

# The Effect of Clustering in Filter Method Results Applied in Medical Datasets

Nadjla Elong, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), Algeria

Sidi Ahmed Rahal, University of Sciences and Technology of Oran Mouhamed Boudiaf (USTO-MB), Algeria

## ABSTRACT

For a deeper and richer analytic processing of medical datasets, feature selection aims to eliminate redundant and irrelevant features from the data. While filter has been touted as one of the simplest methods for feature selection, its applications have generally failed to identify and deal with embedded similarities among features. In this research, a hybrid approach for feature selection based on combining the filter method with the hierarchical agglomerative clustering method is proposed to eliminate irrelevant and redundant features in four medical datasets. A formal evaluation of the proposed approach unveils major improvements in the classification accuracy when results are compared to those obtained via only the applications of the filter methods and/or more classical-based feature selection approaches.

## KEYWORDS

Classification, Feature Selection, Filter Method, HAC Algorithm, Medical Datasets, Similarity Measures

## 1. INTRODUCTION

In vying for a deeper and richer analytic processing of medical datasets, a key challenge in building a superior classification model via machine learning (ML) is the identification of a set of representative features that are inherently embedded in cumulative health datasets. Briefly, this representative set of features should contain mostly relevant and non-redundant features so as to achieve improved accuracy and better classification results for data modeling.

Even so, a preprocessing step for ML is to draw out such a representative set of features embedded in the cumulative datasets; essentially, the primary goal is to eliminate potentially redundant and irrelevant features that may affect the accuracy of the classifier (Chandrashekar & Sahin, 2014). One strategy is to implement an appropriate mix of feature selection methods that result in optimizing feature parsimony. To date, many researchers have adopted the use of filter methods as these are among the simplest methods to implement for data analysis (Karegowda, Manjunath & Jayaram, 2010); however, a key challenge faced in just using the filter methods is its inability to take into consideration the interactions among feature characteristics (Saeys, Inza & Larrañaga, 2007). Consequently, when the obtained set of features is likely to be characterized with redundant and similar features, the accuracy of the classifier may be compromised. This research addresses the downside of adopting

DOI: 10.4018/IJHISI.2021010103

This article, published as an Open Access article on February 4, 2021 in the gold Open Access journal, International Journal of Healthcare Information Systems and Informatics (IJHISI) (converted to gold Open Access January 1, 2021), is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

the filter methods by proposing a novel approach based on integrating critical aspects of the filter methods and hierarchical agglomerative clustering algorithm (HAC).

HAC algorithm is useful in detecting similar features; computationally, this is done by calculating similarities between features and grouping them into clusters. In this sense, the current research aims at investigating the feature selection problem via feature clustering (or variable clustering). Essentially, a representative feature for each cluster based on its relevance is obtained by the adopted filter methods - a fundamental step that can be taken to effectively reduce similar features while decreasing the number of features in the resulting set. To further validate the proposed novel approach, a series of experiments has been conducted with many parameters successively evaluated to achieve better results. These tests have been applied on four medical datasets, namely, the Wisconsin Diagnostic Breast Cancer (or WDBC), Texture, Spectheart and Movement-libras datasets. Accuracy rates of four classifiers are also obtained to determine if the suggested novel approach can in fact effectively enhance classification results.

The organization of the rest of this paper is as follows. Section 2 reviews the extant literature on related works, emphasizing the more popularly applied feature selection methods of past researchers. Section 3 details the proposed novel approach and the experimental methodology reported here. Section 4 highlights a comparative study of four feature selection methods with the proposed novel approach, summarizing the findings and implications of these results. Finally, Section 5 offers concluding remarks with a look at potential study limitations as well as extensible future works.

## 2. BACKGROUND

Past research has investigated the applications of various feature selection methods that are of growing interests to the medical data analytics research community (Polat & Güneş, 2009; Akay, 2009; Shilaskar & Ghatol, 2013; Lavanya & Rani, 2011; Anbarasi, Anupriya & Iyengar, 2010; Inbarani, Azar & Jothi, 2014; Kumar, Ramachandra & Nagamani, 2014; Ibrahim, Ojo & Oluwafisoye, 2018). The wrapper methods, for example, have been commonly adopted as these approaches have provided good accuracy; however, these methods also require high computational time. In 1997, Kohavi & John reported a study of the wrapper methods where they explored the relation between optimal feature subset selection and relevance; as well, their study examined the various strengths v. weaknesses of the wrapper approach.

Filter methods, in contrast, are simpler feature selection methods and are easier to implement. Yet, these methods are limited by the fact that redundant features are often neglected. Even so, more recent filter-based feature selection approach such as the mRMR (minimum Redundancy Max Relevancy) has been designed for improved feature selection of microarray data. Accordingly, in the mRMR algorithm, the features are ranked on the basis of the minimal-redundancy-maximal-relevance criteria. The method tends to select highly predictive but uncorrelated features simultaneously (Radovic, Ghalwash, Filipovic, & Obradovic, 2017). Other prominent feature selection approaches include a Fast Correlation Based Filter (FCBF) solution, FAST and other feature selection methods that used Genetic Algorithms (GAs), including Genetic Programming (GP) and Particle Swarm Optimization (PSO) approaches.

FCBF is a feature selection method based on correlation measure, relevance and redundancy analysis (Yu & Liu, 2003). This method is used in conjunction with an attribute set evaluator; notwithstanding, FCBF has the advantage of being able to identify relevant features as well as redundancy among relevant features. More recently, researchers have applied feature selection methods to explore the attribute space via techniques that are reputed for their global search ability such as the Evolutionary Computation (EC) methods (Xue, Zhang, Browne & Yao, 2015) PSO is a relatively recent EC technique based on swarm intelligence. In comparison with other EC algorithms such as GAs and GP, PSO is computationally less expensive and can converge more quickly (Xue, Zhang & Browne, 2012). Some researchers have applied clustering for feature selection purpose. An

example of such clustering method for feature selection has been proposed in (Hong, Liou, Wang & Vo, 2014). These authors used clustering of the attributes for feature selection. A distance measure for a pair of attributes based on the relative dependency is proposed. As the attributes are grouped into several clusters according to their similarity degrees, an attribute selected from a cluster can thus represent the attributes within the same cluster. An approximate “reduct” could then be formed from the chosen and accumulated attributes. The Hong et al. study achieved good results but had also been limited by its need to specify the number of clusters in advance.

FAST, another clustering algorithm, has been proposed by (Sudhakar, Priya, Chandini & Narasimham, 2016). FAST works in three phases. The first phase entails computing T-relevance test to eliminate irrelevant features. The second phase involves constructing and partitioning the MST (Minimum Spanning Tree) by applying graph theoretical algorithms. The final phase has to do with selecting the most representative feature from each cluster. According to the results the FAST clustering algorithm effectively reduces the dimensionality and also improves accuracy of predictive classifiers

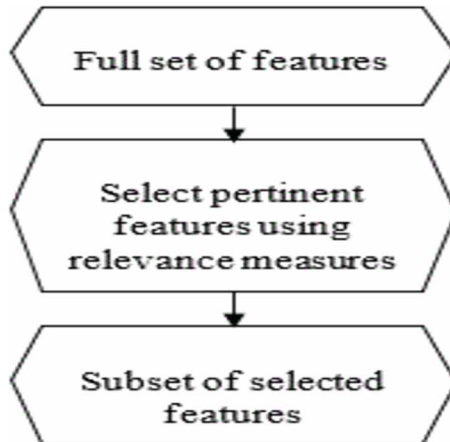
Additionally, a hierarchical feature selection model to reduce the feature space has been proposed in Zhang, Wang & Yoon (2017). This model includes three stages: Stage (1) entails irrelevant features elimination; Stage (2) involves feature clustering based on conditional mutual information (CMI); and Stage (3) pertains to recursive feature elimination. Another work attempting to use Hierarchical Clustering method for feature selection is Park (2013); here, a new similarity measure between two feature groups is defined by directly using the representative feature in each group. A further attempt to use and test the use of hierarchical clustering for feature selection is reported by Ienco & Meo (2008) in which the clustering method adopted is that of Ward’s with a distance measure based on Goodman-Kruskal tau.

In Butterworth, Piatetsky-Shapiro & Simovici (2005), the authors studied an algorithm for feature selection that clusters attributes using a special metric; then, it makes use of the dendrogram of the resulting cluster hierarchy to choose the most relevant attributes. In Nagendruru & Reddy (2015), an efficient feature subset selection technique has been proposed. This algorithm works in two steps: (a) First, features are divided into clusters by using graph-theoretic clustering methods; and (b) Second, the most representative feature that is strongly related to the target classes is selected from each cluster to form a subset of features. Most recently, Chormunge & Jena (2018) proposed a new method where clustering is integrated with correlation measure to produce good feature subset. Specifically, k-means clustering method is used; briefly, this implies that the researchers need to specify “k” number of clusters in advance. Two clusters for each dataset were used in the aforementioned study to obtain the results contrary to our proposed approach in which the number of cluster “k” is not being specified in advance.

While most of the aforementioned methods have attempted to use either the filter methods or one of the clustering of variables for feature selection, our research emphasizes the need to integrate both the filter methods and the use of clustering of variables for feature selection. More specifically, it focuses on enhancing the filter methods and how their results can be improved by using an optimal clustering of variables. In this area, a new selector for relevant features, namely, the Fisher Markov selector, has been proposed in Cheng, Zhou & Cheng (2010). Their work presents a way to represent essential discriminating characteristics together with the “sparsity” as an optimization objective. This may be regarded as a filter method; consequently, any available classifier can also be applied. Given that the proposed method is relevant, applying four classifiers may then also validate the results.

Filter methods use relevance measurements to calculate the relevancy of attributes, but they do not take into account relations between these attributes. The resultant subset can contain redundant features. In order to resolve this challenge, a new approach for feature selection which is divided into two steps is proposed in the current study: first, filtering the set of attributes so as to identify the relevant attributes; and second, using a clustering of variables method to identify the redundant features. More specifically, the filter methods we adopted here will rank features via the following selection metrics: Information Gain, Gain Ratio, ReliefF and OneR, and the HAC method for clustering

Figure 1. Feature selection with filter method (Asir et al., 2016)



features. The purpose of this study is to evaluate the improvement of results achievable via filter methods by applying clustering variables algorithm.

For the experiments, we used Weka (Waikato Environment for Knowledge Analysis), which is a collection of ML algorithms for data mining tasks. More specifically, for experiments on feature selection methods and on the classification task, we apply SPSS statistics for HAC algorithm.

### 3. THE PROPOSED NOVEL APPROACH

For the present study, we propose integrating the clustering variables technique with different filtering methods. This approach is applied with two different similarity measures in order to improve filter method results for classifying medical datasets

#### 3.1 Filter Methods

Filter methods are feature ranking techniques that evaluate the relevance of features by looking at the intrinsic properties of the data independent of the classification algorithm (Mwadulo, 2016). A suitable ranking criterion is used to score the variables and a threshold is used to remove the variable below the threshold (Mwadulo, 2016).

Using a simple filter method, feature selection may be done once before it can be provided as input to the different classifiers (Karegowda et al., 2010). In feature ranking methods, each feature is ranked by a selection metric such as information gain, symmetric uncertainty, gain ratio, or some other meaningful criteria. The top ranked features are selected as relevant features by a pre-defined threshold value (Asir, Appavu & Jebamalar, 2016).

Advantages of filter techniques are that not only can they be scaled easily to very high-dimensional datasets, but they are also computationally simple and fast, besides being independent from the classification algorithm (Saeys et al., 2007). However, many filter techniques pose somewhat similar challenges: (1) these methods ignore the interaction with the classifier, and (2) most of the proposed techniques are “univariate,” which means that each feature is considered separately, by ignoring feature dependencies that may lead to worsening classification performance when compared to other types of feature selection techniques (Saeys et al., 2007).

The general process of filter-based feature selection method is given in Figure 1.

Filter method (Figure 1) 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 other methods. Therefore, it is suitable for high-dimensional space (Asir et al., 2016).

In the following part, we consider these four commonly used ranking measures for further investigation in evaluating our proposed approach: (1) Information Gain (IG) attribute evaluation; (2) Gain Ratio (GR) attribute evaluation; (3) Relief-F (RF) attribute evaluation; and (4) OneR attribute evaluation.

### 3.1.1 Information Gain (IG)

IG (relative entropy, or Kullback-Leibler divergence), in probability theory and information theory, is a measure of the difference between two probability distributions. It evaluates a feature X by measuring the amount of information gained with respect to the class (or group) variable Y, defined as follows:

$$I(X) = H(P(Y)) - H\left(P\left(\frac{Y}{X}\right)\right) \quad (1)$$

Specifically, IG measures the difference between the marginal distribution of observable Y assuming that it is independent of feature  $H(P(Y))$  and the conditional distribution of Y assuming that it is dependent of  $H\left(P\left(\frac{Y}{X}\right)\right)$ . If X is not differentially expressed, Y will be independent of X; thus, X will have small IG value, and vice versa (Yildirim, 2015).

### 3.1.2 Gain Ratio (GR)

The GR is the non-symmetrical measure that is introduced to compensate for the bias of the IG. GR is computed as:

$$GR = \frac{IG}{H(X)} \quad (2)$$

As Equation (2) presents, when the variable Y has to be predicted, we normalize the IG by dividing by the entropy of X, and vice versa.

Owing to this normalization, the GR values always fall in the range [0, 1]. A value of GR = 1 indicates that the knowledge of X completely predicts Y, whereas GR = 0 means that there is no relation between Y and X. In contrast to IG, the GR favors variables with fewer values (Mwadulo, 2016).

### 3.1.3 Relief-F (RF)

RF is an instance-based feature selection method. Accordingly, Saeys, et al. (2007) used RF to evaluate a feature by how well its value distinguishes samples that are from different groups but are similar to each other.

For each feature X, RF selects a random sample and k of its nearest neighbors from the same class and each of different classes. Then, X is scored as the sum of weighted differences in different classes and the same class. If X is differentially expressed, it will show greater differences for samples from different classes, thus yielding a higher score (or vice versa) as noted in Yildirim (2015).

### 3.1.4 OneR

OneR is a simple algorithm. It builds one rule for each attribute in the training dataset; then, it selects the rule with the smallest error. OneR treats all numerically valued features as continuous and uses a

straightforward method to divide the range of values into several disjoint intervals. OneR also handles missing values by treating “missing” as a legitimate value.

OneR is one of the most primitive schemes, producing simple rules based on one feature only. Although it is a minimal form of a classifier, OneR can be useful for determining a baseline performance as a benchmark for other learning schemes (Novaković, 2016).

### 3.2 Proposed Approach for Feature Selection

In this research, the use of clustering method is advocated given that the overarching goal of this study is to see if such an application will improve anticipated results of the filter methods by determining similarities among features, and thereby, eliminating those redundant ones.

Notably, a two-step process is proposed. The process essentially combines filter methods with clustering algorithm; in the first step, the filter method is applied with four known ranking measures, namely, IG, GR, RF and OneR. The objective here is to determine the most relevant features in the datasets. With implementing the first step, a sorted list of features from the most relevant to the less relevant can be generated.

Despite the simplicity of the filter method, it cannot determine dependencies between the features in term of computation time. Thus, the resultant feature subset in the dataset can and will still contain redundant features. For this reason, the second step is applied, which entails the HAC clustering method. The role of applying this algorithm is to determine similarities between each pairs of features in the dataset.

HAC algorithm is commonly used because it is conceptually simple and produces a hierarchy of clusters. Beginning with a large number of initial small clusters, the agglomerative clustering algorithms iteratively select two clusters with the largest affinity under certain measures to merge, until some stopping condition is reached (Zhang, Zhao & Wang, 2013).

Additionally, HAC algorithm calculates similarity matrix based on one of the similarity distances. The output of HAC is a hierarchy of clustered features, presented in a dendrogram. HAC algorithm has an advantage over the k-means, which requires the number k of clusters to be pre-specified prior to using the algorithm; HAC does not require the number of clusters to be pre-specified.

In the extant literature, several agglomerative clustering methods have been proposed based on how the distance between two clusters is computed. Examples include: Single linkage, Complete Linkage, Average Linkage, Ward’s Linkage, Centroid Linkage and Median Linkage.

One fundamental step in HAC is to determine a similarity (dissimilarity) measure to identify the degree of similarity (dissimilarity) between two objects or sets. While similarity is an amount that reflects the strength of relationship between two data items, dissimilarity deals with the measurement of divergence between two data items (Irani, Pise & Phatak, 2016). Based on these two methods, Filter methods and HAC algorithm, we proposed an approach for feature selection. Figure 2 represents a summarized process of this proposed integrated approach.

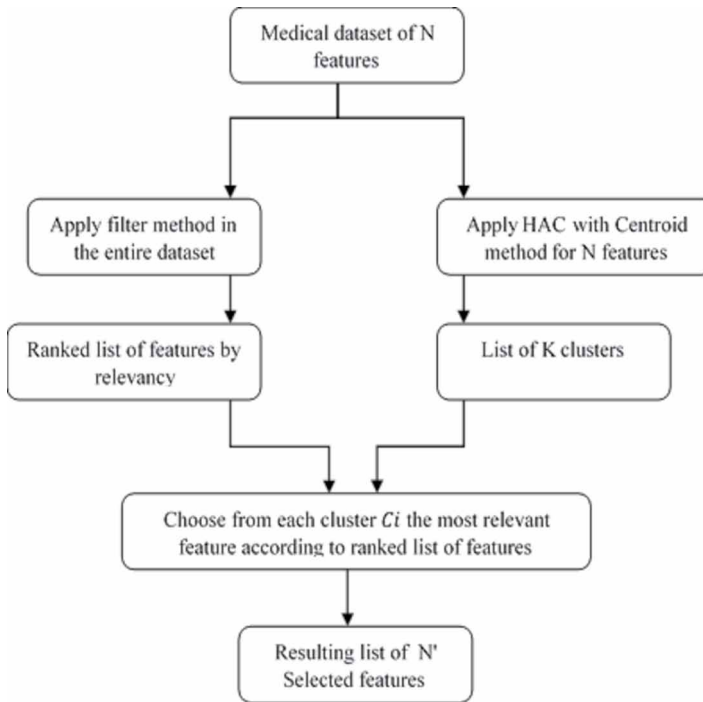
As described, the proposed integrated approach is divided into two steps: First step, filter methods are applied in the given medical datasets. Following this, a ranked list of N features from the most relevant to the less relevant features is generated. It should be noted that in this first step, all set of features would be considered without using the threshold. In the second step, HAC algorithm is applied to cluster the features of the medical datasets. The result of this step is a list of K-clusters, each cluster showcasing the most similar features to each other.

Altogether, from results generated by executing these two steps, a final list of N’ selected features will now be made available by choosing from each resulting cluster from step 2 the most relevant feature according to the ranked list of features obtained from step 1.

It should be noted that the centroid method for HAC algorithm with two similarity measures has been employed in the reported study.

The first similarity measure is the cosine distance between two points. Here, it is computed as one minus the cosine of the included angle between points (treated as vectors); therefore, given an

Figure 2. General schema of proposed approach



m-by-n data matrix X, which is treated as m (1-by-n) row vectors  $x_1, x_2, \dots, x_m$ , the cosine distances between the vector  $x_s$  and  $x_t$  is given by Equation (3) below (Bora, Jyoti, Gupta, & Kumar, 2014):

$$d_{st} = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}} \quad (3)$$

Additionally, a second similarity measure, the Pearson correlation, is also used with its formula given in Equation (4) as:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (4)$$

where  $\mu_X$  denotes the mean and  $\sigma_X$  the standard deviation of X (Berthold & Höppner, 2016).

The Pearson correlation measures the correlation  $\rho$  between two random variables X and Y. We applied the two-steps approach in 4 medical datasets, namely: wdbc, Texture, Spectheart and Movement-libras datasets of 30 attributes, 40 attributes, 44 attributes and 90 attributes respectively. Table 1 provides a brief description of these datasets; notably, the medical datasets are obtained from UCI Machine Learning repository (Dua, Dheeru, & Efi, 2017).

By applying the aforementioned two steps: (a) ranking features based on their relevancy presented in the left branch in the schema, and (b) clustering features based on their similarity presented in the right branch in the schema, the most relevant and similar features may now be identified and

Table 1. Description of medical datasets

Dataset	Number of Attributes	Number of Instances	Number of Classes
Wdbc	30	569	2
Texture	40	5500	11
Spectheart	44	267	2
Movement-libras	90	360	15

generated by choosing from each resulting cluster the most relevant attribute according to a ranked list of attributes.

Our objective is to study the impact of clustering in filtering results, and to show how the hybridization of the filtering-clustering methods for feature selection can improve the limited filtering results and increase the classification accuracy in most cases.

### 3.3 The Classification Methods

Numerous researchers have reported on their investigation into the performance of various classification methods across the data mining literature (Aggarwal, 2014). Among the most common methods used in data classification are decision trees (DTs), rule-based methods (RBM), probabilistic methods (PM), support vector machine (SVM) methods, instance-based (IB) methods, and neural networks (NNs).

A well-known DT method is C4.5, which has been applied in our experiments. CBA (Classification based on Associations) is an example of RBMs whereas Naïve Bayes (NB) classifier has frequently been used in classification as it fits in with many probabilistic methods. The NB classifier is powerful and fast. Another popular classifier is k-NN (k-Nearest Neighbor), which is an IB learning method. Many of the methods cited above faced limitations and challenges; for this reason, recent research has proposed performing data analytics and classification based on a shift from the more traditional means.

A new supervised learning model, designed especially for high-dimensional data and large-scale applications, has been proposed via the discriminative regression approach (Peng, Cheng, & Cheng, 2017). Here, the authors remarked that such a shift is necessary as the classic supervised learning methods have difficulty in achieving good performance for high data dimension. A novel method called weighted NB classifier based on the NB classifier has also been proposed in (Karabatak, 2015) to classify the breast cancer datasets. Another more recent paper also emphasizes the discriminative regression approach to classification (Peng & Cheng, 2019). It estimates a representation for a new example by minimizing the dotting error while explicitly incorporating discrimination information between classes. The experimental results show that the proposed discriminant regression machine outperforms several state-of-the-art classification methods.

Aside from these novel classification methods, the Multilayer Perceptron or MLP has been widely used as a novel NN architecture for classification task. MLP enables a powerful and rich representation with units being arranged in layers comprising an input layer, one or several hidden layers, and an output layer. Owing to the limitation of the scope of this research paper, we are able only to briefly review some of these novel methods that are of interests to the current work although several other novel classification methods have emerged. To validate the experimental results in this study, we have used NB, C4.5, SVM and MLP classifiers.

## 4. RESULTS AND DISCUSSION

As previously noted, we have used clustering of variables to identify similarities between features to improve on results in applying the filter methods; accordingly, we select the most relevant features and eliminate the redundant ones from the resulting list of ranked features and k-clusters of features.



Following these processes, we can now produce our final resulting subset of features, by choosing from each cluster one attribute that is the most relevant. These representative feature(s) characterize the other similar-type features contained in the respective cluster(s). Hence, they automatically provide the same information and a means to further simplify the complicated analytic processes needed in the classification step.

We compare the results based on the accuracy rate; more specifically, we highlight the amelioration in the proposed methods' results when compared to the accuracy rate for the simple filter methods' results. In the first step, we present the results of the experiment of our proposed approach by using the NB classifier. In the second step, we use sequential minimal optimization or SMO (an implementation of SVM classifier), MLP and C4.5 classifiers in order to mount the performances of the proposed approach. To implement these classifiers, the Weka (Waikato Environment for Knowledge Analysis) or an open source Java software ML running environment has been employed for all the selected classification methods.

In order to evaluate the impact of clustering in the filter methods results, a series of tests has been performed. These include: (1) applying the centroid method for clustering with cosine measure, with four filter methods, and (2) with the Pearson correlation.

The limitation of clustering methods is that they cannot determine the optimal number of clusters automatically. As a result, we applied two solutions to determine the number of resulting clusters: (1) Use of fixed number of clusters; for this solution, we choose the value  $N' = N/3$  as the number of resulting subset of features; and (2) Cutting the dendrogram in 3 levels.

These tests were applied for the four medical datasets cited above. The experimental results are summarized in the respective Tables given below. Table 2 highlights the results for the Texture dataset with cosine distance. As shown in Table 2, it can be observed that the clustering method improve filter method results for IG and RF with 13 features, and OneR with 6, 10, and 13 selected features.

**Table 2. Results for texture dataset with cosine distance**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RF	RF+clust	OneR	OneR+clust
6	76.3091%	74.0909%	81.0000%	74.7455%	77.8727%	75.4364%	73.3636%	76.0909%
10	80.0364%	76.1273%	80.3273%	77.0727%	79.2364%	77.5455%	77.5273%	78.6000%
13	79.5818%	80.2545%	81.9091%	80.7091%	79.3455%	80.6364%	78.9455%	79.4545%
20	79.8909%	78.5455%	78.5455%	78.5455%	80.5636%	79.1818%	79.9273%	78.0727%

Table 3 highlights the results for the WDBC dataset with cosine distance. From the application of the proposed approach in the WDBC dataset, the enhancement of results is observed in all cases for 8, 10 and 14 features and only in the case of IG, all selected features.

Table 4 highlights the results for the Spectheart dataset with cosine distance. Results show that accuracy has improved in all cases, except for GR with 22 features.

Table 5 highlights the results for the Movement-libras dataset with cosine distance. Here, we recognize the improvement of classification rate in all cases, between using filter methods alone v. the use of the proposed approach.

A similar experimentation was conducted using centroid method with the four respective datasets with the Pearson correlation. Table 6 highlights the results for the Texture dataset with the Pearson correlation. Here, results show that the accuracy is improved for IG with 4 features, RF with 4 and 16 features, and in two cases with OneR, namely, 4 and 9 features.

**Table 3. Results for WDBC dataset with cosine distance**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RF	RF+clust	OneR	OneR+clust
5	94.3761%	95.2548%	94.3761%	94.3761%	95.4306%	93.4974%	94.3761%	94.0246%
8	93.1459%	96.6608%	94.2004%	96.3093%	94.7276%	95.0791%	94.3761%	95.4306%
10	92.9701%	95.0791%	93.8489%	95.0791%	94.3761%	94.9033%	92.9701%	95.0791%
14	93.4974%	94.9033%	93.3216%	94.3761%	94.3761%	95.0791%	93.3216%	94.3761%

**Table 4. Results for Spectheart dataset with cosine distance**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RELIEFF	RELIEF-F+clust	ONER	ONE-R+clust
6	72.6592%	76.4045%	70.7865%	74.1573%	65.1685%	74.9064%	61.4232%	62.1723%
11	68.1648%	73.7828%	69.6629%	73.7828%	67.7903%	73.4082%	61.4232%	63.6704%
15	70.4120%	74.9064%	67.7903%	74.1573%	67.0412%	74.5318%	62.1723%	65.5431%
22	70.7865%	71.5356%	71.9101%	71.5356%	66.6667%	71.5356%	62.5468%	65.5431%

**Table 5. Results for movement-libras dataset with cosine distance**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RELIEFF	RELIEF-F+clust	ONER	ONE-R+clust
12	33.6111%	72.5000%	38.0556%	69.7222%	30.5556%	74.1667%	47.2222%	71.9444%
23	53.8889%	75.0000%	51.3889%	73.0556%	54.1667%	74.7222%	60.5556%	74.1667%
30	56.3889%	73.0556%	58.0556%	71.1111%	61.3889%	72.7778%	62.5000%	71.9444%
45	68.6111%	75.8333%	66.6667%	76.1111%	64.1667%	75.5556%	68.0556%	74.1667%

**Table 6. Results for texture dataset with the Pearson correlation**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RELIEFF	RELIEF-F+clust	ONER	ONE-R+clust
4	73.6827%	75.0727%	75.9818%	73.8182%	73.5818%	77.4364%	65.7091%	75.9818%
9	78.4000%	77.9091%	82.6545%	80.6545%	78.6545%	78.5091%	77.3636%	78.5818%
13	79.5818%	76.9091%	81.9091%	79.7636%	79.3455%	77.7273%	78.9455%	77.4364%
16	79.2909%	78.3636%	81.7091%	78.7455%	78.9455%	79.0364%	79.1818%	78.1636%

Table 7 highlights the results for the WDBC dataset with the Pearson correlation. Here, results are very satisfying, where it is observed that the accuracy is improved in all cases with the use of proposed approach.

Table 8 highlights the results for the Spectheart dataset with the Pearson correlation. Here, the classification rate is increased in all cases, except in two instances: OneR with 15 features and GR with 22 features.

**Table 7. Results for WDBC dataset with the Pearson correlation**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RELIEFF	RELIEF-F+clust	ONER	ONE-R+clust
6	94.5518%	95.6063%	94.5518%	95.7821%	94.7276%	95.6063%	94.5518%	95.0791%
9	94.2004%	95.4306%	94.2004%	95.0791%	94.0246%	96.1336%	94.2004%	95.6063%
10	92.9701%	94.5518%	93.8489%	94.2004%	94.3761%	95.0791%	92.9701%	94.2004%
16	93.4974%	94.7276%	93.4974%	95.0791%	94.0246%	94.9033%	92.9701%	94.3761%

**Table 8. Results for spectheart dataset with Pearson correlation**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RELIEFF	RELIEF-F+clust	ONER	ONE-R+clust
11	68.1648%	76.4045%	69.6629%	74.5318%	67.7903%	73.4082%	61.4232%	63.6704%
15	70.4120%	75.6554%	67.7903%	74.9064%	67.0412%	73.0337%	62.1723%	61.0487%
18	70.7865%	73.0337%	69.2884%	72.6592%	67.4157%	71.9101%	62.1723%	63.6704%
22	70.7865%	71.5356%	71.9101%	71.5356%	66.6667%	71.5356%	62.5468%	65.5431%

Table 9 highlights the results for the Movement-libras dataset with the Pearson correlation. The applied approach here also provides good results as it is observed that classification accuracy is enhanced in all cases.

Altogether, we see from the obtained results that the proposed approach can enhance classification accuracy. Specifically, this improvement may be achieved via the two different metrics: (a) via the Cosine distance metric; and (b) via the Pearson correlation metric.

Table 10 summarizes the three best results first obtained. With the exception of the Texture dataset, we find that the proposed approach provides a better classification for all datasets. We argue that the IG+Clustering approach, when applied to both the Spectheart and the Movement-libras datasets, dominates with a rate higher than 76% with the two centroid method measures. Similarly, the IG+Clustering approach, when applied to the WDBC dataset with cosine distance, achieved a better performance of 96.66%. However, with the Texture dataset, the GR+Clustering approach achieves only a 1.2% less performance than the GR approach (80.7091% vs. 81.9091%).

To improve the experimental results, the proposed approach with more classifiers, including SMO, MLP and C4.5, was applied to the various datasets. The experimental results show that the proposed approach enhances results for the filter methods for all classifiers as summarized in Tables

**Table 9. Results for movement-libras dataset with Pearson correlation**

Num Att	Method							
	IG	IG+clust	GR	GR+clust	RELIEFF	RELIEF-F+clust	ONER	ONE-R+clust
6	33.0556%	65.2778%	35.0000%	57.5000%	21.6667%	60.0000%	31.9444%	57.2222%
15	39.4444%	72.2222%	40.5556%	71.3889%	39.4444%	76.1111%	55.5556%	71.6667%
26	55.0000%	76.1111%	55.5556%	75.0000%	58.3333%	74.4444%	59.7222%	73.8889%
30	56.3889%	74.1667%	58.0556%	72.7778%	61.3889%	74.1667%	62.5000%	73.0556%

Table 10. Synthesis of the experimental results

Dataset	Attribute	Results	Approach	Centroid Method Measure
Texture	13	81.9091%	GR	cosine distance
	13	80.7091%	GR+Clust	cosine distance
	13	80.6364%	RF+Clust	cosine distance
WDBC	8	96.6608%	IG+Clust	cosine distance
	8	96.3093%	GR+Clust	cosine distance
	9	96.1336%	RF+clust	Pearson correlation
Spectheart	6	76.4045%	IG+Clust	cosine distance
	11	76.4045%	IG+Clust	Pearson correlation
	15	75.6554%	IG+Clust	Pearson correlation
Movement-libras	45	76.1111%	GR+clust	cosine distance
	26	76.1111%	IG+Clust	Pearson correlation
	15	76.1111%	RF+Clust	Pearson correlation

11, 12, 13 and 14. As shown, the various feature selection methods yielding the superior accuracy rates for results drawn on each respective medical datasets have been marked in bold.

Table 11 and Table 12 contain the results for the WDBC and the Texture dataset respectively, with the cosine distance and the Pearson correlation. Results for the four (4) classifiers are displayed. We note that results obtained by the filter methods with IG measure for WDBC is improved by 3.515% (96.661% v. 93.146%) for NB classifier.

Additionally, when adopting the proposed approach, an improvement of more than 1.4% may be seen for the other classifiers. Similarly, for the Texture dataset, an improvement in the accuracy rate for SMO classifier by 2.273% (97.127% v. 94.855%) when RF is applied in the proposed approach.

Table 13 and Table 14 summarize the results obtained for Spectheart and Movement-libras datasets respectively, with the cosine distance and the Pearson correlation. Again, except for the Spectheart dataset with the SMO classifier, results show promising improvements in most cases.

In other cases, an improvement of more than 6% (98.502% v. 91.760%) is obtained via the application of the proposed approach when applying both the RF and MLP classifier. As well, for methods implemented with IG, an improvement of more than 3% (76.405% v. 72.659%) and (98.502% v. 94.008%) may be observed with both the NB and C4.5 classifiers.

As shown in Table 14, for the Movement-libras dataset, a significant improvement in classification accuracy can be observed. Specifically, when applying either the RF and/or NB classifier on the Movement-libras dataset, the accuracy rate is observed to jump by 36% (76.111% v. 39.444%). Relatively impressive improvement can also be observed with the SMO Classifier via OneR application, resulting in a notable improvement of more than 10% (76.944% v. 66.667%).

In previous experiments, the results have only been compared with the simple filter method results. In the current research, a comparative study of results obtained by the proposed method v. other existing and/or emerging ones, as noted in the extant literature, has been conducted. Such methods include the wrapper method (Best-First + WrapperSubSetEval); FCBF + SymSetEval; PSO search with Correlation Feature Selection or CFS; and Best-First with CFS.

Notably, SymSetEval refers to Symmetrical Uncertainty Attribute Set Evaluator, which evaluates the worth of a set of attributes by measuring the symmetrical uncertainty with respect to another set of attributes. As for CFS or Correlation-based Feature Subset Selection, the method evaluates the

Table 11. Summary of classification results for the WDBC dataset

WDBC	Cosine + Centroid					Pearson + Centroid				
Method	NoF	NB	SMO	MLP	C4.5	NoF	NB	SMO	MLP	C4.5
IG	5	94.3761	94.5518	95.0791	95.7821	6	94.5518	94.0246	96.8366	95.7821
	8	93.1459	94.0246	97.0123	95.7821	9	94.2004	94.2004	96.8366	98.594
	10	92.9701	94.3761	96.8366	99.297	10	92.9701	94.3761	96.8366	99.297
	14	93.4974	93.8489	97.188	99.4728	16	93.4974	94.2004	98.2425	99.297
IG+clus	5	<b>95.2548</b>	<b>95.2548</b>	<b>96.4851</b>	<b>98.0668</b>	6	<b>95.6063</b>	<b>96.1336</b>	<b>97.3638</b>	<b>98.2425</b>
	8	<b>96.6608</b>	<b>96.8366</b>	<b>98.4183</b>	<b>98.9455</b>	9	<b>95.4306</b>	<b>96.4851</b>	<b>98.2425</b>	98.2425
	10	<b>95.0791</b>	<b>97.3638</b>	<b>98.594</b>	99.297	10	<b>94.5518</b>	<b>96.3093</b>	<b>98.594</b>	98.2425
	14	<b>94.9033</b>	<b>96.8366</b>	<b>98.7698</b>	99.1213	16	<b>94.7276</b>	<b>96.6608</b>	98.2425	99.1213
GR	5	94.3761	94.5518	95.0791	95.7821	6	94.5518	94.0246	96.8366	95.7821
	8	94.2004	94.2004	96.6608	96.4851	9	94.2004	94.2004	96.8066	96.4851
	10	93.8489	94.3761	96.8366	97.5395	10	93.8489	94.3761	96.8366	97.5395
	14	93.3216	94.0246	98.0668	99.297	16	93.4974	96.6608	98.9455	99.297
GR+clus	5	94.3761	93.8489	<b>95.4306</b>	<b>97.3638</b>	6	<b>95.7821</b>	<b>95.9578</b>	<b>97.891</b>	<b>98.4183</b>
	8	<b>96.3093</b>	<b>96.6608</b>	<b>98.0668</b>	<b>98.9455</b>	9	<b>95.0791</b>	<b>96.6608</b>	<b>98.2425</b>	<b>98.2425</b>
	10	<b>95.0791</b>	<b>96.8366</b>	<b>98.4183</b>	<b>99.1213</b>	10	<b>94.2004</b>	<b>96.6608</b>	96.3093	<b>98.2425</b>
	14	<b>94.3761</b>	<b>97.5395</b>	<b>98.4183</b>	99.1213	16	<b>95.0791</b>	<b>97.5395</b>	98.2425	99.1213
RF	5	95.4306	96.3093	97.0123	97.891	6	94.7276	96.4851	97.0123	97.891
	8	94.7276	96.4851	98.594	98.4183	9	94.0246	96.4851	98.2425	98.4183
	10	94.3761	96.8366	99.1213	98.4183	10	94.3761	96.8366	99.1213	98.4183
	14	94.3761	97.3638	98.9455	98.594	16	94.0246	97.3638	98.4183	99.297
RF+clus	5	93.4974	95.2548	<b>96.3093</b>	97.3638	6	<b>95.6063</b>	<b>96.6608</b>	<b>98.594</b>	<b>98.594</b>
	8	<b>95.0791</b>	<b>96.8366</b>	98.594	<b>98.594</b>	9	<b>96.1336</b>	<b>96.6608</b>	<b>98.4183</b>	<b>98.594</b>
	10	<b>94.9033</b>	<b>97.0123</b>	98.9455	<b>98.594</b>	10	<b>95.0791</b>	<b>98.2425</b>	97.0123	<b>98.594</b>
	14	<b>95.0791</b>	97.0123	98.594	<b>98.9455</b>	16	<b>94.9033</b>	<b>97.7153</b>	<b>98.594</b>	98.9455
OneR	5	94.3761	94.5518	95.0791	95.7821	6	94.5518	94.0246	96.6608	95.7821
	8	94.3761	94.2004	96.8366	98.594	9	94.2004	94.2004	96.6608	98.594
	10	92.9701	94.3761	97.0123	99.297	10	92.9701	94.3761	97.0123	99.297
	14	93.3216	94.0246	98.0668	99.297	16	92.9701	94.7276	98.2425	99.1213
OneR+clus	5	94.0246	<b>95.2548</b>	<b>96.8366</b>	<b>97.188</b>	6	<b>95.0791</b>	<b>96.4851</b>	<b>97.891</b>	<b>98.2425</b>
	8	<b>95.4306</b>	<b>96.6608</b>	<b>98.594</b>	98.2425	9	<b>95.6063</b>	<b>96.6608</b>	<b>98.2425</b>	98.2425
	10	<b>95.0791</b>	<b>97.3638</b>	<b>98.7698</b>	99.297	10	<b>94.2004</b>	<b>96.6608</b>	<b>98.7698</b>	98.2425
	14	<b>94.3761</b>	<b>97.3638</b>	98.0668	99.1213	16	<b>94.3761</b>	<b>97.3638</b>	<b>98.4183</b>	99.1213

worth of a subset of attributes by considering the individual predictability of each feature along with the degree of redundancy between them.

Table 15 summarizes the results obtained from implementing the proposed method as compared with results obtained by existing methods drawn from the extant literature. We highlight best results

Table 12. Summary of classification results for the texture dataset

Texture	Cosine + Centroid					Pearson + Centroid				
Method	NoF	NB	SMO	MLP	C4.5	NoF	NB	SMO	MLP	C4.5
IG	6	76.3091	85.9273	90.3636	95.9636	4	73.6727	80.2545	82.0909	91.4727
	10	80.0364	93.3455	98.2182	98.3455	9	78.4	91.8182	97.7273	97.7818
	13	79.5818	93.7455	99.3636	98.5273	13	79.5818	93.7455	99.3636	98.5273
	20	79.8909	97.2182	99.8545	98.6909	16	79.2909	95.2182	99.7818	98.7636
IG+clus	6	74.0909	85.0727	87.6727	95.7273	4	<b>75.0727</b>	<b>81.4</b>	<b>83.3091</b>	<b>92.7273</b>
	10	76.1273	90.5636	96.2182	97.4545	9	77.9091	<b>92.5818</b>	96.8364	<b>97.8364</b>
	13	<b>80.2545</b>	<b>96.3455</b>	99.3636	<b>98.8182</b>	13	76.9091	<b>95.5636</b>	<b>99.4364</b>	98.5091
	20	78.5455	<b>98.0727</b>	<b>99.9273</b>	<b>98.8182</b>	16	78.3636	<b>97.0545</b>	<b>99.8909</b>	98.5636
GR	6	81	90.8727	93.0182	97.6909	4	75.9818	84.6545	85.2545	92.8
	10	80.3273	94.0727	97.9091	98.2727	9	82.6545	93.4545	97.1091	98.1091
	13	81.9091	94.7273	99.1455	98.4182	13	81.9091	94.7273	99.1455	98.4182
	20	78.5455	96.7455	99.9636	98.8182	16	81.7091	96.1818	99.6727	98.6364
GR+clus	6	74.7455	86.8182	89.1818	96.2	4	73.8182	79.3636	78.7636	91.2727
	10	77.0727	92.3455	95.5091	97.8364	9	80.6545	<b>94.2909</b>	<b>97.5273</b>	<b>98.3455</b>
	13	80.7091	<b>96.9636</b>	<b>99.4909</b>	<b>98.6727</b>	13	79.7636	<b>95.6364</b>	<b>99.5636</b>	<b>98.6727</b>
	20	78.5455	<b>98.0727</b>	99.9273	98.8182	16	78.7455	<b>97.0909</b>	<b>99.8545</b>	<b>98.7273</b>
RF	6	77.8727	85.2727	88.7455	95.2909	4	73.5818	78.5273	80.5273	89.3091
	10	79.2364	91.4909	96.4909	97.6545	9	78.6545	90.2727	95.1455	97.2909
	13	79.3455	94.9636	99.5091	98.4	13	79.3455	94.9636	99.5091	98.2727
	20	80.5636	98	99.9636	98.7273	16	78.9455	94.8545	99.6545	98.4364
RF+clus	6	75.4364	<b>85.5818</b>	<b>91.8727</b>	<b>95.6545</b>	4	<b>77.4364</b>	<b>82.6727</b>	<b>84.0727</b>	<b>92.8364</b>
	10	77.5455	<b>94.3455</b>	<b>98.0909</b>	<b>98.0182</b>	9	78.5091	<b>93.0909</b>	<b>97.2364</b>	<b>97.9336</b>
	13	<b>80.6364</b>	<b>96.4182</b>	99.3273	<b>98.8545</b>	13	77.7273	<b>95.9091</b>	99.4364	<b>98.4727</b>
	20	79.1818	<b>98.3818</b>	99.9273	<b>98.7818</b>	16	<b>79.0364</b>	<b>97.1273</b>	<b>99.8909</b>	<b>98.8727</b>
OneR	6	73.3636	84.9455	91.6364	95.6545	4	65.7091	77.6364	81.2545	89.3636
	10	77.5273	93.0909	98.0909	98.0182	9	77.3636	92.2364	97.4182	97.7455
	13	78.9455	94.2545	99.0364	98.4182	13	78.9455	94.2545	99.0364	98.3636
	20	79.9273	96.6182	99.9818	98.7455	16	79.1818	95.1818	99.8	98.5091
OneR+clus	6	<b>76.0909</b>	<b>86.6364</b>	<b>92.7818</b>	95.5273	4	<b>75.9818</b>	<b>82.9273</b>	<b>85.0727</b>	<b>93.5091</b>
	10	<b>78.6</b>	<b>95.4182</b>	<b>98.7091</b>	98.0182	9	<b>78.5818</b>	<b>93.6545</b>	96.8364	<b>98.1091</b>
	13	<b>79.4545</b>	<b>96.7273</b>	98.9636	<b>98.4909</b>	13	77.4364	<b>95.4545</b>	98.8	<b>98.4909</b>
	20	78.0727	<b>97.8</b>	99.8909	<b>99</b>	16	78.1636	<b>96.5636</b>	99.6727	<b>98.7818</b>

obtained as tabulated. Note that Best-First + SubsetEval, FCBF+ SymmetricalEval, PSO + CFS and CFS + Best-First feature selection methods used in the table automatically select the number of features. Thus, in each case, we provide the accuracy rate obtained by the proposed approach based on the number of selected features and the observed performance accuracy.

Table 13. Summary of classification results for Spectheart dataset

Spectheart	Cosine + Centroid					Pearson + Centroid					
	Method	NoF	NB	SMO	MLP	C4.5	NoF	NB	SMO	MLP	C4.5
IG	6	72.6592	79.4007	82.7715	89.8876	11	68.1648	79.4007	86.5169	90.6367	
	11	68.1648	79.4007	86.5159	90.6367	15	70.412	79.4007	89.5131	94.0075	
	15	70.412	79.4007	89.5131	94.0075	18	70.7865	79.4007	93.633	94.0075	
	22	70.7865	79.4007	91.0112	98.1273	22	70.7865	79.4007	91.0112	98.1273	
IG+clus	6	<b>76.4045</b>	79.4007	<b>84.6442</b>	85.0187	11	<b>76.4045</b>	79.4007	<b>90.2622</b>	<b>94.7566</b>	
	11	<b>73.7828</b>	79.4007	<b>89.1386</b>	<b>96.2547</b>	15	<b>75.6554</b>	79.4007	<b>93.2584</b>	<b>97.7528</b>	
	15	<b>74.9064</b>	79.4007	<b>89.8876</b>	<b>98.5019</b>	18	<b>73.0337</b>	79.4007	<b>96.2547</b>	<b>98.1273</b>	
	22	<b>71.5356</b>	79.4007	<b>97.0037</b>	<b>98.5019</b>	22	<b>71.5356</b>	79.4007	<b>97.3783</b>	97.7528	
GR	6	70.7865	79.4007	82.7715	86.1423	11	69.6629	79.4007	84.2697	88.3895	
	11	69.6629	79.4007	84.2697	88.3895	15	67.7903	79.4007	87.6404	96.2547	
	15	67.7903	79.4007	87.6404	96.2547	18	69.2884	79.4007	93.633	98.1273	
	22	71.9101	79.4007	96.6292	98.1273	22	71.9101	79.4007	96.6292	98.1273	
GR+clus	6	<b>74.1573</b>	79.4007	<b>84.6442</b>	82.7715	11	<b>74.5318</b>	79.4007	<b>91.0112</b>	<b>94.7566</b>	
	11	<b>73.7828</b>	79.4007	<b>89.1386</b>	<b>96.6292</b>	15	<b>74.9064</b>	79.4007	<b>92.1348</b>	<b>97.3783</b>	
	15	<b>74.1573</b>	79.4007	<b>92.8839</b>	<b>97.7528</b>	18	<b>72.6592</b>	79.4007	<b>95.1311</b>	98.1273	
	22	<b>71.5356</b>	79.4007	<b>97.0037</b>	<b>98.5019</b>	22	71.5356	79.4007	<b>97.3783</b>	97.7528	
RF	6	65.1685	79.4007	79.4007	81.2734	11	67.7903	79.4007	85.7678	90.6367	
	11	67.7903	79.4007	85.7678	90.6367	15	67.0412	79.4007	87.2659	95.5056	
	15	67.0412	79.4007	87.2659	95.5056	18	67.4157	79.4007	89.5131	97.0037	
	22	66.6667	79.4007	91.7603	98.1273	22	66.6667	79.4007	91.7603	98.1273	
RF+clus	6	<b>74.9064</b>	79.4007	<b>83.5206</b>	<b>82.397</b>	11	<b>73.4082</b>	79.4007	<b>87.6404</b>	<b>95.5056</b>	
	11	<b>73.4082</b>	79.4007	<b>91.7603</b>	88.764	15	<b>73.0337</b>	79.4007	<b>93.633</b>	91.7603	
	15	<b>74.5318</b>	79.4007	<b>93.2584</b>	<b>97.7528</b>	18	<b>71.9101</b>	79.4007	<b>94.382</b>	<b>97.7528</b>	
	22	<b>71.5356</b>	79.4007	<b>98.5019</b>	98.1273	22	<b>71.5356</b>	79.4007	<b>96.6292</b>	86.5169	
OneR	6	61.4232	79.4007	80.1498	88.015	11	61.4232	79.4007	86.8914	92.1348	
	11	61.4232	79.4007	86.8914	92.1348	15	62.1723	79.4007	88.764	97.7528	
	15	62.1723	79.4007	88.764	97.7528	18	62.1723	79.4007	93.633	97.7528	
	22	62.5468	79.4007	94.7566	97.7528	22	62.5468	79.4007	94.7566	97.7528	
OneR+clus	6	<b>62.1723</b>	79.4007	<b>81.2734</b>	83.5206	11	<b>63.6704</b>	79.4007	<b>91.3858</b>	83.5206	
	11	<b>63.6704</b>	79.4007	<b>91.3858</b>	83.5206	15	61.0487	79.4007	<b>91.7603</b>	92.5094	
	15	<b>65.5431</b>	79.4007	<b>93.633</b>	81.2734	18	<b>63.6704</b>	79.4007	<b>95.5056</b>	97.3783	
	22	<b>65.5431</b>	79.4007	<b>96.2547</b>	92.8839	22	<b>65.5431</b>	79.4007	93.633	97.7528	

From Table 15, it may be argued that the proposed approach generally yields improved accuracy for classifying medical datasets; more specifically, nine cases among all of the applied classifiers and feature selection methods have been noted to achieve superior classification accuracy performance. Notwithstanding, in other cases, some combination of feature selection methods and classifiers such as the wrapper methods and various classifiers for the Textual dataset can be seen to be superior to the proposed approach.

Table 14. Summary of classification results for the movement-libras dataset

Movlibras	Cosine + Centroid					Pearson + Centroid				
Method	NoF	NB	SMO	MLP	C4.5	NoF	NB	SMO	MLP	C4.5
IG	12	33.6111	35	62.2222	81.6667	6	33.0556	27.7778	56.3889	80.2778
	23	53.8889	56.6667	85.2778	89.4444	15	39.4444	40.8333	67.5	83.0556
	30	56.3889	61.1111	87.5	90.2778	26	55	58.6111	86.3889	89.7222
	45	68.6111	75.5556	92.7778	92.7778	30	56.3889	61.1111	87.5	90.2778
IG+clus	12	<b>72.5</b>	<b>68.8889</b>	<b>88.6111</b>	<b>92.2222</b>	6	<b>65.2778</b>	<b>60.2778</b>	<b>77.7778</b>	<b>88.8889</b>
	23	<b>75</b>	<b>72.2222</b>	<b>90.5556</b>	<b>93.3333</b>	15	<b>72.2222</b>	<b>70.8333</b>	<b>89.4444</b>	<b>92.2222</b>
	30	<b>73.0556</b>	<b>75.2778</b>	<b>90.2778</b>	<b>93.8889</b>	26	<b>76.1111</b>	<b>76.1111</b>	<b>92.5</b>	<b>92.7778</b>
	45	<b>75.8333</b>	<b>81.1111</b>	89.7222	<b>93.3333</b>	30	<b>74.1667</b>	<b>76.6667</b>	<b>91.1111</b>	<b>91.9444</b>
GR	12	38.0556	36.9444	67.5	82.7778	6	35	32.2222	54.1667	80
	23	51.3889	60	84.4444	91.3889	15	40.5556	40	70.2778	87.2222
	30	58.0556	60.5556	87.5	90.2778	26	55.5556	61.1111	85.2778	90.5556
	45	66.6667	73.8889	92.2222	93.0556	30	58.0556	60.5556	87.5	90.2778
GR+clus	12	<b>69.7222</b>	<b>68.3333</b>	<b>87.7778</b>	<b>94.1667</b>	6	<b>57.5</b>	<b>52.5</b>	<b>77.2222</b>	<b>86.3889</b>
	23	<b>73.0556</b>	<b>72.5</b>	<b>90.5556</b>	<b>92.5</b>	15	<b>71.3889</b>	<b>71.1111</b>	<b>90.8333</b>	<b>93.8889</b>
	30	<b>71.1111</b>	<b>74.7222</b>	<b>90.2778</b>	<b>93.8889</b>	26	<b>75</b>	<b>74.1667</b>	<b>90.5556</b>	<b>93.0556</b>
	45	<b>76.1111</b>	<b>80.8333</b>	89.7222	92.5	30	<b>72.7778</b>	<b>76.1111</b>	<b>91.9444</b>	<b>93.3333</b>
RF	12	30.5556	34.1667	59.1667	83.8889	6	21.6667	21.6667	28.3333	71.1111
	23	54.1667	64.7222	89.1667	90.8333	15	39.4444	38.0556	68.0556	83.0556
	30	61.3889	67.5	88.3333	91.3889	26	58.3333	61.6667	88.3333	91.3889
	45	64.1667	74.7222	91.9444	93.3333	30	61.3889	67.5	88.3333	91.3889
RF+clus	12	<b>74.1667</b>	<b>66.1111</b>	<b>88.6111</b>	<b>92.2222</b>	6	<b>60</b>	<b>56.1111</b>	<b>76.9444</b>	<b>87.7778</b>
	23	<b>74.7222</b>	<b>75</b>	<b>90.5556</b>	<b>94.1667</b>	15	<b>76.1111</b>	<b>70.5556</b>	<b>89.4444</b>	<b>93.0556</b>
	30	<b>72.7778</b>	<b>75.8333</b>	<b>91.1111</b>	<b>94.7222</b>	26	<b>74.4444</b>	<b>76.3889</b>	<b>90</b>	<b>93.8889</b>
	45	<b>75.5556</b>	<b>81.9444</b>	<b>91.6667</b>	<b>94.1667</b>	30	<b>74.1667</b>	<b>76.9444</b>	<b>89.4444</b>	<b>94.1667</b>
OneR	12	47.2222	54.1667	85.5556	90.2778	6	31.9444	30	55.5556	81.1111
	23	60.5556	68.6111	88.6111	92.2222	15	55.5556	58.8889	86.3889	91.6667
	30	62.5	70.8333	90.8333	91.9444	26	59.7222	66.6667	90	92.2222
	45	68.0556	79.4444	94.4444	93.3333	30	62.5	70.8333	90.8333	91.9444
OneR+clus	12	<b>71.9444</b>	<b>70</b>	<b>87.2222</b>	<b>91.1111</b>	6	<b>57.2222</b>	<b>53.0556</b>	<b>81.3889</b>	<b>91.3889</b>
	23	<b>74.1667</b>	<b>76.1111</b>	<b>92.2222</b>	<b>92.7778</b>	15	<b>71.6667</b>	<b>67.7778</b>	<b>92.5</b>	<b>93.3333</b>
	30	<b>71.9444</b>	<b>76.3889</b>	90.8333	<b>93.6111</b>	26	<b>73.8889</b>	<b>76.9444</b>	<b>90.8333</b>	<b>93.3333</b>
	45	<b>74.1667</b>	<b>80.5556</b>	89.7222	92.7777	30	<b>73.0556</b>	<b>76.1111</b>	89.4444	<b>93.6111</b>

## 5. CONCLUSION

In this paper, we proposed a novel hybrid feature selection approach based on an integration of commonly deployed filter methods and the hierarchical agglomerative clustering or HAC method. The primary objective of adopting this unique approach is to augment results typically obtained with just applying the filter methods; more importantly, the underlying rationale is to eliminate redundant and irrelevant features effectively and meaningfully with the intelligent clustering of variables.



Table 15. A comparative study of the proposed approach with various feature selection methods

Dataset		FS Method				
		Wrapper method	FCBF+ SymSetEval	PSO+CFS	BestFirst+CFS	Proposed Approach
WDBC	NB	94.3761% 6 features	95.2548% 7 features	94.2004% 11 features	95.2548% 10 features	<b>95.7821%</b> <b>6 features</b>
	SMO	96.4851% 6 features	96.3093% 7 features	96.6608% 11 features	96.6608% 10 features	<b>96.6608%</b> <b>6 features</b>
	MLP	97.7153% 6 features	98.4183% 7 features	98.594% 11 features	98.594% 10 features	<b>98.594%</b> <b>6 features</b>
	C4.5	98.9455% 6 features	98.2425% 7 features	99.1213% 11 features	98.9455% 10 features	<b>99.297%</b> <b>10 features</b>
Texture	NB	77.5455% 22 features	82.6364% 7 features	82.6545% 14 features	<b>83.2545%</b> <b>10 features</b>	79.188% 20 features
	SMO	<b>98.4545%</b> <b>22 features</b>	92.2909% 7 features	97.4727% 14 features	95.8% 10 features	97.1273% 16 features
	MLP	<b>99.9455%</b> <b>22 features</b>	96.0364% 7 features	99.6727% 14 features	98.2727% 10 features	98.2727% 10 features
	C4.5	<b>98.8727%</b> <b>22 features</b>	98.1273% 7 features	98.7273% 14 features	98.5636% 10 features	98.8545% 13 features
Spectheart	NB	74.5318% 3 features	72.6592% 7 features	71.1611% 20 features	71.9101% 21 features	<b>76.4045%</b> <b>6 features</b>
	SMO	<b>79.4007%</b> <b>3 features</b>	79.4007% 7 features	79.4007% 20 features	79.4007% 21 features	79.4007% features
	MLP	81.2734% 3 features	86.1423% 7 features	93.633% 20 features	<b>95.8801%</b> <b>21 features</b>	95.5056% 18 features
	C4.5	81.2734% 3 features	89.1386% 7 features	97.3783% 20 features	98.1273% 21 features	<b>98.1273%</b> <b>18 features</b>
Movement-libras	NB	61.9444% 11 features	74.4444% 10 features	70.8333% 37 features	72.2222% 26 features	<b>76.1111%</b> <b>15 features</b>
	SMO	63.6111% 11 features	69.7222% 10 features	80.0000% 37 features	77.2222% 26 features	<b>96.94444%</b> <b>26 features</b>
	MLP	88.611% 11 features	90.2778% 10 features	90.2778% 37 features	<b>94.1667%</b> <b>26 features</b>	92.2222% 23 features
	C4.5	93.0556% 11 features	91.9444% 10 features	93.6111% 37 features	92.2222% 26 features	<b>94.1667%</b> <b>23 features</b>

In this study, four relevance measurements: IG, GR, RF and OneR, and two similarity measures (Cosine distance and the Pearson correlation) have been chosen to evaluate and compare among the results of the variously applied classification models. The purpose is to evaluate if the classification accuracy using NB, SMO, MLP and C4.5 classifiers may be enhanced using a hybridization of filter methods with agglomerative clustering method. Results have shown a positive impact of the clustering technique onto application of the filtering methods with the proposed approach providing a superior classification for all data generally. However, we note that IG+Clustering dominates with better performance.

## 5.1 Limitations

The strength of our work is the use of a variety of similarity and relevance measurements to have clear and valid results about the impact of clustering in filter results. However, the proposed approach can cause high computation time especially for high dimensional datasets.

A key limitation of our research is to address the question of how to automatically select the number of selected features; moreover, a further limitation is the challenge to validate evolving data analytics environments in the real world as increasingly massive amounts of diverse formatted data, information and knowledge have evolved on various medical conditions over the prior and coming decades.

## 5.2 Feature Work

Our vision for future research is to extend the proposed novel approach and applications in multiple ways, for example, by developing an effective way to automatically select the optimal number of features, promoting a greater integration of nature-inspired feature selection methods with classical-based filter methods, and the applications of these data analytic methodologies across many more real-world domains in which databases may have already been accumulated in the context of presumably secured data infrastructural environments, which have evolved from advances in artificial intelligence and other emerging computational technologies.

## REFERENCES

- Aggarwal, C. C. (2014). *Data classification: algorithms and applications*. CRC Press. doi:10.1201/b17320
- Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, 36(2), 3240–3247. doi:10.1016/j.eswa.2008.01.009
- Anbarasi, M., Anupriya, E., & Iyengar, N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, 2(10), 5370–5376.
- Asir, D., Appavu, S., & Jebamalar, E. (2016). Literature Review on Feature Selection Methods for High-Dimensional Data. *International Journal of Computers and Applications*, 136(1), 9–17. doi:10.5120/ijca2016908317
- Berthold, M. R., & Höppner, F. (2016). *On clustering time series using euclidean distance and pearson correlation*. arXiv preprint arXiv:1601.02213
- Bora, M., Jyoti, D., Gupta, D., & Kumar, A. (2014). *Effect of different distance measures on the performance of K-means algorithm: an experimental study in Matlab*. arXiv preprint arXiv:1405.7471
- Butterworth, R., Piatetsky-Shapiro, G., & Simovici, D. A. (2005). *On feature selection through clustering*. Paper presented at the Data Mining, Fifth International Conference on IEEE. doi:10.1109/ICDM.2005.106
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16–28. doi:10.1016/j.compeleceng.2013.11.024
- Cheng, Q., Zhou, H., & Cheng, J. (2010). The fisher-markov selector: Fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(6), 1217–1233. doi:10.1109/TPAMI.2010.195 PMID:21493968
- Chormunge, S., & Jena, S. (2018). Correlation based feature selection with clustering for high dimensional data. *Journal of Electrical Systems and Information Technology*.
- Dua, D., & Efi, K. T. (2017). *UCI Machine Learning Repository*. <http://archive.ics.uci.edu/ml>
- Hong, T.-P., Liou, Y.-L., Wang, S.-L., & Vo, B. (2014). Feature selection and replacement by clustering attributes. *Vietnam Journal of Computer Science*, 1(1), 47–55. doi:10.1007/s40595-013-0004-3
- Ibrahim, M. A., Ojo, O. A., & Oluwafisoye, P. A. (2018). On feature selection methods for accurate classification and analysis of emphysema ct images. *Int. J. Med. Imaging*, 5(6), 70. doi:10.11648/j.ijmi.20170506.12
- Ienco, D., & Meo, R. (2008). Exploration and reduction of the feature space by hierarchical clustering. *Proceedings of the 2008 SIAM International Conference on Data Mining*. doi:10.1137/1.9781611972788.53
- Inbarani, H. H., Azar, A. T., & Jothi, G. (2014). Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. *Computer Methods and Programs in Biomedicine*, 113(1), 175–185. doi:10.1016/j.cmpb.2013.10.007 PMID:24210167
- Irani, J., Pise, N., & Phatak, M. (2016). Clustering techniques and the similarity measures used in clustering: A survey. *International Journal of Computers and Applications*, 134(7).
- Karabatak, M. (2015). A new classifier for breast cancer detection based on Naïve Bayesian. *Measurement*, 72, 32–36. doi:10.1016/j.measurement.2015.04.028
- Karegowda, A. G., Manjunath, A., & Jayaram, M. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, 2(2), 271–277.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273–324. doi:10.1016/S0004-3702(97)00043-X
- Kumar, G. R., Ramachandra, G., & Nagamani, K. (2014). An efficient feature selection system to integrating svm with genetic algorithm for large medical datasets. *International Journal (Toronto, Ont.)*, 4(2), 272–277.
- Lavanya, D., & Rani, D. K. U. (2011). Analysis of feature selection with classification: Breast cancer datasets. *Indian Journal of Computer Science and Engineering*, 2(5), 756–763.

- Mwadulo, M. W. (2016). A Review on Feature Selection Methods For Classification Tasks. *International Journal of Computer Applications Technology and Research*, 5(6), 395–402. doi:10.7753/IJCATR0506.1013
- Nagendrudu, S., & Reddy, V. R. (2015). Enhanced Clustering of High Dimensional Data Using Fast Cluster Based Feature Selection. *International Journal of Science. Engineering and Computer Technology*, 5(5), 113.
- Novaković, J. (2016). Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research*, 21(1).
- Park, C. H. (2013). *A feature selection method using hierarchical clustering*. In *Mining Intelligence and Knowledge Exploration*. Springer.
- Peng, C., Cheng, J., & Cheng, Q. (2017). A supervised learning model for high-dimensional and large-scale data. *ACM Transactions on Intelligent Systems and Technology*, 8(2), 30. doi:10.1145/2972957
- Peng, C., & Cheng, Q. (2019). *Discriminative Regression Machine: A Classifier for High-Dimensional Data or Imbalanced Data*. arXiv preprint arXiv:1904.07496
- Polat, K., & Güneş, S. (2009). A new feature selection method on classification of medical datasets: Kernel F-score feature selection. *Expert Systems with Applications*, 36(7), 10367–10373. doi:10.1016/j.eswa.2009.01.041
- Radovic, M., Ghalwash, M., Filipovic, N., & Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 9. doi:10.1186/s12859-016-1423-9 PMID:28049413
- Saeyns, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19), 2507–2517.
- Shilaskar, S., & Ghatol, A. (2013). Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. *Expert Systems with Applications*, 40(10), 4146–4153. doi:10.1016/j.eswa.2013.01.032
- Sudhakar, C., Priya, C. S., Chandini, S., & Narasimham, C. (2016). Cluster Based Feature Subset Selection (CFSS) for High-Dimensional Data. *International Journal of Applied Engineering Research*, 11(2), 1369–1372.
- Xue, B., Zhang, M., & Browne, W. N. (2012). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics*, 43(6), 1656–1671. doi:10.1109/TSMCB.2012.2227469 PMID:24273143
- Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4), 606–626. doi:10.1109/TEVC.2015.2504420
- Yildirim, P. (2015). Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease. *International Journal of Machine Learning and Computing*, 5(4), 258–263. doi:10.7763/IJMLC.2015.V5.517
- Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. *Proceedings of the 20th international conference on machine learning (ICML-03)*.
- Zhang, Q., Wang, H., & Yoon, S. W. (2017). A Hierarchical Feature Selection Model using Clustering and Recursive Elimination Methods. *Paper presented at the IIE Annual Conference Proceedings*.
- Zhang, W., Zhao, D., & Wang, X. (2013). Agglomerative clustering via maximum incremental path integral. *Pattern Recognition*, 46(11), 3056–3065. doi:10.1016/j.patcog.2013.04.013

*Nadja Elong is fourth-year PhD student in computer engineering. Her main research interest centers on how to reduce the number of features to enhance classification accuracy. She is also concerned about improving feature selection methods. For her doctoral thesis, she wanted to ameliorate filter method results for feature selection, by the use of well-known clustering method. She holds master's degree in computer engineering from the university of Ibn Khaldoun, Tiaret, Algeria.*

*Rahal Sid Ahmed is a Professor in the Department of Computer Science at university of sciences and technology of Oran, Algeria. He is member of SSD laboratory (Signal, System and Data). His research interests are: Object-Oriented, Data Mining, Agents and Expert Systems. He published many research papers, and he participates in many conferences.*