

# Literature Review on Feature Selection Methods for High-Dimensional Data

D. Asir Antony Gnana  
Singh

Anna University, BIT Campus,  
Tiruchirappalli, India.

S. Appavu alias  
Balamurugan

K.L.N College of Information  
Technology, Sivagangai, India.

E. Jebamalar Leavline  
Anna University, BIT Campus,  
Tiruchirappalli, India.

## ABSTRACT

Feature selection plays a significant role in improving the performance of the machine learning algorithms in terms of reducing the time to build the learning model and increasing the accuracy in the learning process. Therefore, the researchers pay more attention on the feature selection to enhance the performance of the machine learning algorithms. Identifying the suitable feature selection method is very essential for a given machine learning task with high-dimensional data. Hence, it is required to conduct the study on the various feature selection methods for the research community especially dedicated to develop the suitable feature selection method for enhancing the performance of the machine learning tasks on high-dimensional data. In order to fulfill this objective, this paper devotes the complete literature review on the various feature selection methods for high-dimensional data.

## General Terms

Literature review on feature selection methods, study on feature selection, wrapper-based feature selection, embedded-based feature selection, hybrid feature selection, filter-based feature selection, feature subset-based feature selection, feature ranking-based feature selection, attribute selection, dimensionality reduction, variable selection, survey on feature selection, feature selection for high-dimensional data, introduction to variable and feature selection, feature selection for classification.

## Keywords

Introduction to variable and feature selection, information gain-based feature selection, gain ratio-based feature selection, symmetric uncertainty-based feature selection, subset-based feature selection, ranking-based feature selection, wrapper-based feature selection, embedded-based feature selection, filter-based feature selection, hybrid feature selection, selecting feature from high-dimensional data.

## 1. INTRODUCTION

In the digital era, handling the massive data is a challenging task among the researchers since the data are accumulated through various data acquisition techniques, methods, and devices. These accumulated massive raw data degrade the performance of the machine learning algorithms in terms of causing overfitting, spending more time to develop the machine learning modes and degrading their accuracy since the raw data are noisy in nature and have more number of features known as high-dimensional data. In general, the high-dimensional data contains irrelevant and the redundant features. The irrelevant features cannot involve in the learning process and the redundant features contain same information hence they miss lead the learning process. Therefore, these issues can be tackled by the feature selection. The feature

selection is a process of removing the redundant and the irrelevant features from a dataset to improve the performance of the machine learning algorithms. The feature selection is also known as variable selection or attribute selection. The features are also known as variables or attributes. The machine learning algorithms can be roughly classified into two categories one is supervised learning algorithm and another one is unsupervised learning algorithm. The supervised learning algorithms learn the labeled data and construct learning models that are known as classifiers. The classifiers are employed for classification or prediction to identify or predict the class-label of the unlabeled data. The unsupervised learning algorithms learn the unlabeled data and construct the learning models that known as clustering models. The clustering models are employed to cluster or categorize the given data for predicting or identifying their group or cluster. Mostly, the feature selections are employed for the supervised learning algorithms since they suffered by the high-dimensional space. Therefore, this paper presents a complete literature review on various feature selection methods for high-dimensional data.

The rest of this paper is organized as follows: Section 2 describe the feature selection process. In Section 3, survey on feature selection is conducted. Section 4 summarizes the survey on feature section. Section 5 concludes this paper.

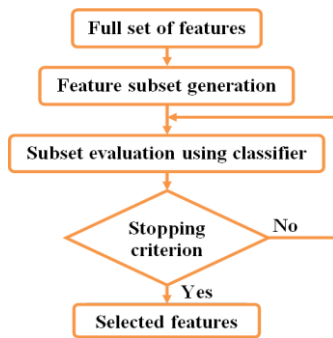
## 2. FEATURE SELECTION

Feature selection is a process of removing the irrelevant and redundant features from a dataset in order to improve the performance of the machine learning algorithms in terms of accuracy and time to build the model. The process of feature selection is classified into two categories namely feature subset selection and feature ranking methods based on how the features are combined for evaluation. The feature subset selection approach generates the possible number of combinations of the feature subsets using any one of the searching strategies such as a greedy forward selection, greedy backward elimination, etc. to evaluate the individual feature subset with a feature selection metric such as correlation, consistency, etc. In this method, space and the computational complexity involved are more due to the subset generation and evaluation [2].

In feature ranking method, each feature is ranked by a selection metric such as information gain, symmetric uncertainty, gain ratio, etc. and the top ranked features are selected as relevant features by a pre-defined threshold value. This approach is computationally cheaper and space complexity is less compared to subset approach. However, it does not deal with redundant features.

Further, the process of feature selection is classified into four categories namely wrapper, embedded, filter, and hybrid

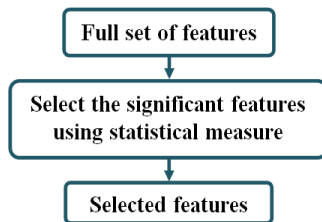
methods based on how the supervised learning algorithm is employed in the feature selection process.



**Figure 1 Feature selections with wrapper approach**

Wrapper approach incorporates the supervised learning algorithm for validating the generated feature subsets using any one of the searching strategies as shown in Figure 1. It yields high classification accuracy only for the particular learning algorithm adopted. Hence, it does not possess a high generality and the computational complexity is higher than embedded and filter methods.

The embedded approach uses a part of supervised learning algorithm for feature selection process and it produces better accuracy only for the learning algorithm used in the selection process. Hence, it does not have a high generality and it is computationally expensive than the filter and lesser than the wrapper method.



**Figure 2 Feature selection with filter approach**

The filter approach (Figure 2) selects the features without the influence of any supervised learning algorithm. Hence, it works for any classification algorithm and achieves more generality with less computational complexity than the wrapper and embedded methods. Therefore, it is suitable for high-dimensional space. The combination of wrapper and filter approach is known as hybrid method [1].

### 3. SURVEY ON FEATURE SELECTION

As the feature selection is employed in various machine learning applications, it has remarkable literature records made by the research community. Feature selection is a preprocessing technique to select the significant features from a dataset by removing the irrelevant and redundant features for improving the performance of the machine learning algorithms. The feature selection process can be categorized into various methods based on how the features are combined for evaluation in the feature selection process and how the supervised learning algorithm is used to evaluate the features in the features selection process. This paper reviews the literature on various features selection methods and explores their merits and demerits.

### 3.1 Feature Selection Based on Combining the Features for Evaluation

This section reviews various methods of feature selection based on how the features are combined for evaluation in order to select the significant features from a dataset. They are classified into feature subset-based and feature ranking-based methods.

#### 3.1.1 Feature subset-based methods

In the feature subset-based method, the features are combined as possible combinations of feature subsets using any one of the searching strategies. Then, the feature subsets are evaluated using any one of the statistical measures or the supervised learning algorithms to observe the significance of each subset and the most significant subset is selected as the significant feature subset for a given dataset. If the subset is evaluated using the supervised learning algorithm, then this method is known as wrapper method.

The best example for the feature subset-based method is correlation-based feature subset selection (CRFS) developed by Hall [3]. In this approach, two correlation measures are considered; one is feature-class correlation and another one is feature-feature correlation. Initially, N numbers of features are combined as possible combinations of feature subsets using heuristic-based best-first search, then each subset is evaluated with the two correlation measures as mentioned above. The subset that has lesser feature-feature correlation and higher feature-class correlation compared to other feature subsets is considered as the selected significant feature subset for the classification task. Liu & Setiono [4] proposed a feature subset-based feature selection method namely consistency-based feature subset selection (COFS). This method uses the class consistency as an evaluation metric in order to select the significant feature subset from the given dataset. These methods are the filter-based methods since they do not use the supervised learning algorithm to validate the subsets and they use the statistical measure for evaluating the feature subsets.

In general, the exhaustive or complete search has to generate  $2^N$  number of subsets to produce the maximum number of possible combinations of feature subsets from the N number of features for evaluation. Therefore, this exhaustive searching strategy is computationally quite expensive hence the heuristic searching strategies such as simulated annealing (SA), tabu searching (TS), ant colony optimization (ACO), genetic algorithm (GA), particle swarm optimization (PSO), etc. [5] are used by some of the researchers to get the optimal solution by generating less number of feature subsets for evaluation. In the heuristic searching, the heuristic function obtains the prior knowledge to guide the search process to generate the subsets and these subsets are evaluated using supervised machine learning algorithm. These factors make the feature subset-based methods computationally expensive and also these methods seem to be the wrapper approach.

Some researchers used the simulated annealing search for generating the feature subset for evaluations. For example, Lin et al used the simulated annealing search to generate the feature subsets and evaluated them by supervised learning algorithm namely back-propagation network (BPN) to choose the better feature subset [6]. Meiri & Zahavi used simulated annealing-based feature selection for marketing application [7]. In several feature selection methods, the tabu search is used for subset generation such as Zhang & Sun developed a tabu search-based feature selection. In this method, the subsets generated by tabu search are evaluated using the classification error criteria to find the better feature subset [8].

Tahir et al formed the feature subsets using tabu search then these subsets are evaluated using K-nearest neighbor classifier (kNN) with the classification error as evaluation criteria to obtain the significant feature subset [9].

A number of feature selection processes used the ant colony optimization as the searching criteria for subset generation. Aghdam et al employed the ant colony optimization search to form the feature subsets and they are validated by the nearest neighbor classifier for text classification application [10]. Kanan & Faez proposed a feature selection method using ant colony optimization for face recognition system. In this approach, the nearest neighbor classifier is adopted for evaluating the generated subset using ant colony optimization-based learning [11]. Sivagaminathan & Ramakrishnan developed an ant colony optimization-based feature selection with artificial neural networks (ANN) for medical diagnosis system. In this method, the generated feature subsets are validated using ANN [12]. Sreeja & Sankar presented an ant colony optimization-based feature selection with instance-based pattern matching-based classification (PMC) [13].

In certain feature selection research works, the genetic algorithm is adopted to generate the feature subsets for evaluation and the supervised machine learning algorithm is used to evaluate the generated subsets. Welikala et al presented a feature selection using genetic algorithm with support vector machine (SVM) for mining the medical dataset [14]. Erguzel et al used the genetic algorithm and artificial neural network for electroencephalogram (EEG) signal classification [15]. Oreski & Oreski proposed a feature selection method based on genetic algorithm with neural networks for credit risk assessment [16]. Li et al developed a genetic algorithm with support vector machine for hyper-spectral image classification [17]. Das et al formulated a genetic algorithm with support vector machine-based feature selection for handwritten digit recognition application [18]. Wang et al applied the genetic algorithm for subset generation with support vector machine in feature selection process for data classification applications [19].

In the literature, some researches employed the particle swarm optimization to generate the feature subsets and to validate them by supervised machine learning algorithm to identify the significant feature subset. Xue et al designed a particle swarm optimization (PSO)-based feature selection for classification. In this method, the feature subsets generated by PSO are evaluated using supervised learning algorithm [20]. Chen et al presented a feature selection method using particle swarm optimization search for sleep disorder diagnosis system [21]. Yang et al developed a particle swarm optimization-based feature selection for land cover classification [22].

From the subset-based feature selection literature, it is observed that the exhaustive or complete search leads to high computational complexity as it generates  $2^N$  number of subsets from N number of features for evaluation. This searching strategy cannot be a better choice for high-dimensional space. The heuristic search methods also lead to more computational complexity, because they need prior knowledge and each generated subset needs to develop a classification model for evaluating them to obtain the optimal feature subset in an iterative manner, hence these searching strategies are not suitable for high-dimensional space. However, these heuristic search methods follow a wrapper-based approach. Therefore, these methods are computationally expensive and they can only produce higher classification accuracy for the specific classification algorithm used to validate the subset, so they cannot achieve high generality.

### *3.1.2 Feature ranking-based methods*

In the feature-ranking based approach, each feature of a dataset is weighted based on any one of the statistical or information-theoretic measures and the features are ranked based on their weight. Then the higher ranked features are selected as the significant features using a predefined threshold that determines the number of features to be selected from a dataset. The best example for the feature ranking-based method is chi-square-based feature selection (CQFS). In this method, Liu & Setiono used the chi-square statistic measure to weight the features in order to rank them for selecting the significant features [24]. In the similar way, the information-theoretic measures such as information gain, symmetric uncertainty, gain ratio, etc. are employed to weight the individual feature and rank them for selection.

Further, it is observed that the feature ranking-based methods use the statistical measures or information-theoretic measures to weight the individual feature only by observing the relevancy between the individual feature and the target-class. Hence, these methods take less runtime but fail to remove the redundant features [2]. The feature ranking-based methods follow a filter-based approach since these methods do not involve the supervised learning algorithm to evaluate the significance of the features. Consequently, these methods are independent of the supervised learning algorithm hence they achieve more generality and less computational complexity. Thus, the feature ranking-based methods can be a good choice for selecting the significant features from the high-dimensional space with suitable redundancy analysis mechanism.

## **3.2 Feature Selection Based on the Supervised Learning Algorithm Used**

This section reviews various methods of feature selection based on the machine learning algorithm used. They are categorized as wrapper, embedded, filter, and hybrid methods.

### *3.2.1 Wrapper-based methods*

Wrapper-based approach generates the feature subsets using any one of the searching techniques and evaluates these subsets using the supervised learning algorithm in terms of classification error or accuracy [25]. The wrapper method seems to be a “brute force” method. This approach is illustrated in Figure 1. Kohavi & John developed a wrapper-based feature selection method for selecting the significant features from the dataset [26]. This method consists of search engine for subset generation and classification algorithm to evaluate the subset. Further, they compare the performance of this method in terms of classification accuracy with hill-climbing and best-first searching strategies using decision tree and naïve Bayes classifiers. However, they observed that wrapper method has the problems such as searching overhead, overfitting, and increased runtime.

In wrapper approach, the searching is an overhead since the searching technique does not have the domain knowledge. In order to overcome the searching time overhead, Inza et al used estimation of Bayesian network algorithm for feature subset selection using naïve Bayes and ID3 (Iterative Dichotomiser 3) [27]. In general, the searching method may lead to increase in computational complexity, since the training data is split for evaluation. In order to overcome this issue, Grimaldi et al used an aggregation principle with sequential search [28]. Dy & Brodley developed a wrapper-based approach for unsupervised learning using order identification (recognizing the number of clusters in the data) with the expectation maximization (EM) clustering algorithm

using maximum likelihood (ML) criterion [29]. Aha & Bankert presented a wrapper-based method with beam search and IB1 classifier [30]. Also, they compared its performance with the well known sequential search algorithms for feature selection such as forward sequential selection (FSS) and backward sequential selection (BSS). They observed that the beam search outperforms the FSS and BSS.

The Maldonado & Weber developed a wrapper approach-based feature selection by combining support vector machine (SVM) with kernel functions. This method uses the sequential backward selection for feature subset generation and these subsets are validated in terms of classification error to identify the best subset [31]. In order to minimize the searching overhead, Gütlein et al used the search algorithm namely ORDERED-FS that orders the features in terms of resubstitution error to identify their irrelevancy [32]. Kabir et al developed a wrapper-based constructive approach for feature selection (CAFS) using neural network (NN). In this method, the correlation measure is used to remove the redundancy in the searching strategy for improving the performance of NN [33]. Stein et al proposed an ant colony optimization-based feature selection with wrapper model. In this approach, the ant colony optimization is used as a searching method in order to reduce the searching overhead such as blind search or forward selection or backward elimination searching methods [34]. Furthermore, to minimize the searching overhead, Zhuo et al presented a wrapper-based feature selection using genetic algorithm with support vector machine for classifying the hyper-spectral images [35].

In the wrapper approach, overfitting can be overcome by post-pruning, jitter, and early stopping methods. Post-pruning is carried out while developing the decision tree [36]. In jitter method, the noisy data that make the learning process more difficult are eliminated in order to fit the training data thereby the overfitting is eliminated [37]. In early stopping method, overfitting is eliminated using neural network by stopping the training process when performance on a validation set starts to deteriorate [38] [39]. The researchers have tried to reduce the overfitting by early stopping method using genetic algorithm-based searching with early stopping (GAWES) [40].

Further, it is observed that the wrapper-based methods are suffered by the searching overhead, overfitting [41] and have more computational complexity with less generality since they use the supervised learning algorithm for evaluating the generated subsets by the searching method. Therefore, these methods are not suitable choice for the high-dimensional space.

### *3.2.2 Embedded-based methods*

The embedded-based methods use a part of the learning process of the supervised learning algorithm for feature selection. Embedded-based methods reduce the computational cost than the wrapper method [42]. This embedded method can be roughly categorized into three namely pruning method, built-in mechanism, and regularization models. In the pruning-based method, initially all the features are taken into the training process for building the classification model and the features which have less correlation coefficient value are removed recursively using the support vector machine (SVM) [43]. In the built-in mechanism-based feature selection method, a part of the training phase of the C4.5 [36] and ID3 [44] supervised learning algorithms are used to select the features. In the regularization method, fitting errors are minimized using the objective functions and the features with near zero regression coefficients are eliminated [45] [46].

Neumann et al developed an embedded-based feature selection method for selecting the significant features from synthetic and real world datasets. In their approach, linear and non linear SVMs are employed in the selection process using the deference of convex functions algorithm (DCA) [47]. Xiao et al proposed an embedded-based method to select the significant features from audio signals for emotion classification. This method was implemented based on the principle of evidence theory with mass function and the identified most relevant features are added incrementally for classification [48]. Maldonado et al developed an embedded method to select the significant features from imbalanced data for classification with several objective functions [49].

Further, it is observed that the embedded methods are computationally efficient than the wrapper methods and computationally costlier than the filter methods hence they cannot be suitable choice for high-dimensional space and they have poor generality since the embedded methods use the supervised learning algorithm.

### *3.2.3 Filter-based methods*

The filter-based approaches are independent of the supervised learning algorithm therefore offer more generality and they are computationally cheaper than the wrapper and embedded approaches. For processing the high-dimensional data, the filter methods are suitable rather than the wrapper and embedded methods.

Generally, the process of feature selection aimed at choosing the relevant features. The best example is Relief [50] that was developed with the distance-based metric function that weights each feature based on their relevancy (correlation) with the target-class. However, Relief is ineffective as it can handle only the two-class problems and also does not deal with redundant features. The modified version of the Relief known as ReliefF [51] can handle the multi-class problems and deal with incomplete and noisy datasets too. However, it fails to remove the redundant features. Holte developed a rule-based attribute selection known as OneR which forms one rule for each feature and selects the rule with the smallest error [52]. Yang & Moody proposed a joint mutual information-based approach (JMI) for classification. It calculates the joint mutual information between the individual feature and the target-class to identify the relevant features, and a heuristic search is adopted for optimization when the number of features is more. The features containing similar information and lesser relevancy to the target-class are treated as redundant features that are to be eliminated [53].

Peng et al proposed a mutual information-based max-relevancy min-redundancy (MRMR) feature selection. To identify the feature relevancy, the mutual information is computed between the individual feature and target-class, and to identify the redundant feature, the mutually exclusive condition is applied [54]. Battiti developed a mutual information-based feature selection method (MIFS). In this method, mutual information measure is used to determine the relevancy between the individual feature and the target-class. The features having similar information are considered as redundant features that are to be removed [55]. Fleuret presented a feature selection scheme namely conditional mutual info maximization (CMIM) that recursively chooses the features that have maximum mutual information with the target-class for classification [56].

Meyer & Bontempi proposed a filter-based approach that uses double input symmetrical relevance (DISR) metric for feature selection. This approach returns the selected features that

contain more information about the target-class than the information about other features [57]. Lin & Tang introduced an information theory-based conditional infomax feature extraction (CIFE) algorithm to measure the class-relevancy and redundancy for feature selection [58]. Brown et al used the conditional redundancy (CondRed) metric for selecting the significant features from the dataset [59].

In the recent past, the clustering technique is also adopted in feature selection. Song et al developed a feature selection framework and adopted the graph-based clustering technique to identify the similarity among the features for removing the redundant features [60]. Dhillon et al developed a feature selection algorithm based on information theory for text classification. In this approach, the hierarchical clustering is used to cluster the features or terms of documents for identifying their dependencies [61]. Li et al incorporated the clustering algorithm with the chi-square statistical measure to select the features from statistical data [62]. Cai et al developed a spectral clustering-based feature selection (MCFS) for selecting the significant features from the datasets [63]. Chow & Huang employed the supervised clustering technique and mutual information for identifying the salient features from synthetic and real world datasets [64]. Mitra et al presented a feature selection approach by adopting the graph-based clustering approach to identify the similarity among the features for redundancy analysis [65]. Sotoca & Pla developed a feature selection method for classification based on feature similarity with hierarchical clustering [66].

Further, it is observed that the filter-based methods are computationally better than the wrapper [67] and embedded [68] methods. Therefore, the filter-based methods can be a suitable choice for high-dimensional space. The filter-based methods achieve high generality since they do not use the supervised learning algorithm.

### 3.2.3 Hybrid Methods

The hybrid methods are the combination of filter and wrapper-based approaches [69]. In general, processing the high-dimensional data is a difficult task with the wrapper method therefore the authors Bermejo et al developed a hybrid feature selection method known as filter-wrapper approach. In this approach, they used a statistical measure to rank the features based on their relevancy then the higher ranked features are given to the wrapper method so that the number of evaluations required for the wrapper method is linear. Thus, the computational complexity is reduced using hybrid method for medical data classification [70]. Ruiz et al developed a gene (feature) selection algorithm for selecting the significant genes for the medical diagnosis system. They used a statistical ranking approach to filter the features from high-dimensional space and the filtered features are fed into the wrapper approach. This combination of the filter and wrapper approach was used to distinguish the significant genes causing cancer disease in the diagnosis process [71].

Xie et al developed a hybrid approach for diagnosing the erythema-squamous diseases. In this approach, F-score measure is used to rank the features to identify the relevant features (filter approach). The significant features are selected from the ranked features with the sequential forward floating search (SFFS) and SVM (wrapper method) [72]. Kannan & Faez presented a hybrid feature selection framework. In this approach, ant colony optimization (ACO)-based local search (LS) is used with the symmetric uncertainty measure to rank the features [73]. Xie et al designed a hybrid approach with F-score to identify the relevant attributes from a disease

dataset. For feature subset generation from the relevant features, the searching strategies such as sequential backward floating search (SBFS), extended sequential forward search (ESFS), and sequential forward floating search (SFFS) are also employed [74]. Naseriparsa et al proposed a hybrid method using information gain and genetic algorithm-based searching method combined with a supervised learning algorithm [75]. Huda et al developed a hybrid feature selection method by combining the mutual information (MI) and artificial neural network (ANN) [76]. Gunal presented a hybrid feature selection method by combining filter and wrapper method for text classification. In this method, information gain measure is used for ranking the significant features and the genetic algorithm is used as the searching strategy with support vector machine [77].

D'Alessandro et al proposed a hybrid approach for epileptic seizure prediction, in which ranking with genetic algorithm-based wrapper approach was implemented [78]. Yang et al developed a hybrid method for classifying the micro array data. In this method, the information gain and correlation metric are used for filter method and an improved binary particle swarm optimization (BPSO) method is used with the supervised learning algorithm as the wrapper method to improve the performance of the classification algorithm. The performance of this method is evaluated using kNN and SVM classifiers [79]. To avoid the computational cost of the wrapper method, Bermejo presented a hybrid method by combining the filter and wrapper methods. In this method, the GRASP meta-heuristic based on stochastic algorithm is used as filter method for reducing the wrapper computation [80]. Foithong et al also designed a hybrid feature selection method by combining the filter and the wrapper methods. In this method, the mutual information criterion is used for filtering the relevant features and the supervised learning algorithm is adopted as the wrapper method for evaluating features obtained from the filter method [81].

Further, it is observed that the hybrid methods are computationally intensive than the filter methods since they combine the wrapper and filter methods and have less generality compared to the filter methods since they use the supervised learning algorithm in feature selection process. These hybrid methods take more computational time than the filter-based methods.

## 4. SUMMARY

This section summarizes the feature selection methods that are categorized based on how the features are combined in the selection process namely feature subset-based and feature ranking-based and based on how the supervised learning algorithm used namely wrapper, embedded, hybrid, and filter.

The subset-based methods generate the feature subsets using any one of the searching strategies for evaluation. The exhaustive or complete search is used to generate the subset that leads to high computational complexity since maximum  $2^N$  number of possible combination of the subsets to be generated from the N number of features to evaluate them. This is a "brute force" method so this is not suitable for high-dimensional space. The heuristic search such as SA, TS, ACO, GA, and PSO are employed to reduce the number of feature subset generation for evaluation using the heuristic function. The subset-based feature selection methods using the heuristic search lead to more computational complexity, because they need prior knowledge and each generated subset need to develop a classification model to evaluate them.

However, this heuristic search methods follow a wrapper-based approach therefore these methods are computationally expensive and they can only produce higher classification accuracy for the specific classification algorithm used to obtain the fitness or heuristic function. Therefore, these methods cannot achieve high generality. The ranking-based methods take less computation time and achieve high generality since they do not use the supervised learning algorithm. They cannot remove the redundant features since they only compute the correlation or similarity between the individual feature and the target-class. Therefore, they can be a suitable choice for high dimensional space with a suitable redundancy analysis mechanism.

The wrapper, embedded, and hybrid methods are computationally inefficient than the filter approach. In addition, they do not have high generality since they use the supervised learning algorithm in feature selection process. Therefore, the filter methods are the best choice for the high-dimensional data. Further, the filter methods are preferred because they can perform better with any classification algorithm since they possess better generality and require less computational complexity. The ranking-based approaches are better than the feature subset-based methods since the subset-based methods require more space and computational complexity. Therefore, the ranking-based methods are the best choice for selecting the relevant features from the high-dimensional space.

In the feature selection literature, some researchers have succeeded in effectively removing the irrelevant features, but failed to handle the redundant features. On the other hand, some other researchers dealt with removing the irrelevant features and redundant features. Furthermore, the state-of-the-art feature selection methods reported in literature use the rule-based metric and nearest neighbor principles. Both of the methods eliminate the irrelevant features, but fail to treat the redundant features. Some of the methods use the information-theoretic-based metric to calculate the relevancy between the feature and the target-class for relevancy analysis and to calculate the independency among features for redundancy analysis.

In most of the information-theoretic-based approaches, the same metric is used for both redundancy and irrelevancy analysis. Some of these approaches perform pair wise analysis to identify the independency among the features for redundancy analysis, resulting in increased time complexity. They do not have special mechanism for treating the redundant features, yet they do moderate redundancy analysis. Most of the clustering-based approaches use the hierarchical clustering for feature selection and they deal with specific types of datasets. However, the hierarchical clustering is expensive for high-dimensional datasets and less effective in high-dimensional space due to the dimensionality phenomenon. Hence, simple, scalable, and faster K-means clustering algorithm can be used [82] for relevancy analysis in feature selection.

## 5. CONCLUSION

This paper analyzed several feature selection methods that are proposed by various researchers. From the earlier research works, it is observed that the feature ranking-based methods are better than the subset-based methods in terms of memory space and computational complexity and the ranking-based methods do not reduce the redundancy. Further, the wrapper, embedded, and hybrid methods are computationally inefficient than the filter method and they have poor

generality. Therefore, the feature selection can be developed for high-dimensional data using the filter approach with ranking method for selecting the significant features from the high-dimensional space. In addition, to overcome the limitations of the ranking method the redundancy analysis mechanism can be adopted with a suitable clustering approach.

## 6. REFERENCES

- [1] Saeys, Y, Inza, I & Larrañaga, P 2007, 'A review of feature selection techniques in bioinformatics. *Bioinformatics*', vol. 23, no. 19, pp.2507-2517
- [2] Bolón-Canedo, V, Sánchez-Marroño, N & Alonso-Betanzos, A, 2013, 'A review of feature selection methods on synthetic data', *Knowledge and information systems*, vol. 34, no.3, pp.483-519.
- [3] Hall, MA 1999, *Correlation-based feature selection for machine learning*, Ph.D. thesis, The University of Waikato, NewZealand.
- [4] Liu, H & Setiono, R 1996, 'A probabilistic approach to feature selection-a filter solution', *Proceedings of Eighteenth International Conference on Machine Learning*, Italy, pp. 319-327.
- [5] Lisnianski, A, Frenkel, I & Ding, Y, 2010, 'Multi-state system reliability analysis and optimization for engineers and industrial managers', Springer, New York.
- [6] Lin, S.W, Tseng, TY, Chou, SY & Chen, SC 2008, 'A simulated-annealing-based approach for simultaneous parameter optimization and feature selection of back-propagation networks', *Expert Systems with Applications*, vol. 34, no.2, pp.1491-1499.
- [7] Meiri, R & Zahavi, J 2006, 'Using simulated annealing to optimize the feature selection problem in marketing applications', *European Journal of Operational Research*, vol.171, no.3, pp.842-858.
- [8] Zhang, H & Sun, G 2002, 'Feature selection using tabu search method', *Pattern recognition*, vol. 35, no.3, pp.701-711.
- [9] Tahir, MA, Bouridane, A & Kurugollu, F 2007, 'Simultaneous feature selection and feature weighting using Hybrid Tabu Search/K-nearest neighbor classifier' *Pattern Recognition Letters*, vol. 28, no.4, pp.438-446.
- [10] Aghdam, MH, Ghasem-Aghaee, N & Basiri, ME 2009, 'Text feature selection using ant colony optimization', *Expert systems with applications*, vol. 36, no.3, pp.6843-6853.
- [11] Kanan, HR & Faez, K 2008, 'An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system', *Applied Mathematics and Computation*, vol. 205, no.2, pp.716-725.
- [12] Sivagaminathan, RK & Ramakrishnan, S 2007, 'A hybrid approach for feature subset selection using neural networks and ant colony optimization', *Expert systems with applications*, vol. 33, no.1, pp.49-60.
- [13] Sreeja, NK & Sankar, A 2015, 'Pattern Matching based Classification using Ant Colony Optimization based Feature Selection', *Applied Soft Computing*, vol. 31, pp.91-102.

- [14] Welikala, R.A, Fraz, MM, Dehmeshki, J, Hoppe, A, Tah, V, Mann, S, Williamson, TH & Barman, SA 2015, 'Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy', *Computerized Medical Imaging and Graphics*, vol. 43, pp.64-77.
- [15] Erguzel, TT, Ozekes, S, Tan, O & Gultekin, S 2015, 'Feature Selection and Classification of Electroencephalographic Signals an Artificial Neural Network and Genetic Algorithm Based Approach', *Clinical EEG and Neuroscience*, vol. 46, no.4, pp.321-326.
- [16] Oreski, S & Oreski, G 2014, 'Genetic algorithm-based heuristic for feature selection in credit risk assessment', *Expert systems with applications*, vol. 41, no.4, pp.2052-2064.
- [17] Li, S, Wu, H, Wan, D & Zhu, J, 2011, 'An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine', *Knowledge-Based Systems*, vol. 24, no.1, pp.40-48.
- [18] Das, N, Sarkar, R, Basu, S, Kundu, M, Nasipuri, M & Basu, DK 2012, 'A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application', *Applied Soft Computing*, vol.12, no.5, pp.1592-1606.
- [19] Wang, Y, Chen, X, Jiang, W, Li, L, Li, W, Yang, L, Liao, M, Lian, B, Lv, Y, Wang, S & Wang, S 2011, 'Predicting human microRNA precursors based on an optimized feature subset generated by GA-SVM', *Genomics*, vol. 98, no.2, pp.73-78.
- [20] Xue, B, Zhang, M & Browne, WN 2013, 'Particle swarm optimization for feature selection in classification: A multi-objective approach', *IEEE Transactions on Cybernetics*, vol. 43, no.6, pp.1656-1671.
- [21] Chen, LF, Su, CT, Chen, KH & Wang, PC 2012, 'Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis', *Neural Computing and Applications*, vol. 2, no. 8, pp.2087-2096.
- [22] Yang, H, Du, Q & Chen, G 2012, 'Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5 no.2, pp.544-554.
- [23] Lin, SW, Ying, KC, Chen, SC & Lee, ZJ 2008, 'Particle swarm optimization for parameter determination and feature selection of support vector machines', *Expert systems with applications*, vol. 35, no. 4, pp.1817-1824.
- [24] Liu, H & Setiono, R 1995, 'Chi2: Feature selection and discretization of numeric attributes', *Proceedings of the IEEE Seventh International Conference on Tools with Artificial Intelligence*, Washington DC, USA, pp. 388-391.
- [25] Dash, M & Liu, H 1997, 'Feature selection for classification', *Intelligent data analysis*, vol. 1, no.1, pp.131-156.
- [26] Kohavi, R & John, GH 1997, 'Wrappers for feature subset selection', *Artificial intelligence*, vol. 97, no.1, pp.273-324.
- [27] Inza, I, Larrañaga, P, Etxebarria, R & Sierra, B 2000, 'Feature subset selection by Bayesian network-based optimization', *Artificial intelligence*, vol. 123, no. 1, pp.157-184.
- [28] Grimaldi, M, Cunningham, P & Kokaram, A 2003, 'An evaluation of alternative feature selection strategies and ensemble techniques for classifying music', *Proceedings of Fourteenth European Conference on Machine Learning and the Seventh European Conference on Principles and Practice of Knowledge Discovery in Databases*, Dubrovnik, Croatia
- [29] Dy, JG & Brodley, CE 2000, 'Feature subset selection and order identification for unsupervised learning', *proceedings In Proceedings of the Seventeenth International Conference on Machine Learning*, p. 247-254.
- [30] Aha, DW & Bankert, RL 1996, 'A Comparative Evaluation of Sequential Feature Selection Algorithms', Springer, New York.
- [31] Maldonado, S & Weber, R 2009, 'A wrapper method for feature selection using support vector machines', *Information Sciences*, 179(13), pp.2208-2217.
- [32] Gütlein, M, Frank, E, Hall, M & Karwath, A 2009, 'March. Large-scale attribute selection using wrappers', *Proceeding of IEEE Symposium on Computational Intelligence and Data Mining*, Nashville, TN, USA, pp. 332-339.
- [33] Kabir, MM, Islam, MM & Murase, K 2010, 'A new wrapper feature selection approach using neural network', *Neurocomputing*, vol. 73, no. 16, pp.3273-3283.
- [34] Stein, G, Chen, B, Wu, AS & Hua, KA 2005, 'March. Decision tree classifier for network intrusion detection with GA-based feature selection' *Proceedings of the forty-third ACM Annual Southeast regional conference*, Kennesaw, GA, USA, vol. 2, pp. 136-141.
- [35] Zhuo, L, Zheng, J, Li, X, Wang, F, Ai, B & Qian, J 2008, 'A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine' *Proceedings of Geoinformatics and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images*, pp. 71471J-71471J.
- [36] Quinlan JR 2014, 'C4.5: programs for machine learning', Morgan Kaufmann publishers, San Mateo, California.
- [37] Koistinen, P & Holmström, L 1991, 'Kernel regression and backpropagation training with noise', *Proceedings of IEEE International Joint Conference on Neural Networks*, pp. 367-372.
- [38] Baluja, S 1994, 'Population-based incremental learning a method for integrating genetic search based function optimization and competitive learning', *Technical Report No. CMU-CS-94-163*, Carnegie Mellon University, Pittsburgh, Pa.
- [39] Buntine, W 1991, 'Theory refinement on Bayesian networks' *Proceedings of the Seventh conference on*

- Uncertainty in Artificial Intelligence, Barcelona, Spain, pp. 52-60.
- [40] Loughrey, J & Cunningham, P 2005, 'Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets' Proceedings of Research and Development in Intelligent Systems, Springer London. pp. 33-43.
- [41] Freeman, C, Kulić, D & Basir, O 2015, 'An evaluation of classifier-specific filter measure performance for feature selection, Pattern Recognition, vol.48, no.5, pp.1812-1826.
- [42] Chandrashekar G & Sahin, F 2014, 'A survey on feature selection methods', Computers & Electrical Engineering, vol. 40, no.1, pp.16-28.
- [43] Guyon, I, Weston, J, Barnhill, S & Vapnik, V, 2002, 'Gene selection for cancer classification using support vector machines, Machine learning, vol. 46, no. 1-3, pp.389-422.
- [44] Quinlan, JR, 1986, 'Induction of decision trees', Machine learning, vol. 1, no.1, pp.81-106.
- [45] Tibshirani, R, Saunders, M, Rosset, S, Zhu, J & Knight, K 2005, 'Sparsity and smoothness via the fused lasso', Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, no. 1, pp. 91-108.
- [46] Ma, S & Huang, J 2008, 'Penalized feature selection and classification in bioinformatics', 'Briefings in bioinformatics, vol. 9, no. 5, pp.392-403.
- [47] Neumann, J, Schnörr, C & Steidl, G 2004, 'SVM-based feature selection by direct objective minimisation', Proceeding of the twenty-sixth DAGM Symposium on Pattern Recognition, Germany, pp. 212-219.
- [48] Xiao, Z, Dellandrea, E, Dou, W & Chen, L 2008, 'ESFS: A new embedded feature selection method based on SFS', Rapports de recherche.
- [49] Maldonado, S, Weber, R & Famili, F 2014, 'Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines', Information Sciences, vol. 286, pp.228-246.
- [50] Kira, K & Rendell, LA 1992, 'A practical approach to feature selection', Proceedings of the ninth international workshop on Machine learning, Aberdeen, Scotland, UK (pp. 249-256).
- [51] Kononenko, I 1994, 'Estimating attributes: analysis and extensions of RELIEF'. Proceeding of European Conference on Machine Learning, Catania, Italy, pp. 171-182.
- [52] Holte, RC 1993, 'Very simple classification rules perform well on most commonly used datasets, Machine learning, vol.11, no.1, pp.63-90.
- [53] Yang, HH & Moody, JE 1999, 'Data Visualization and Feature Selection: New Algorithms for Nongaussian Data', Advances in Neural Information Processing Systems, vol. 99, pp. 687-693.
- [54] Peng, H, Long, F & Ding C 2005, '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.
- [55] Battiti, R 1994, 'Using mutual information for selecting features in supervised neural net learning', 'IEEE Transactions on Neural Networks, vol. 5, no. 4, pp.537-550.
- [56] Fleuret, F 2004, 'Fast binary feature selection with conditional mutual information', The Journal of Machine Learning Research, vol. 5, pp.1531-1555.
- [57] Meyer, PE & Bontempi, G 2006, 'On the use of variable complementarity for feature selection in cancer classification', Applications of Evolutionary Computing, pp. 91-102.
- [58] Lin, D & Tang, X 2006, 'Conditional infomax learning: an integrated framework for feature extraction and fusion', Proceeding of ninth European Conference on Computer Vision, Graz, pp. 68-82.
- [59] Brown, G, Pocock, A, Zhao, MJ & Luján, M 2012, 'Conditional likelihood maximisation: a unifying framework for information theoretic feature selection', The Journal of Machine Learning Research, vol.13, no.1, pp.27-66.
- [60] Song, Q, Ni, J & Wang, G 2013, 'A fast clustering-based feature subset selection algorithm for high-dimensional data', IEEE Transactions on Knowledge and Data Engineering, vol. 25, no.1, pp.1-14.
- [61] Dhillon, IS, Mallela, S & Kumar, R 2003, 'A divisive information theoretic feature clustering algorithm for text classification', The Journal of Machine Learning Research, vol. 3, pp.1265-1287.
- [62] Li, Y, Luo, C, & Chung, SM 2008, 'Text clustering with feature selection by using statistical data. IEEE Transactions on Knowledge and Data Engineering, vol. 20, no.5, pp.641-652.
- [63] Cai, D, Zhang, C & He, X 2010, 'Unsupervised feature selection for multi-cluster data', Proceedings of the sixteenth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, pp. 333-342.
- [64] Chow, TW & Huang, D 2005, 'Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information', IEEE Transactions on Neural Networks, vol.16, no.1, pp.213-224.
- [65] Mitra S, & Acharya T 2005, 'Data mining: multimedia, soft computing, and bioinformatics' John Wiley & Sons, New Jersey.
- [66] Sotoca, JM & Pla, F 2010, 'Supervised feature selection by clustering using conditional mutual information-based distances', Pattern Recognition, vol. 43, no.6, pp.2068-2081
- [67] Freeman, C, Kulić, D & Basir, O 2015, 'An evaluation of classifier-specific filter measure performance for feature selection, Pattern Recognition, vol.48, no.5, pp.1812-1826.
- [68] Frénay, B, Doquire, G & Verleysen, M 2014, 'Estimating mutual information for feature selection in the presence of label noise, Computational Statistics & Data Analysis, vol. 71, pp.832-848.
- [69] Tabakhi, S, Moradi, P & Akhlaghian, F 2014, 'An unsupervised feature selection algorithm based on ant



- colony optimization', *Engineering Applications of Artificial Intelligence*, vol. 32, pp.112-123.
- [70] Bermejo, P, Gámez, J & Puerta, J 2008, 'On incremental wrapper-based attribute selection: experimental analysis of the relevance criteria, *Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, France*, pp.638-645.
- [71] Ruiz, R, Riquelme, JC & Aguilar-Ruiz, JS 2006, 'Incremental wrapper-based gene selection from microarray data for cancer classification' *Pattern Recognition*, vol. 39, no. 12, pp.2383-2392.
- [72] Xie, J, Xie, W, Wang, C & Gao, X 2010, 'A Novel Hybrid Feature Selection Method Based on IFSFFS and SVM for the Diagnosis of Erythematous-Squamous Diseases', *Proceedings of Workshop on Applications of Pattern Analysis, Cumberland Lodge, Windsor, UK*, pp. 142-151.
- [73] Kannan, SS & Ramaraj, N 2010, 'A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm', *Knowledge-Based Systems*, vol. 23, no. 6, pp.580-585.
- [74] Xie, J, Lei, J, Xie, W, Shi, Y & Liu, X 2013, 'Two-stage hybrid feature selection algorithms for diagnosing erythematous-squamous diseases', *Health Information Science and Systems*, vol.1, no.10, pp.2-14.
- [75] Naseriparsa, M, Bidgoli, AM & Varae, T 2013, 'A Hybrid Feature Selection method to improve performance of a group of classification algorithms', *International Journal of Computer Applications*, vol. 69, no. 17, pp. 0975 – 8887.
- [76] Huda, S, Yearwood, J & Stranieri, A 2011, 'Hybrid wrapper-filter approaches for input feature selection using maximum relevance-minimum redundancy and artificial neural network input gain measurement approximation (ANNIGMA)', *Proceedings of the Thirty-Fourth Australasian Computer Science Conference, Australia*, vol. 113, pp. 43-52.
- [77] Gunal, S 2012, 'Hybrid feature selection for text classification', *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 20, no.2, pp.1296-1311.
- [78] D'Alessandro, M, Esteller, R, Vachtsevanos, G, Hinson, A, Echauz, J & Litt, B 2003, 'Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients', *IEEE Transactions on Biomedical Engineering*, vol. 50, no.5, pp.603-615.
- [79] Yang CS, Chuang LY, Ke CH, & Yang CH, 2008, 'A hybrid feature selection method for microarray classification, *IAENG International Journal of Computer Science*, vol. 35, no. 3, pp. 1-3.
- [80] Bermejo, P, Gámez, JA & Puerta, JM 2011, 'A GRASP algorithm for fast hybrid (filter-wrapper) feature subset selection in high-dimensional datasets', *Pattern Recognition Letters*, vol. 32, no.5, pp.701-711.
- [81] Foithong, S, Pinngern, O & Attachoo, B 2012, 'Feature subset selection wrapper based on mutual information and rough sets', *Expert Systems with Applications*, vol. 39, no.1, pp.574-584.
- [82] Coates A, Ng AY (2012) Learning Feature Representations with K-Means. In: Montavon, G., Orr, G.B., and Müller, K.-R. (eds.) *Neural Networks: Tricks of the Trade*. pp. 561–580. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-35289-8\_30