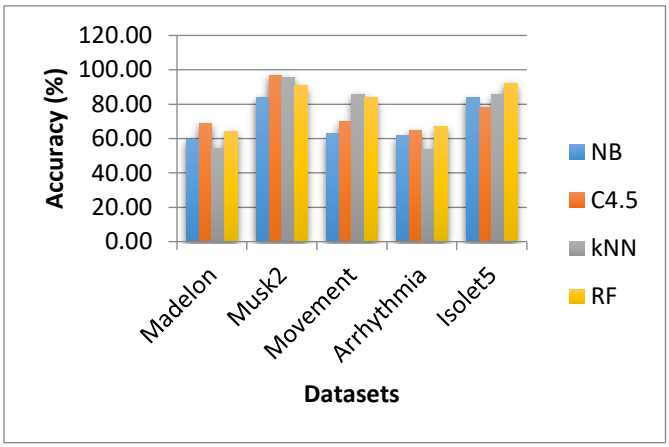


Fig. 2. Performance comparison of various classifiers on total attributes for different datasets

According to the experiments, it can be established that the classification accuracy of the attribute selected by the ranking based selection algorithm presented in this paper outperforms on the five benchmark datasets using four well-known classifiers. The accuracies with random forest algorithm on selected attributes are obviously higher than the other three classifiers when classifying the different data. The proposed ranking based FR with random forest algorithm can reduce more than 50% of attributes in three datasets. Based on the relatively small number of best selected attributes, the FR algorithm achieves the highest classification accuracy.

## V. Conclusion

This paper presented a dimensional reduction algorithm based on ranking method to select the effective attribute subsets in high dimensional datasets. The algorithm first uses PCC to rapidly decrease the dimension, and then eliminates the redundant features using Euclidean distance measure with larger feature distance. The various classifiers are applied as the classification algorithm to select the attributes with the highest accuracy rate in this approach. Experiments on five benchmark datasets from UCI machine learning repository show that the ranking based reduction algorithm achieves better classification results with smaller attributes compared to total attributes evaluated on different classifiers. The presented approach eliminates the workload to select the method for discretization and preprocessing. The results of the experiment indicate that the presented approach has advantage on large-scale datasets, has low expectation of accuracy and cares about the classification efficiency.

As a future direction, the presented ranking approach is combined with wrapper method for other high dimensional data classification problems. In addition, different types of distance measurements instead of Euclidean distance can be used to calculate the redundancy between two features. It is possible to apply this approach to very high dimensional and small samples dataset especially microarray gene expression datasets.

## References


[1] P. R. Anukrishna and V. Paul, "A Review on Feature Selection for High Dimensional Data," in *International Conference on Inventive Systems and Control (ICISC-2017)*, Coimbatore, India, Jan. 2017, pp. 1–4, doi: 10.1109/ICISC.2017.8068746.

[2] T. Saw and P. H. Myint, "Feature Selection to Classify Healthcare Data using Wrapper Method with PSO Search," *IJITCS*, vol. 11, no. 9, pp. 31–37, Sep. 2019, doi: 10.5815/ijitcs.2019.09.04.

[3] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," in *Proceedings of the Twentieth International Conference on Machine Learning*, Washington DC, 2003, pp. 856–863.

[4] M. S. Kumbhar and S. Patil, "An Efficient & Effective Feature Subset Selection for High Dimensional Data," in *International Conference on Trends in Electronics and Informatics (ICEI)*, Tirunelveli, India, May 2017, pp. 486–491, doi: 10.1109/ICOEI.2017.8300975.

[5] B. Senliol, G. Gulgezen, L. Yu, and Z. Cataltepe, "Fast Correlation Based Filter (FCBF) with a different search strategy," in *2008 23rd International Symposium on Computer and Information Sciences*, Istanbul, Turkey, Oct. 2008, pp. 1–4, doi: 10.1109/ISCIS.2008.4717949.

[6] P. Pramokchon and P. Piamsa-nga, "Effective threshold estimation for filter-based feature selection," in *2016 International Computer Science and Engineering Conference (ICSEC)*, Chiang Mai, Thailand, Dec. 2016, pp. 1–6, doi: 10.1109/ICSEC.2016.7859919.

[7] P. Yildirim, "Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease," *IJMLC*, vol. 5, no. 4, pp. 258–263, Aug. 2015, doi: 10.7763/IJMLC.2015.V5.517.

[8] B. Schowe and K. Morik, "Fast-Ensembles of Minimum Redundancy Feature Selection," in *Ensembles in Machine Learning Applications*, vol. 373, O. Okun, G. Valentini, and M. Re, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 75–95.

[9] J. Song, W. Zhao, Q. Liu, and X. Wang, "Hybrid feature selection for supporting lightweight intrusion detection systems," *J. Phys.: Conf. Ser.*, vol. 887, p. 012031, Aug. 2017, doi: 10.1088/1742-6596/887/1/012031.

[10] Y. Perez-Riverol, M. Kuhn, J. A. Vizcaíno, M.-P. Hitz, and E. Audain, "Accurate and fast feature selection workflow for high-dimensional omics data," *PLoS ONE*, vol. 12, no. 12, p. e0189875, Dec. 2017, doi: 10.1371/journal.pone.0189875.

[11] S. Li and S. Oh, "Improving feature selection performance using pairwise pre-evaluation," *BMC Bioinformatics*, vol. 17, no. 1, p. 312, Dec. 2016, doi: 10.1186/s12859-016-1178-3.

[12] E. P. Xing, M. I. Jordan, and R. M. Karp, "Feature Selection for High-Dimensional Genomic Microarray Data," vol. 1, pp. 601–608, 2001.

[13] J. Novakovic, P. Strbac, and D. Bulatovic, "Toward optimal feature selection using ranking methods and classification algorithms," *Yugosl J Oper Rres*, vol. 21, no. 1, pp. 119–135, 2011, doi: 10.2298/YJOR1101119N.

[14] C. Ding and H. Peng, "Minimum Redundancy Feature Selection from Microarray Gene Expression Data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[15] E. C. Blessie and E. Karthikeyan, "Sigmis: A Feature Selection Algorithm Using Correlation Based Method," *Journal of Algorithms & Computational Technology*, vol. 6, no. 3, pp. 385–394, Sep. 2012, doi: 10.1260/1748-3018.6.3.385.

[16] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "MIFS-ND: A mutual information-based feature selection method," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6371–6385, Oct. 2014, doi: 10.1016/j.eswa.2014.04.019.

[17] 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, Jan. 2016, doi: 10.1016/j.neucom.2014.12.123.

[18] D. Dua and C. Graff, "UCI Machine Learning Repository," *University of California, Irvine, School of Information and Computer Sciences*, 2017, [Online]. Available: http://archive.ics.uci.edu/ml.

[19] E. Frank, M. A. Hall, and I. H. Witten, *The WEKA workbench*. Morgan Kaufmann, 2016.

[20] L. BREIMAN, "Random Forests," *Machine Learning, Kluwer Academic Publishers. Manufactured in The Netherlands*, vol. 45, no. 1, pp. 5–32, Apr. 2001.