International Journal of Data and Network Science 8 (2024) 1117–1128

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Hybrid feature selection based ScC and forward selection methods

Luai Al-Shalabi[a*]

[a]*Faculty of Computer Studies, Arab Open University, Ardiya Industrial Area, 92400, Kuwait*

C H R O N I C L E

Article history:
Received: August 15, 2023
Received in revised format: September 25, 2023
Accepted: November 26, 2023
Available online: November 26, 2023

Keywords:
Feature Selection
ScC
Forward Selection
Machine Learning
Classification

A B S T R A C T

Operational data is always huge. A preprocessing step is needed to prepare such data for the analytical process so the process will be fast. One way is by choosing the most effective features and removing the others. Feature selection algorithms (FSAs) can do that with a variety of accuracy depending on both the nature of the data and the algorithm itself. This inspires researchers to keep on developing new FSAs to give higher accuracies than the existing ones. Moreover, FSAs are essential for reducing the cost and effort of developing information system applications. Merging multiple methodologies may improve the dimensionality reduction rate retaining sensible accuracy. This research proposed a hybrid feature selection algorithm based on ScC and forward selection methods (ScCFS). ScC is based on stability and correlation while forward selection is based on Random Forest (RF) and Information Gain (IG). A lowered subset generated by ScC is fed to the forward selection method which uses the IG as a decision criterion for selecting the attribute to split the node of the RF to generate the optimal reduct. ScCFS was compared to other known FSAs in terms of accuracy, AUC, and F-score using several classification algorithms and several datasets. Results showed that the ScCFS excels other FSAs employed for all classifiers in terms of accuracy except FLM where it comes in second place. This proves that ScCFS is the pioneer in generating the reduced dataset with remaining high accuracies for the classifiers used.

1. Introduction

Feature selection algorithms are essential in machine learning where huge data is needed to solve critical real-life problems. Deciding which features are beneficial in prediction can help experts in the domain overcome serious obstacles by developing new systems for the given problem.

Working with big data is time-consuming and could lead to a low amount of accuracy. Many important algorithms are qualified to minimize the dimensionality of a given dataset such as Rough Set (RS) theory, Weight-Guided (WG), and Stability-correlation and Correlation (ScC). Merging more than one methodology, which is called hybrid, may also improve the performance of the feature selection method to determine the optimal feature subset. The three mentioned feature selection algorithms will be used in this research and their results will be compared with the results of the proposed one to test its efficiency. Ignoring irrelevant features is important to minimize the size of the dataset which after analysis will improve the performance of the data model. This can be examined by different available classification algorithms. Classification algorithms use different classification engines that may give different results for the same real-life problem. For this, it is more applicable to apply several classification algorithms for the same problem and choose the best of them which is not necessary to be the best for all datasets. The following classification algorithms will be used in this research: Fast Large Margin (FLM), Gradient-Boosted Trees (GBT), Logistic Regression (LR), and random forest (Cormen, 2009).

* Corresponding author.
E-mail address: lshalabi@aou.edu.kw (L. Al-Shalabi)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.11.022

Filter, wrapper, and embedded are the available three types of feature selection algorithms. (Das et al., 2018; Al-Shalabi, 2022; Niu et al., 2020; Guyon & Elisseeff, 2003). As explained by Ma and Xia (2017), filter algorithms emphasize the attributes of a dataset that are relevant to the problem itself. The most relevant attributes will be chosen. Such kinds of algorithms are simple, fast, and produce effective reducts. The second type is wrapper algorithms. Such algorithms use classification algorithms to estimate the importance of an attribute. If the accuracy of the reduct is high then that attribute is important and it will be kept, otherwise, it is not and it should be removed. All attributes should be tested one by one to form the final reduct. This repeated training process is classified as an expensive and complex algorithm but efficient. For more details, the reader may refer to (Zhang et al., 2015; Lazer et al., 2012; Kohavi & John, 1997). The third type is embedded algorithms. In this type, the accuracy of attributes is produced by the machine learning model. In the embedded methods, to save computation time, the feature selection process is embedded in the learning process stage such that the development of the machine learning algorithms and the selection of important features are accomplished simultaneously.

Filter methods choose significant features based on statistical techniques. Unlike wrapper methods, filter methods are simple because they are independent of machine learning algorithms that need high computational time for the training process. For this reason, the focus of this research was to use ScC filter algorithm to reduce the size of the given dataset before feeding the reduced subset to the wrapper forward selection to build the optimal solution. ScC will help in reducing the complexity of the forward feature selection, especially for large datasets. FLM, GBT, LR, and RF were used to test the importance of the proposed algorithm for feature selection. To do so, RS, ScC, WG, and the proposed algorithm were used to generate the reduct of the benchmark datasets then the accuracy, F-score, and AUC results were generated for these reducts. A comparison between the results was accomplished which showed that the proposed algorithm is a pioneer.

The main contributions of this study are three-fold:

- A pioneer hybrid feature selection algorithm based on ScC filter method and forward selection wrapper method is proposed considering the speed of a filter method and the relevance search of a wrapper method.
- Statistical analysis of the proposed algorithm using a variety of datasets.
- A comparison between the proposed algorithm and other well-known algorithms.

The rest of this article is structured as follows. Section 2 illustrates related works. Section 3 explains the feature selection methods, classification algorithms, and evaluation methods and metrics used in this study. Section 4 explains the proposed algorithm. Section 5 expresses the results achieved. Section 6 represents the conclusion part of the work.

2. Related Work

In literature, there are many feature selection algorithms, and all of them are datasets-based efficiency. There are no algorithms superior to all kinds of datasets. Comparison between feature selection algorithms is still important to determine the best for the given dataset. There are many measures to do so such as accuracy, recall, precision, AUC, and F-score. Building a classification model for the reduced dataset is important to decide the acceptance of the feature selection method based on its performance. Feature selection in the field of machine learning has been studied by many researchers for a very long time (Cai et al., 2018). Pabuccu and Barbu (2023) investigate the use of feature selection methods for improving the forecasting performance of many common machine learning algorithms. Suresh et al. (2022) investigated different methods of feature selection to improve the performance of a machine-learning model in the recognition of the onset of Autonomic Dysreflexia. Kim et al. (2022) developed an efficient feature selection method using the Reinforcement Learning algorithm by selecting features that are effective for classification in most datasets. Tianyi et al. (2023) stated that filter methods are computationally more efficient than wrapper methods because they do not involve the construction of machine-learning models. They also stated that they are easily adapted to high-dimensional datasets with low computational costs. Al-Shalabi (2022) has proposed a feature selection method (ScC) based on correlation and stability. Kang et al. (2023) identified in their study the most important features needed to improve the recognition of dental caries and to get the most prediction model via a combination of GINI and mRMR techniques. Phyu and Oo (2016) proposed a new feature subset selection algorithm based on a conditional mutual information approach. They concluded that the selected reduct was appropriately important for the learning process. Hoque et al. (2018) built a filter method that totaled the scores of several available filter methods. Many published articles comparing different filter methods based on accuracy measures can be found in the literature (Liu, 2004; Peng et al., 2005; Darshan & Jaidhar, 2018; Bolón-Canedo et al., 2014; Inza et al., 2004). Fleuret (2004) used accuracy and runtime measures to compare different filter methods. Cherrington et al. (2019) studied numerous filter methods based on the idea of ranked scores and how threshold value can affect the results. Pourpanah et al. (2019) proposed a feature selection method called FAM-BSO that combined the Fuzzy ARTMAP model and BSO feature selection method. Sinayobye et al. (2019) proposed a hybrid classification method that joins a correlation-based filter method and machine learning classifiers. Bommert et al. (2020) showed that no filter method constantly bests the other methods. Hu et al. (2021) proposed a sequential forward selection method based on separability (SFSS) that has superior accuracy and low runtime. Filter and wrapper feature selection methods were investigated by Wah et al. (2018) and Xue et al. (2015) using accuracy measures. Other works done by Zhu et al. (2007) and Mohtashami and Eftekhari (2019) compared several wrapper methods based on accuracy. Aphinyanaphongs et al. (2014) compared filter and wrapper methods for text classification datasets. The three types of feature selection (filter, wrapper, and embedded methods) were compared based on the information gain, gain ratio, and correlation performance

(Chaudhary et al., 2013) using the naïve Bayes classifier. The three methods were also studied by Bolón-Canedo et al. (2013) by comparing the accuracy of several classifiers. Al-Shalabi (2019) proposed a dimensionality reduction method for noisy datasets based on grouping the data records into noisy and not-noisy groups and then treating each group separately before merging them again. Haq et al. (2021) combined multiple feature selection methods to generate the best reduct. A scalable algorithm based on Apache Spark cluster (DQPFS) shows important results for big data feature selection was proposed by Soheili and Moghadam (2020). Gong et al. (2022) proposed a hybrid feature selection method based on feature subsets generated by factor analysis. The method generates feature subsets from the maximum load of each feature through factor analysis. Then, both sequential forward selection and minimal redundancy and maximal relevance are employed to remove the redundancy of each feature subset. Xiao et al. (2021) proposed a hybrid feature selection method that generates the set of candidate features by scoring and ranking and then by using a heuristic method to generate the final subset. Yongbin et al. (2023) proposed a hybrid feature selection method based on artificial immune algorithm optimization and metaheuristic-based search strategy to generate effective features from high-dimensional data. Alhenawi et al. (2023) proposed a hybrid feature selection method for microarray data processing that joins an ensemble filter with an Improved Intelligent Water Drop algorithm. Yin et al. (2023) proposed a hybrid feature selection method for multi-class network anomalies using a multilayer perceptron network and a combination of information gain and random forest methods. Other work can be found in (Asghari et al., 2023; Kamalov et al., 2023; Mienye & Sun, 2023).

3. Methods

The preparation for this study starts by making all resources available: the datasets, the state-of-the-art feature selection algorithms, the classifiers that will be used to build the model, and the performance measures that will be considered. All of them are needed to test the performance of the proposed algorithm.

3.1 Data

A collection of nine famous different sizes datasets from several domains was determined and prepared for statistical analysis. Table 1 summarizes them as shown below.

Table 1
Description of the Datasets.

Dataset	# features	# examples
Phishing	30	11055
SpamBase	57	4601
Messidor	19	1151
SGC	20	1000
Austra	14	690
Pop-Failure	20	540
Iono	34	351
Heart Disease	13	270
Sonar	60	208

3.2 Feature Selection Methods

Feature selection methods discover the key features from a given dataset. The reduced dataset usually has high accuracy, the same or higher than the accuracy of the original dataset. Having a few features always improves the efficiency of the algorithm. All feature selection methods used in this research are explained below.

Rough set theory: It is one of the important feature selection algorithms used in the field of data mining and machine learning. It uses the concept of a discernibility matrix to generate a reduced list of features (Velusamy and Manavalan, 2012). The matrix finds both the vital and non-vital features. The theory can generate several reduced lists called reducts. Each list may have a different combination of features and the intersection between these lists is called the core. Let's say that *SF* is a list of all significant features, *CF* is the list of all conditional features, *DS* is the original dataset and γ is the quality of classification, then $\gamma_{SF}(Reduct) = \gamma_{CF}(DS)$ (Velayutham and Thangavel, 2011). Generated reducts are defined as follows:

$$Reduct_{\mathrm{all}} = \{SF \mid SF \subseteq CF, \gamma_{SF}(Reduct) = \gamma_{CF}(DS)\}, \tag{1}$$

The minimal reduct *MR* is in high demand if its quality of classification is also high. It satisfies $MR \subseteq Reduct_{\mathrm{all}}$, such that:

$$Reduct_{\mathrm{all}} = \{SF \mid SF \subseteq Reduct_{\mathrm{all}}, \forall Y \in Reduct_{\mathrm{all}}, |SF| \leq |Y|\}, \tag{2}$$

The theory was used by many researchers for many real-life problems. RS was used in many studies such as predicting the performance of students at a university level (Al-Shalabi, 2016), crime perception (Al-Shalabi, 2017), solving missing data in datasets (Al-Shalabi, 2019), and evaluation of the COVID-19 vaccine (Al-Shalabi, 2022).

Stability-correlation and correlation: It is a feature selection algorithm developed by merging the stability and correlation aspects (Al-Shalabi, 2022) to generate a minimal set of significant features with high accuracy. Let's $DS = (T, CF, V, f)$, where DS is the original dataset, T is a finite set of tuples, CF is a finite set of features, $V = \cup F \in CF$, Vn is a domain of feature F, and f: $T \times CF \rightarrow V$ is a total function such that $f(t,F) \in Vn$ for every $F \in CF$, $t \in T$. Let $\gamma(Red_{\min})$ is the accuracy of the minimal reduct produced by ScC where $Red_{\min} \subseteq CF$ and $\gamma(Red_{\min}) \geq \gamma(DS)$ then $Red_{\min}$ is called the reduct of DS and is denoted by $Red(DS)$.

Weight guided: It is one of the feature selection algorithms used in the field of machine learning. Attributes are given weights, and these weights are used as input to the algorithm. The algorithm then is used to determine the order of features added to the reduction list where the highest weight is added first. The algorithm repeatedly adds weights if there are features still not added or if an addition to the reduction list does not increase the performance (Ringsquandl et al., 2015). Al-Shalabi and Tahhan (2020) have used this method in their work.

Forward Selection (FS): Forward selection is a search strategy to find important features (Reunanen, 2006). It is an iterative wrapper algorithm that starts with having no feature in the subset. In each iteration, one feature is added to the subset that best improves the model. The new subset is trained by a learning method, and the performance measure is calculated. The algorithm keeps iterated until no new feature improves the performance of the model. In this work, RF works well in training datasets and IG is one of the measures used to choose the best-split feature in the RF. Cross-validation is also used for the performance measure of the subset.

3.3 The Classification Methods (CM)

Classification methods are important to build and test a model for a specific real-world classification problem. Different methods build a classification model in different ways, so they give different performances. The performance of the model also depends on the input data type, size, structure, noise, and others and can be evaluated by many measures as explained earlier. LR, FLM, RF, and GBT were used to test the performance of the proposed feature selection method by creating its classification model as well as the other feature selection methods (RS, WG, and ScC). RapidMiner tool, which is a famous tool for data mining was used for this purpose using a 10-fold cross-validation.

Logistic regression: It is a supervised machine-learning method that models the probability of classification labels. It is a popular method for analyzing a given yes/no question and introducing answers to determine whether a dependent variable supports the result. Peng et al. shows that LR has ordinal, binary, binomial, and multinomial types (Peng et al., 2002).

Fast large margin method: This is a popular method for a linear support vector. It can be classified as linear SVM (Nayak et al., 2020). The method is used with massive datasets. It is an SVM-Like method, and it is fast with only O(n) complexity.

Random forest: It was proposed by Breiman (2001) as one of the important methods used to build a collection of tree-based classifiers. It is formed by choosing randomly a small set of input variables to split on. This process will be implemented at each node. It employs regression as well as classification problems. Various decision trees are generated during the training process and a common class of distinct trees is chosen for classification (Al-Shalabi et al, 2006; Quinlan, 1989).

Gradient-boosted tree: This is an advanced decision tree method that enhances the accuracy of sequentially produced trees. The trees are grown using gradient descent optimization. GBT can handle linear and non-linear relationships as well as complex interactions between features (Luekiangkhamla et al., 2023).

3.4 Performance Metrics

The previously introduced ML classification techniques were applied to datasets mentioned in Section 3. During the process of feature selection, the four FSAs mentioned earlier were used for the classification problems. To evaluate and compare the performance of the proposed model, the accuracy, F-score, and AUC metrics were used. Besides that, IG and cross-validation are used in generating and testing the performance of the proposed hybrid method. They are all summarized below.

Accuracy: It evaluates the model of a given reduct. The higher the value is, the more accurate the model and consequently the more useful the reduct. It can be calculated by dividing the number of correct classification tuples (true positives (TP) and true negatives (TN)) by the total number of input tuples (TP, TN, false positives (FP) and false negatives (FN)) as shown in formula 3.

$$\text{Acc} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

AUC: It measures the capability of the classifier to distinguish between positive and negative classes. The higher value of the AUC indicates the superior functioning of the classifier.

F-score: Combines the recall (R) and precision (P) measures. Precision computes the percentage between the positive tuples to all tuples projected as positive. Whereas the recall computes the percentage between the positive tuples to all tuples that should have been projected as positive as shown below.

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$F - score = 2 \times \frac{P \times R}{P + R} \tag{6}$$

Information gain and information gain ratio: Information gain is a mathematical procedure used in machine learning for feature selection. It represents the amount of information acquired during a certain decision. It uses entropy to measure the uncertainty amount which is then used for making decisions to decide about the best split in decision trees and random forest. Entropy explains the amount of information: if it is low then the amount of information is high which gives a better split and vice versa. Mathematically, ***for the given dataset* DS, *the information gain*** of a feature **F** with **c** classes is calculated using Eq. (7).

$$IG(F) = Ent(DS) - \sum_{n=1}^{n} \frac{DS_i}{DS} + Ent(DS_i) \tag{7}$$

where *Ent(DS)* describes the entropy of class labels, DS_i/DS describes the proportion of the number of occurrences of each value on feature *F*, and $Ent(DS_i)$ is the entropy of i^{th} feature value calculated by splitting the dataset *DS* using the feature *F*. The entropy is calculated as in Eq. (8):

$$Ent(DS) = - \sum_{i=1}^{c} p_i \, log_2(p_i) \tag{8}$$

where p_i is the probability of class *i* in the dataset *DS* for *c* classes. If the feature *F* has a high information gain value, then it is significant to the target variable.

The Split Information (SI) value is an integer positive value that explains the possibility of splitting a branch from a node. It represents an intrinsic value that will be used to eliminate the bias in the Information Gain Ratio (IGR) calculation. The SI is calculated as in Eq. (9):

$$SplitInformation(X) = - \sum_{n=1}^{n} \frac{N(t_i)}{N(t)} + log_2 \frac{N(t_i)}{N(t)} \tag{9}$$

where *X* represents a random variable with possible values $x_1, x_2, \ldots, x_i$, and $N(t_i)$ is the number of times that t_i occurs divided by the total number of occurrences *N(t)* where *t* is the set of occurrences. The information gain ratio is the ratio between the information gain and the Split Information value which considers both the IG and the feature's number of results to determine the best feature to split the branch from the node. IGR is calculated as in Eq. (10):

$$IGR(F) = \frac{IG(F)}{SplitInformation(X)} \tag{10}$$

Cross Validation: This method is used to estimate the accuracy of the learned model. The dataset is partitioned into *n* equal subsets. One subset is chosen as the test subset and the remaining *n*-1 subsets are used for training. The cross-validation procedure is repeated *n* times and each of the *n* subsets is employed exactly once as the test data. The average of the *n* results from the *n* iterations is generated to give a single evaluation for the unseen dataset.

4.The Proposed Method

A new hybrid feature selection algorithm based on ScC and forward feature selection algorithms was proposed in this work. The goal is to minimize the dimensions of the original dataset by considering the most relevant features and to enhance the detection of the classification output. To the best of my knowledge, this study is new, and its results are greatly important and comparable to other methods. The proposed method first utilizes the ScC filter-based method that implements both the correlation and the stability for its simplicity and ability to identify the most significant features within the input dataset. Selected features are passed to the forward feature selection wrapper-based method that implements both the IGR and the RF. The subsequent wrapper method was selected for its ability to explore the complex search space of possible feature subsets. Both ScC and forward feature selection techniques have proven their capability to enhance feature selection and classification performance for several combinatorial problems. In the context of this study, the elements of ScC and the forward feature selection methods were chosen. Firstly, the combination of correlation and stability creates a vector with highly significant features that fed to the correlation process again to create the most relevant subset based on specific criteria. In the second part, the forward feature selection method receives the vector of the highest-informative features from the ScC and iteratively selects them one by one. Each feature is evaluated individually using the IGR. In each iteration, RF trains the subset that is initially empty, and IGR makes the decision to keep or reject it. The feature that enhances the model performance when merged with the previously selected features is picked and added to the subset. The process can only stop if no further improvement is observed in the model or when no more features are to be tested. The third part, the optimal subset is evaluated by several classifiers.

Let *DS* be the original dataset $DS = (R, A, D, fun)$, where R is a finite set of records in the original dataset, A is a finite set of attributes, $D = \cup a \in A, V_n$ is a domain of attribute a, and $fun: R \times A \rightarrow D$ is a function such that $fun(r,a) \in D_n$ for every $a \in A$, $r \in R$. Let $\gamma(Subset)$ be the accuracy of the subset generated by ScCFS where $Subset \subseteq A$ and $\gamma(Subset) \geq \gamma(DS)$ then *Subset* is called the optimal subset of *DS* and is denoted by *Subset*(*DS*). The process of the proposed ScCFS algorithm can be outlined through the following steps and is shown in Fig. 1:

1) ScC highly correlated features: the original dataset *DS* will be entered into the ScC algorithm, and each feature will be tested based on its correlation with the target feature and its stability. All features that satisfy a predefined threshold will be added to the reduced subset named *Subset*.
2) Wrapper forward feature selection: building this method by combining the random forest classifier for training the input received from step 1 with the information gain ratio as a criterion for selecting the most informative feature. The process will be iteratively conducted until it discovers optimal subsets of features.
3) Random forest: it uses a rule-based approach and is based on ensemble learning. It improves accuracy by reducing the variance and overfitting. Moreover, it can work with continuous and categorical variables. The random forest was also chosen because it can handle noise in the dataset. It is very efficient in producing a robust subset of features. The only difficulties while using it are its complexity and long training time. The proposed method overcomes these difficulties by reducing the search space by the ScC algorithm.
4) Information gain ratio: It acts as a measure that will evaluate the importance of the feature for splitting the tree by calculating the weight of features with respect to the classification feature.
5) Cross-validation: It is used to measure the performance of each subset in each iteration to accept or reject the feature. It gives the average model of fitness which will be used for predictive performance.
6) Model evaluation: The model's performance is evaluated using LR, FLM, RF, and GBT validation techniques to ensure the generalization fitness of the chosen features.
7) Iterative refinement: the process can be refined, and the methodology continuously iterates and adds new features until it finds the best feature subset.

The following algorithm describes the proposed model:

Algorithm 1: Feature selection
Input: *ODS*
Output: The reduct *Subset*

Let $A=\{A_1, A_2, \ldots, A_n\}$ be the list of attributes of the *ODS*, S is the stability of the attribute and r is the correlation with respect to the predictive attribute, $S1$ and $S2$ are the first and the second subsets where $S2$ is generated from $S1$. Let $F=\{F_1, F_2, \ldots, F_n\}$ be the list of $S2$ features, and let $Per(Sub)$ be the evaluation metric for the iteratively updated subset *Sub*. The performance is given by training the model using RF which uses IGR to split the node. *Subset* is the optimal reduct generated by the proposed algorithm.

```
α=0.35// Stability value
β=0.1// first round correlation value
λ=0.05 // second round correlation value
PP=0 // predictive performance initially is 0
Subset←∅
Sub←∅, S1←∅, S2←∅

for each value Ai in A
        if SAi < α or (SAi ≥ α and rAi ≥ β) then
                S1←S1 ∪ Ai
for each feature Fi in F
    Sub ← Sub ∪ Fi
    if Per(Sub) > PP
        PP←Per(Sub)
        Subset← Sub
    else
        Sub← Sub - Fi
return Subset
```

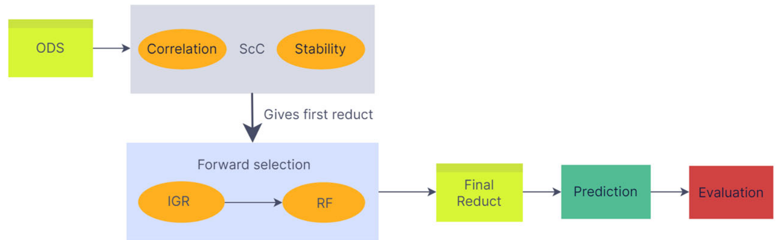

Fig. 1. The framework of the proposed hybrid feature selection model

5. Experimental results

This section explores a comprehensive analysis to present the efficiency of ScCFS. Nine benchmark datasets from different fields and different sizes were used. A 10-fold cross-validation method was employed to train and test these datasets. The ScCFS method was compared against WG, RS, and ScC feature selection methods using LR, FLM, RF, and GBT classifiers in terms of accuracy, reduction rate, AUC, and F-score. The reduct that gives the best evaluation, mainly accuracy, is the preferred one. Table 2 shows the reduction rates of all datasets and their percentages whereas Table 3 shows the accuracies given by the five classification methods and are illustrated in Fig. 2. Table 4 shows the performance of the F-score and is illustrated in Fig. 3 whereas Table 5 shows the result of the AUC, and it is illustrated in Fig. 4. Best-performing results highlighted in bold.

Table 2
The reduction rate of all algorithms used.

		RS		WGFS		ScC		ScCFS	
Dataset	**Original**	**Reduced**	**%**	**Reduced**	**%**	**Reduced**	**%**	**Reduced**	**%**
Austra	14	7	78.6	12	85.7	9	64.3	13	**93.9**
Heart Disease	13	10	**76.9**	7	53.9	8	61.5	10	**76.9**
Phishing	30	7	23.3	29	**96.7**	25	83.3	25	83.3
Sonar	60	38	63.3	50	83.3	39	65	56	**93.3**
iono	34	18	52.9	31	**91.2**	26	76.5	29	85.3
SGC	20	18	90	17	85	19	**95**	19	**95**
SpamBase	57	40	70.2	45	78.95	51	**89.5**	51	**89.5**
Messidor	19	15	79	15	79	16	**84.2**	16	**84.2**
Pop-Failure	20	18	**90**	15	75	18	**90**	18	**90**

Results of the models generated by the classifiers that trained the reducts produced by the ScCFS are highlighted. For austra reduct, the accuracy of FLM was the highest among all other feature selection methods and was also higher than the accuracy of the original dataset. For heart disease reduct, GBT led the accuracy. For phishing reduct, the accuracy of the RF excelled. All the classifiers produced the highest accuracies for sonar, spam, and pop-failure reducts. For iono reduct, the accuracy given by LR and RF was the highest. By training the SGC reduct, all the classifiers produced higher accuracies except for the GBT with a very low difference (0.1%). For messidor reduct, the accuracy is led by RF over other methods.

Table 3
The accuracy of each classifier using the reduct generated by each Feature Selection Method (FSM).

CM/ FSM	**FSM**	**Austra**	**Heart Disease**	**Phishing**	**Sonar**	**Iono**	**SGC**	**SpamBase**	**Messidor**	**Pop-Failure**	**points**	**priority**
LR	Original	0.8531	0.7942	0.9307	0.2788	0.8500	0.7343	0.6164	0.6231	0.9157		
	ScCFS	**0.8682**	0.8717	0.9047	**0.7288**	**0.8600**	**0.7018**	**0.8851**	0.5303	**0.9157**	**6**	**1**
	RS	0.8579	0.6758	**0.9313**	0.5258	0.7800	0.6644	0.5548	**0.6708**	**0.9157**	3	3
	WGFS	0.6953	0.7650	0.5570	0.6621	0.6500	0.6958	0.6768	0.5394	**0.9157**	1	4
	ScC	0.8678	**0.9092**	0.9047	0.7000	**0.8600**	**0.7018**	**0.8851**	0.5303	**0.9157**	5	2
FLM	Original	0.5556	0.8567	0.9307	0.7288	0.8400	0.7018	0.7146	0.7561	0.9222		
	ScCFS	**0.8682**	0.7800	0.9126	**0.7333**	0.7900	**0.7018**	**0.8227**	0.5410	**0.9290**	5	2
	RS	0.6396	0.6542	**0.9310**	0.5409	0.7700	0.6574	0.4247	**0.6522**	0.9157	2	3
	WGFS	0.6396	0.6900	0.5570	0.5500	0.6500	0.6888	0.7688	0.6442	0.9157	0	4
	ScC	0.7819	**0.8700**	0.9126	**0.7333**	**0.8400**	**0.7018**	**0.8227**	0.5410	**0.9290**	**6**	**1**
RF	Original	0.8782	0.7675	0.9253	0.6833	0.9500	0.7204	0.6621	0.5303	0.9157		
	ScCFS	0.8682	0.7950	**0.9256**	**0.7273**	**0.9400**	**0.7018**	**0.8813**	**0.5915**	**0.9157**	**7**	**1**
	RS	0.8579	0.6883	0.9237	0.7136	**0.9400**	**0.7018**	0.7329	0.5793	**0.9157**	3	2
	WGFS	0.7010	0.7167	0.5570	0.5758	0.8500	**0.7018**	0.7123	0.5714	**0.9157**	2	3
	ScC	**0.8733**	**0.8442**	**0.9256**	0.7167	0.9100	**0.7018**	**0.8813**	**0.5915**	**0.9157**	**7**	**1**
GBT	Original	0.8326	0.8050	0.9405	0.7470	0.9000	0.7474	0.6271	0.5305	0.9157		
	ScCFS	**0.8682**	**0.8600**	0.9246	**0.6939**	0.9200	0.7018	**0.6903**	0.6191	**0.9157**	**5**	**1**
	RS	0.8477	0.7033	**0.9332**	0.6833	0.9200	0.7018	0.6058	**0.7121**	**0.9157**	3	3
	WGFS	0.6701	0.7517	0.5570	0.6773	0.8300	**0.7028**	0.6167	0.6707	**0.9157**	2	4
	ScC	**0.8586**	0.8208	0.9246	0.6773	**0.9300**	0.7018	**0.6903**	0.6191	**0.9157**	4	2

Rather than accuracy, the F-score and AUC show the importance of ScCFS through their performance results. As shown in Tables 4 and 5, the best priority is given for each feature selection algorithm based on the number of higher F-score or AUC values as points out of 9 (datasets). In all the classifiers, the score of the F-score put ScCFS in the highest priority. Out of the nine datasets, 7 had the highest values of F-measure for RF and GPT whereas they were 6 for FLM and 4 for LR. The same procedure was applied to AUC values and even accuracy. Table 5 shows that ScCFS led other methods for all classifiers except for RF. Table 3 shows that the proposed method excels other methods for all classifiers except FLM where it comes in second place.

The results of the experiments proved that ScCFS is the pioneer in choosing the best reduct that offers high accuracy, F-measure, and AUC. ScCFS surpassed the other experimented feature selection algorithms mentioned.

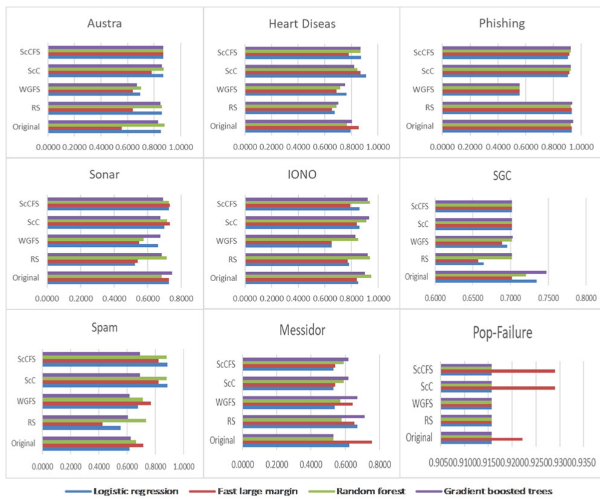

Fig. 2. The accuracy rate of the hybrid approach using all the datasets

Table 4
The F-score rates for the reducts of the nine datasets given by the four classifiers

CM/ FSM	FSM	Austra	Heart Disease	Phishing	Sonar	Iono	SGC	SpamBase	Messidor	Pop-Failure	points	priority
LR	ScCFS	0.869	0.901	0.926	**0.790**	**0.898**	**0.815**	0.889	0.611	**0.979**	**4**	**1**
	RS	0.862	0.7402	**0.939**	0.640	0.830	0.792	0.420	0.682	0.956	1	2
	WGFS	0.762	0.8052	0.715	0.640	0.788	0.813	0.770	0.683	0.956	0	3
	ScC	**0.890**	**0.9244**	0.919	0.717	0.855	0.789	**0.906**	**0.693**	0.956	**4**	**1**
FLM	ScCFS	**0.869**	0.842	0.927	**0.757**	**0.846**	0.811	**0.902**	**0.745**	**0.986**	**6**	**1**
	RS	0.719	0.7123	**0.939**	0.617	0.825	0.786	0.096	0.698	0.956	1	2
	WGFS	0.695	0.7494	0.715	0.419	0.788	**0.812**	0.774	0.722	0.956	1	2
	ScC	0.803	**0.8938**	0.925	0.749	0.839	0.789	0.865	0.697	0.962	1	2
RF	ScCFS	**0.869**	0.845	**0.931**	**0.785**	**0.956**	**0.830**	0.795	**0.693**	**0.956**	**7**	**1**
	RS	0.862	0.7531	0.932	0.763	0.955	0.825	0.812	0.562	**0.956**	1	3
	WGFS	0.722	0.7665	0.715	0.487	0.890	0.825	0.787	0.631	**0.956**	1	3
	ScC	0.843	**0.8712**	0.930	0.738	0.928	0.825	**0.905**	0.444	**0.956**	3	2
GBT	ScCFS	**0.869**	**0.964**	0.945	**0.747**	**0.941**	**0.847**	**0.830**	0.725	**0.976**	**7**	**1**
	RS	0.854	0.7613	**0.974**	0.720	**0.941**	0.825	0.755	**0.747**	0.956	3	2
	WGFS	0.761	0.7826	0.715	0.683	0.878	0.822	0.763	0.716	0.956	0	3
	ScC	0.866	0.8339	0.941	0.699	0.907	0.789	0.658	0.715	0.956	0	3

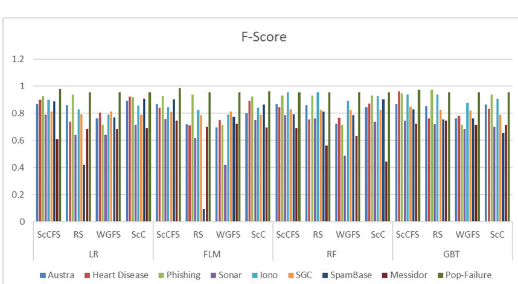

Fig. 3. The F-Score results

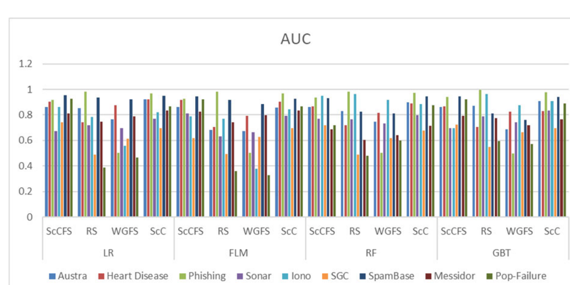

Fig. 4. The AUC results

Table 5
The AUC rates for the reducts of the nine datasets given by the four classifiers.

CM/ FSM	FSM	Austra	Heart Disease	Phishing	Sonar	Iono	SGC	SpamBase	Messidor	Pop-Failure	Points	priority
LR	ScCFS	0.860	0.904	0.919	0.674	**0.860**	**0.744**	**0.956**	0.809	**0.926**	**4**	**1**
	RS	0.851	0.743	**0.980**	0.718	0.782	0.491	0.936	0.747	0.388	1	2
	WGFS	0.763	0.874	0.502	0.695	0.558	0.615	0.921	0.787	0.464	0	3
	ScC	**0.923**	**0.923**	0.968	**0.769**	0.822	0.697	0.948	**0.834**	0.867	**4**	**1**
FLM	ScCFS	**0.860**	**0.918**	0.925	**0.809**	0.788	0.617	**0.945**	0.824	**0.921**	**5**	**1**
	RS	0.683	0.707	**0.980**	0.630	0.771	0.495	0.918	0.740	0.362	1	3
	WGFS	0.674	0.794	0.502	0.662	0.378	0.625	0.887	0.798	0.327	0	4
	ScC	0.859	0.904	0.969	0.792	**0.842**	**0.697**	0.925	**0.832**	0.867	3	2
RF	ScCFS	0.860	0.868	0.934	0.771	0.947	**0.719**	0.931	0.685	0.720	1	3
	RS	0.828	0.718	**0.982**	0.765	**0.963**	0.490	0.826	0.606	0.478	2	2
	WGFS	0.748	0.817	0.502	0.731	0.917	0.619	0.813	0.643	0.598	0	4
	ScC	**0.899**	**0.891**	0.974	**0.797**	0.887	0.678	**0.943**	**0.714**	**0.876**	**6**	**1**
GBT	ScCFS	0.860	**0.866**	0.941	0.694	0.694	**0.725**	**0.945**	**0.793**	**0.920**	**5**	**1**
	RS	0.870	0.706	**0.996**	0.787	**0.964**	0.549	0.811	0.776	0.597	2	2
	WGFS	0.688	0.827	0.499	0.742	0.874	0.664	0.761	0.718	0.572	0	3
	ScC	**0.908**	0.829	0.978	**0.834**	0.908	0.697	0.941	0.765	0.888	2	2

6. Conclusion

In this research article, a new important hybrid feature selection algorithm labeled ScCFB was introduced. This algorithm is based on two important elements: the ScC method and forward feature selection. Important features were identified based on the power of both ScC filter method and the forward feature selection wrapper method. The reduced subset generated by ScC is fed to the forward feature selection method to generate the final reduct. To evaluate the significance of the proposed algorithm, it was compared to RS, ScC, and WG feature selection algorithms in terms of accuracy, F-measure, and AUC. The experiments included nine well-known datasets and LR, FLM, RF, and GPT classifiers. Results showed that the proposed algorithm is merit as it gives better results than the others.

Finally, it is important to conclude that the proposed algorithm is the pioneer in helping governments, public, and private sectors to enhance their model performance. The proposed algorithm can be extended to manage semi-structured, unstructured, and unbalanced datasets.

Acknowledgment

The author wishes to thank his wife, daughters, and friends for their continuous support.

Competing Interests

There are no conflicts.

Data availability

The datasets are derived from different free data repositories. They are available to be downloaded.

References

Alhenawi, E., Al-Sayyed, R., Hudaib A., & Mirjalili, S. (2023). Improved intelligent water drop-based hybrid feature selection method for microarray data processing, *Computational Biology and Chemistry*, *103*, 107809, https://doi.org/10.1016/j.compbiolchem.2022.107809.

Al-Shalabi, L., Shaaban, Z., & Kasasbeh, B. (2006). Data mining: A Preprocessing Engine. *Journal of Computer Science, 2*(9), 735-739.

Al-Shalabi, L. (2016). Data Mining Application: Predicting Students' Performance of ITC program in the Arab Open University in Kuwait – The Blended Learning. *International Journal of Computer Science and Information Security (IJCSIS)*, *14*(12), 827-833.

Al-Shalabi, L. (2017). Perceptions of Crime Behavior and Relationships: Rough Set Based Approach. *International Journal of Computer Science and Information Security (IJCSIS)*, *15*(3), 413-420.

Al-Shalabi, L. (2019). Rough Set-Based Reduction of Incomplete Medical Datasets by Reducing the Number of Missing Values. *The International Arab Journal of Information Technology (IAJIT)*, *16*(2).

Al-Shalabi, L., & Tahhan, Y. (2020). A WGFS-Based Approach to Extract Factors Influencing the Marketing of Korean Language in GCC. *Journal of Computing and Information Technology*, *28*(3), 165–181.

Al-Shalabi, L. (2022). New Feature Selection Algorithm Based on Feature Stability and Correlation. *IEEE Access*, *10*, 4699-4713. doi: 10.1109/ACCESS.2022.3140209.

AL-Shalabi, L. (2022). Evaluation of COVID-19 Vaccine Refusal among AOU Students in Kuwait and their Families and their Expected Inclination Towards the Acceptance or Refusal of the Vaccine. *International Journal of Statistics in Medical Research*, *11*, 147–161.

Aphinyanaphongs, Y., Fu, L.D., Li, Z., Peskin, E.R., Efstathiadis, E., Aliferis, C.F., & Statnikov, A. (2014). A comprehensive empirical comparison of modern supervised classification and feature selection methods for text categorization. *J. Assoc. Information Science Technology, 65*(10), 1964–1987.

Asghari, S., Nematzadeh, H., Akbari, & Motameni, H. (2023). Mutual information-based filter hybrid feature selection method for medical datasets using feature clustering. *Multimed Tools Appl*. https://doi.org/10.1007/s11042-023-15143-0

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge Information System, 34*(3), 483–519.

Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J.M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Science, 282*, 111–135.

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics and Data Analysis, 143*: 106839.

Breiman, L. (2001). Random Forests. *Machine Learning,* 45, 5–32.

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79.

Chaudhary, A., & Kolhe, S. (2013). Performance Evaluation of feature selection methods for Mobile devices. *Int. Journal of Engineering Research and Applications, 3*(6), 587-594.

Cherrington, M., Thabtah, F., Lu, J., & Xu, Q. (2019). Feature Selection: Filter Methods Performance Challenges. *2019 International Conference on Computer and Information Sciences (ICCIS)*, pp. 1-4. doi: 10.1109/ICCISci.2019.8716478.

Cormen, T.H. (2009). Introduction to Algorithms, MIT Press.

Darshan, S.S., & Jaidhar, C. (2018). Performance evaluation of filter-based feature selection techniques in classifying portable executable files. *Procedia Computer Science, 125*, 346–356.

Das, A.K., Sengupta, S., & Bhattacharyya, S. (2018). A group incremental feature selection for classification using a rough set theory-based genetic algorithm. *Applied Soft Computing*, 65, 400–411.

Kamalov, F., Sulieman, H., Moussa, S, Reyes, J.A., & Safaraliev, M. (2023). Nested ensemble selection: An effective hybrid feature selection method. *Heliyon, 9*(9), e19686, https://doi.org/10.1016/j.heliyon.2023.e19686

Fleuret, F. (2004). Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research, 5*, 1531–1555.

Gong, L., Xie, S., Zhang, Y., Wang, M., & Wang, X. (2022). Hybrid feature selection method based on feature subset and factor analysis. *IEEE Access, 10*, 120792-120803.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157-1182.

Haq, A.U., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications, 168*, 114444.

Hoque, N., Singh, M., & Bhattacharyya, D.K. (2018). EFS-MI: An ensemble feature selection method for classification. *Complex Intelligent System., 4*(2), 105–118.

Hu, M., Tsang, E.C.C., Guo, Y., & Xu, W. (2021). Fast and Robust Attribute Reduction Based on the Separability in Fuzzy Decision Systems. *IEEE Trans Cybern.*, doi: 10.1109/TCYB.2020.3040803.

Inza, I., Larrañaga, P., Blanco, R., & Cerrolaza, A.J. (2004). Filter versus wrapper gene selection approaches in DNA microarray domains. *Artificial Intellectual Medicine, 31*(2), 91–103.

Kang, I-A., Njimbouom, SN., & Kim, J-D. (2023). Optimal Feature Selection-Based Dental Caries Prediction Model Using Machine Learning for Decision Support System. *Bioengineering*, *10*(2), 245. https://doi.org/10.3390/bioengineering10020245Kothari, CR. (2007). Quantitative techniques, New Delhi, UBS Publishers Ltd.

Kim, M., Bae, J., Wang, B., Ko, H., & Lim, JS. (2022). Feature Selection Method Using Multi-Agent Reinforcement Learning Based on Guide Agents. *Sensors (Basel), 23*(1), 98. doi: 10.3390/s23010098.

Kohavi, R., & John, G.H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(1–2), 273–324. http://www.robotics.stanford.edu/~gjohn.

Lazar, C., Taminau, J., Meganck, S., Steenhoff, D., Coletta, A., Molter, C., de Schaetzen, V., Duque, R., Bersini, H., & Nowe, A. (2012). A survey on filter techniques for feature selection in gene expression microarray analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform, 9*(4), 1106–1119.

Liu, Y. (2004). A comparative study on feature selection methods for drug discovery. *Journal of Chemical and Information Computational Science, 44*(5), 1823–1828.

Luekiangkhamla, A., Panagant, N., Bureerat, S., & Pholdee, N. (2023). Application of various machine learning models for fault detection in the refrigeration system of a brewing company. *Engineering and Applied Science Research, 50*(2), 149–154. https://ph01.tci-thaijo.org/index.php/easr/article/view/251060

Ma, B., & Xia, Y. (2017). A tribe competition-based genetic algorithm for feature selection in pattern classification. *Appl. Soft Computing, 58*, 328–338.

Mienye, I.D., & Sun, Y. (2023). A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection. *Applied Sciences*, *13*(12):7254. https://doi.org/10.3390/app13127254

Mohtashami, M., & Eftekhari, M. (2019). A hybrid filter-based feature selection method via hesitant fuzzy and rough sets

concepts. *Iranian Journal of Fuzzy Systems, 16*(2), 165–182.
Nayak, SK., Pradhan, BK., Banerjee, I., & Pal, K. (2020). Analysis of heart rate variability to understand the effect of cannabis consumption on Indian male paddy-field workers. *Biomed Signal Process Control, 62*:102072.
Niu, T., Wang, J., Lu, H., Yang, W., & Du, P. (2020). Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications, 148*: 113237.
Pabuccu, H., & Barbu, A. (2023). Feature Selection for Forecasting. *arXiv preprint arXiv:2303.02223*.
Peng, C.Y.J., Lee, K.L., & Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3-14. http://dx.doi.org/10.1080/00220670209598786, 2002.
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, 27*(8), 1226–1238.
Phyu, T.Z., & Oo, N.N. (2016). Performance Comparison of Feature Selection Methods. *The 3rd International Conference on Control, Mechatronics and Automation (ICCMA 2015), 42*(06002). doi: 10.1051/matecconf/20164206002
Pourpanah, F., Shi, Y., Lim, C., Hao, Q., & Tan, C.J. (2019). Feature selection based on brainstorm optimization for data classification. *Applied Soft Computing Journal, 80*, 761–775.
Quinlan, J.R. (1989). Unknown Attribute Values in Induction. *In Proceedings of the Sixth International Workshop on Machine Learning,* 164-168.
Reunanen, J. (2006). Search Strategies. *Feature Extraction Foundations and Applications Stud-Fuzz 207*, 119–136, Springer.
Ringsquandl, M., Lamparter, S., Brandt, S., Hubauer, T., & Lepratti, R. (2015). Semantic-guided feature selection for industrial automation systems. *The Semantic Web - ISWC 2015,* pp. 225-240. http://dx.doi.org/10.1007/978-3-319-25010-6_13, 2015.
Sinayobye, J.O., Kaawaase, K.S., Kiwanuka, F.N., & Musabe, R. (2019). Hybrid Model of Correlation Based Filter Feature Selection and Machine Learning Classifiers Applied on Smart Meter Data Set. *IEEE/ACM Symposium on Software Engineering in Africa (SEiA)*, pp. 1-10. doi: 10.1109/SEiA.2019.00009.
Soheili, M., & Moghadam, A.E. (2020). DQPFS: Distributed quadratic programming based feature selection for big data. *Journal of Parallel and Distributed Computing, 138*, 1-14.
Suresh, S., Newton, DT., Everett, TH., Lin, G., & Duerstock, BS. (2022). Feature Selection Techniques for a Machine Learning Model to Detect Autonomic Dysreflexia. *Front Neuroinform*, *10*(16),901428. doi: 10.3389/fninf.2022.901428.
Tianyi, Z. Yingzhe, Z., & Zhe, W. (2023). Feature selection-based machine learning modeling for distributed model predictive control of nonlinear processes. *Computers & Chemical Engineering*, *169*, 108074. https://doi.org/10.1016/j.compchemeng.2022.108074.
Velayutham, C., & Thangavel, K. (2011). Unsupervised Quick Reduct Algorithm Rough Set Theory. *Journal of Electronic Science and Technology, 9*(3).
Velusamy, K., & Manavalan, R. (2012). Performance Analysis of Unsupervised Classification based on Optimization. *International Journal of Computer Applications, 42*(19), 22-27.
Wah, Y.B., Ibrahim, N., Hamid, H.A., Abdul-Rahman, S., & Fong, S. (2018). Feature selection methods: the case of filter and wrapper approaches for maximizing classification accuracy. *Pertanika Journal of Science and Technology., 26*(1), 329–340.
Xiao, W., Ji, P., & Hu, J. (2021). RnkHEU: A Hybrid Feature Selection Method for Predicting Students' Performance. *Scientific Programming*, *2021*, 1-16. https://doi.org/10.1155/2021/1670593
Xue, B., Zhang, M., & Browne, W.N. (2015). A comprehensive comparison of evolutionary feature selection approaches to classification. *International Journal of Computational Intelligence and Application, 14*(2).
Yin, Y., Jang-Jaccard J., Xu W., Singh, A., Zhu, J., Sabrina, F., & Kwak, J. (2023). IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. *Journal of Big Data,10*, https://doi.org/10.1186/s40537-023-00694-8
Yongbin, Z., Wenshan, L., & Tao, L. (2023). hybrid Artificial Immune optimization for high-dimensional feature selection, *Knowledge-Based Systems, 260*, 110111, https://doi.org/10.1016/j.knosys.2022.110111.
Zhang, Y., Gong, D., Hu, Y., & Zhang, W. (2015). Feature selection algorithm based on bare bones particle swarm optimization. *Neurocomputing, 148*, 150–157.
Zhu, Z., Ong, Y.S., & Dash, M. (2007). Wrapper-filter feature selection algorithm using a memetic framework. *IEEE Transactions on Systems, Man, and Cybernetics, 37*(1), 70–76.