Feature Selection Method Based on Consecutive Forward Selection and Backward Elimination Concepts Using a Weighted Vector

Luai Al-Shalabi
Faculty of Computer Studies, Arab Open University, Kuwait
lshalabi@aou.edu.kw

Abstract: *Feature selection is an essential preprocessing task in many disciplines, including Machine Learning (ML) and the Internet of Things (IoT), and it is the most demanding process for data analysis. This process attempts to identify and remove as much irrelevant and redundant information as possible in a controlled manner. Existing algorithms still have limitations in selecting the most informative features maintaining high classification accuracy results. This study proposed a consecutive Forward selection and Backward Elimination algorithm (FBWV) that enhances feature selection by applying the forward selection concept, backward elimination concept, weighted chi-square vector, and custom decision threshold value. The FBWV model framework was optimized through data preprocessing and parameter tuning. The effectiveness of the proposed method was evaluated by comparing it with other state-of-the-art Feature Selection Algorithms (FSA), namely, Rough Set (RS), Weight-Guided (WG) feature selection, and Stability-correlation and Correlation (ScC). The reduced subsets were trained by several classifiers using different measures, including accuracy, F-measure, reduction rate, and AUC. The results revealed that the FBWV effectively reduced the size of the given datasets. It achieved the highest accuracies of 85.28%, 88.33%, 96.26%, 81.36%, 96%, 74.39%, 81.89%, 65.26%, and 98.69% for Austra, Heart Disease, Phishing, Sonar, Iono, SGC, and SpamBase, respectively. The Messidor and Pop-Failure datasets outperform the other FSAs. Moreover, it achieved the highest F-measure and AUC rates of 97.94% each for the Pop-Failure dataset. The FBWV proved the capability of handling different types of datasets and reduced computational complexity, storage, and cost.*

Keywords: *Feature selection, forward-backward, chi-square, classification.*

Received October 11, 2024; accepted March 3, 2025
https://doi.org/10.34028/iajit/22/4/13

1. Introduction

Recently, dimensionality reduction methods have become very important in fields such as machine learning, Artificial Intelligence (AI), data science, data mining, big data, and the Internet of Things (IoT). Among the methods available, feature selection is of great interest. WG, RS, and ScC are methods that attempt to minimize the size of a dataset to its minimum possible dimension by retaining its value, credibility, and accuracy. As Al-Shalabi [3] mentioned, the accuracy of the reduced subset could be the same, greater, or less than that of the original dataset. Features such as the name of a person, redundant features, and features with high variation rates in their values usually have low significance values, which reflect their irrelevancy to the given dataset. Removing them will most likely improve the accuracy of the dataset. Moreover, the dataset will be more understood and more suitable for further analysis, such as classification.

Building a classification system with few features is highly beneficial to stakeholders such as customers, doctors, patients, and others. This system minimizes effort, costs, and time for developers and users. Accuracy is one of the important measures to consider when we build risky systems using reduced subsets such as medical, stock market, and banking systems. None of the feature selection methods are appropriate for all kinds of datasets. The process is dataset-based. Therefore, several methods should be tested, and the best methods should be adopted for the given dataset. As in many previous studies, the number of features and the accuracy measure are the main players. Researchers are looking for more reduction rates with higher accuracy values, which is a challenge.

This study proposed a filter consecutive forward selection and backward elimination model for feature selection. The model is based on a weighted vector generated via the chi-square method and a custom hreshold value. Different classifiers are used to test the algorithm's efficiency by training the reduced subset generated by the proposed method. The accuracy of each classifier and other performance metrics were calculated. If the accuracy is similar (closely similar to or higher) to the accuracy of the original dataset, then the subset is accepted. This implies that the proposed model correctly selects the most significant features. Nine benchmark datasets were employed to evaluate the effectiveness of the proposed model. The results were statistically analyzed via four metrics, namely, accuracy, F-measure, AUC, and the reduction rate; three feature selection methods, namely, RS, ScC, and WG; and four classifiers, namely, logistic regression (LR), Fast Large Margin (FLM), Random Forest (RF), and Gradient-Boosted Trees (GBT).

The main contributions of this study are threefold:

- A consecutive forward selection and backward elimination model is based on a weighted vector generated from the chi-square results and a custom threshold value. This model reduces the dimension of a dataset to the optimal level by having the most relevant features.
- A comprehensive statistical evaluation of the proposed model using well-known benchmark datasets.
- A comprehensive comparison of the results (accuracy, reduction rate, F-measure, and AUC) of the four well-known feature selection methods and the four famous classifiers mentioned earlier.

The rest of this article is organized as follows. Section 2 describes the literature review. Section 3 describes the methodology. Section 4 describes the proposed approach. Section 5 describes the comparison of results achieved. Section 6 presents the conclusions of the work.

2. Literature Review

In the literature, numerous feature selection methods have been proposed to minimize the dimensionality of datasets. There are three types of feature selection methods, namely, filter, wrapper, and embedded methods, which are explained by many researchers [3, 20]. Filter methods have low computational complexity [41]. As discussed by Guyon and Elisseeff, each feature in the filter method is determined by a score generated via calculations. Scores that exceed a definite threshold value represent the most informative features and are only selected [22]. Filter methods are fast and simple and can work with thousands of features. On the other hand, wrapper methods were explained by Patel *et al.* [39]. They are used with a classifier to estimate the efficiency of the reduced subset. It is applied to all assumed reduced subsets and the reduct with the highest efficiency is chosen. On the basis of the accuracy of the classifier, a decision is made to remove or add the features to the subset. The process is repeated until the decision is taken to end the process. The technique is considered expensive because of this long training process [30]. Compared with filter methods, wrapper approaches are more efficient, as a classifier is required to repeat the learning of each reduced subset, but this leads to complexity in the method [55]. The embedded methods use a machine learning model to find the best attributes basis of the model's accuracy. Such methods reduce overfitting via built-in penalization functions. They combined the abilities of both filter and wrapper methods.

The comparison of several feature selection methods is important because it provides deep insight into the best methods for a given dataset. The classification accuracy is one way to test the performance of the method used. Other metrics are also considered, as mentioned earlier. Some feature selection methods are based on correlation values, such as the ScC filter method proposed by Al-Shalabi [3]. The method integrates the stability of the feature and its correlation value. The CFS method is another method proposed by Hall [23] and can be applied to classification and regression problems. Hoque *et al.* [25] developed a filter method that sums the scores of various filter methods and compares them to those of single filter methods. Nge *et al.* [36] proposed filter FSAs that use the γ-metric for evaluation to select important features. A new filter feature selection method called CONMI was constructed by Huanhuan *et al.* [27], which uses the normalized mutual information and the Pearson correlation coefficient for classification tasks. Many researchers have used classification accuracy to compare filter methods [1, 29, 34, 37, 45, 52]. Pirgazi *et al.* [41] merged filter and wrapper methods based on the Harris–Hawks optimization algorithm to identify the optimal subset of features. A hybrid filter-genetic feature selection approach to solve the high-dimensional microarray dataset problem, which ultimately enhances the precision of cancer classification, was proposed by Ali and Saeed [2]. The FAM-BSO feature selection model for data classification was proposed by Pourpanah *et al.* [42]. They combined the fuzzy ARTMAP model and the BSO feature selection method. Cherrington *et al.* [15] examined and analyzed several filter methods based on ranking procedures considering the idea of ranked scores and how threshold determination can affect the results of several filter methods. Bommert *et al.* [11] investigated many filter methods for accuracy and runtime. They concluded that no filter technique always outperforms the other techniques. A hybrid classification model that combines a correlation-based filter method and machine learning classifiers was proposed by Sinayobye *et al.* [48]. A Sequential Forward Selection method based on Separability (SFSS) was proposed by Hu *et al.* [26]. The SFSS has high accuracy and a low computational time. Cilia *et al.* presented a comprehensive comparison between filter and wrapper techniques for feature selection in the field of handwritten character recognition. The results confirmed that the filter and wrapper methods perform similarly [18]. Maseno and Wang [32] proposed a sequential FSA using an Extreme Learning Machine (ELM) and an SVM. To select the informative features, the algorithm, as an estimator, is applied in wrapper sequential forward selection. Comparisons between various wrapper methods on the basis of classification accuracy were conducted by Zhu *et al.* [56] and

Mohtashami and Eftekhari [34]. Xue *et al.* [54] compared wrapper and filter methods on the basis of classification accuracy. Al-Shalabi [3] proposed a hybrid feature selection method based on ScC and forward selection techniques to generate the most informative subset. He used numerous classifiers and compared the accuracy measures. Chaudhary *et al.* [16] examined filter, wrapper, and embedded methods and compared the performance of the gain ratio, correlation, and information gain approaches with that of the naïve Bayes classifier. Boln-Canedo *et al.* [10] studied filter, wrapper, and embedded feature selection methods. They used various classifiers and compared the accuracy measures. A dimensionality reduction method for noisy datasets was proposed by Al-Shalabi [5], and the results were promising. Haq *et al.* [24] used multiple feature selection methods to generate an optimal reduced subset by combining features selected via these methods. DQPFS is a scalable algorithm proposed by Soheili and Moghadam [49]. DQPFS is based on the Apache Spark cluster computing model and yields significant feature selection results for big data. Bermego *et al.* [9] suggested an adaptation of the CMIM feature selection method for multilabel feature selection. Their work efficiently approximates the conditional multivariate mutual information of each nominee attribute concerning the whole set of features.

Table 1. Accuracy comparisons between different methods.

Reference	FSA-classifier	Dataset	Accuracy
[3]	ScC-GBT	Iono	93
	ScC-GBT	Phishing	92.64
[23]	CFS-C4.5	Iono	90.94
	CFS-KNN	Sonar	79.79
[25]	EFS-KNN	TUIDS	95
[27]	CONMI_FS-KNN	Multiple datasets	88.83
	CONMI_FS-SVM		88.98
[29]	SCF-SVM	Sonar	86.41
[2]	A hybrid filter-genetic feature selection-RF	Breast cancer	93.81
[56]	WFFSA – SVM	Iono	95.19
[54]	ScCFS-RF	Iono	94
	ScCFS-RF	Sonar	73.33
This work	FBWV-GBT	Iono	96
	FBWV-GBT	Phishing	96.26
	FBWV-GBT	Sonar	81.36

Deep Learning (DL) is a technique that extracts relevant features from domain datasets. It achieves high performance in various domains and tasks, including classification, feature selection, and clustering. It works better with big data. Alsini *et al.* [7] studied the prediction of whether a person is emotionally feeling too much or a logical thinker utilizing Bi-LSTM. The performance of the study was 91.57%. Jenifa *et al.* [28] proposed a personality recognition model that leverages Generative Artificial Intelligence-based Learning Principles (GAILP), which combines text-based features extracted via natural language processing (NLP) techniques with DenseNet feature extraction from user profile images. The accuracy of the model was greater than 97%. Saeidi [47] proposed a model to identify the personality characteristics of WhatsApp users via a Long Short-Term Memory (LSTM) neural network by investigating the most commonly used emojis. The accuracy of the model was 95.48% when a random forest classifier was used.

Table 1 shows the accuracy comparisons between the different feature selection methods explained earlier. The intersection between these feature selection methods and the proposed method is that all the most informative features are selected by implementing various techniques.

3. Methodology

This section explains all the necessary information and practical issues related to the design and testing of the proposed algorithm.

3.1. Datasets

Nine well-known datasets from various domains and different sizes were used in the analysis. The numbers of features, examples, and classes for each dataset are listed in Table 2. The scope of this research is limited to binary classification. All the datasets are balanced except the Pop-Failure dataset. Nevertheless, it has not been balanced to compare its results accurately with those of previous works, which also have not balanced it.

Table 2. The datasets.

Dataset	# of features	# of examples	# of classes
Sonar	60	208	2
SpamBase	57	4601	2
Iono	34	351	2
Phishing	30	11055	2
SouthGerman Credit (SGC)	20	1000	2
Pop-Failure	20	540	2
Messidor	19	1151	2
Austra	14	690	2
Heart Disease	13	270	2

3.2. Dimensionality Reduction Methods

The feature selection process finds the most important features from the list of features in a dataset, and they are called the reduced subset. This reduced subset is significant in improving the efficiency of the learning process and consequently reduces the running time of the learning process. Three state-of-the-art dimensionality reduction methods were used to evaluate the proposed method. To extend the comparison results, the accuracy, reduction rate, F-measure, and AUC evaluation methods were used. The feature selection methods used are RS, ScC, and WG, whose properties are briefly summarized next. The performance comparisons between them reveal their limitations in selecting appropriate features.

1) Rough set: one of the most important features of rough set theory is the discernibility matrix used to

generate a reduced subset of features [51]. The matrix identifies both the essential and unessential features. RS generates many reduced subsets that are called reducts. Each has a different size with common significant features among all the reduced subsets called the core. The generated reduced subset *R* of essential conditional features *X,* where *X* is a subset of all conditional features *C* of the original dataset *D,* gives the same quality of classification γ as the original dataset: $\gamma_X(R)=\gamma_C(D)$ [50]. All generated subsets are defined as follows:

$$Rall = \{X|X\subseteq C, \gamma X(R) = \gamma C\ (D) \quad (1)$$

Usually, the most reduced subset is the preferred one. This reduced subset must satisfy $R_{min}\subseteq Rall$, where;

$$Rmin = \{X|X\subseteq Rall, \forall Y \in Rall, |X| \leq |Y|\} \quad (2)$$

2) Stability correlation and correlation: the ScC method is a novel FSA whose aims are intended to increase dataset reduction rate and performance accuracy by merging the stability and correlation aspects [3]. Given a dataset $DS=(O, F, V, f)$, where O is a finite set of instances, F is a set of features, $V=\cup a\in F$, Vn is a domain of feature a, $f: O\times F\rightarrow V$ is a function such that $f(x,a)\in Vn$ for every $a\in F$, $x\in O$, and *Red* is the reduced dataset. Let $\text{A}(Red)$ be an estimate of the accuracy of the reduct produced by ScC. In the dataset (*DS*), the minimal subset *Red*, where $Red\subseteq F$ such that $\text{A}(Red)$ is maximized, is called the reduct of *DS* and is denoted by reduct(*DS*).
3) Weight guided: the attribute weights are the input to the weight-guided feature selection method to determine the order of features added to the reduced subset. The features that have the highest weights are added first to the reduced subset. The algorithm is stopped if one of the following conditions is satisfied: if an addition to the reduced subset does not improve the efficiency or if all attributes are already added [46].

Table 3. The advantages and disadvantages of the RS, WG, and ScC methods.

FSA	Advantages	Disadvantages
RS [50]	Handle. uncertainty in data.	Scalability limit as it is computationally intensive.
	easily can interpret the selected features	The Discretization of the continuous features is needed.
	Does not require prior knowledge about data.	Noisy data affects the performance of the method.
WG [46]	Weight can be adjusted to suit different types of data.	It is weight dependent which may affect its performance.
	It uses weight to determine the important features.	High complexity of calculating weights.
	It improves performance by reducing overfitting.	Wrongly using the weights will increase the risk of overfitting.
ScC [3]	The generated reduct can easily interpreted.	The wrong value of the stability parameter will lead to an overfitting.
	It combines correlation and stability to guarantee its robustness to noise.	The performance of ScC heavily depends on the threshold value. The wrong choice of this value may lead to inappropriate reduct.
	It does not require any knowledge of data.	Recalculating the correlation in sequence stages increases the complexity of the method.

Table 3 summarizes the advantages and disadvantages of the RS, WG, and ScC feature selection algorithms.

3.3. The Classification Algorithms (CA)

Four classifiers were employed to evaluate the performance of the feature selection methods mentioned earlier (including the proposed method). The RapidMiner tool, which uses a 10-fold cross-validation technique for the learning procedure, was used to implement the four classifiers described next, and the default settings were employed otherwise.

1) Logistic regression is a supervised ML algorithm. It analyzes a dataset and answers yes/no questions to decide whether an attribute supports an obvious result. There are many types of logistic regression, such as ordinal, binary, binomial, and multinomial [40].
2) A fast large margin algorithm is based on a linear support vector. It is one of the important algorithms in machine learning and is used with large datasets. The results generated by FLM are similar to the results achieved by LR [53].
3) A random forest algorithm that was proposed by Breiman [13] creates a set of tree-based classifiers. It utilizes many types of problems, such as regression and classification. During the training process, several decision trees are created. The algorithm selects the mutual class of individual trees for classification [6, 44].
4) Gradient-boosted trees increase the accuracy of sequentially produced trees. It is important to reduce the speed while increasing the accuracy. GBTs are called shallow learning because they use a two-layer procedure [35].

3.4. Performance Metrics

Four performance metrics were used to test the performance of the proposed algorithm as explained below:

1) The reduction rate represents the percentage of the dimensions of the original dataset that are minimized. It is calculated by dividing the number of reduced attributes (m) by the total number of attributes (n) of the given dataset as in the following formula:

$$Red - rate = m/n \quad (3)$$

2) Accuracy: Accuracy is a metric for evaluating a given classification model. It can be described by the proportion of the number of correct predictions to the total number of input examples. It is calculated via the following formula:

$$Acc = ((TP + TN)/(TP + FP + TN + FN) \quad (4)$$

where *TP* and *TN* are the true positives and the true negatives, whereas *FP* and *FN* are the false positives and the false negatives.

3) Area under the ROC curve (AUC): AUC is the metric used to measure the classifier's ability to discern between positive and negative classes of a dataset. A higher value of the AUC signifies better performance of the classifier. The reader may refer to [38] for extra information.
4) F-Measure: F-Measure is a combined metric that measures the recall (R) and precision (P) metrics into a single measure. The precision metric calculates the ratio between the positive instances and all instances predicted as positive. The recall metric calculates the ratio between the positive instances and all instances that should have been predicted as positive. It is calculated via the following formula:

$$F-measure = 2((P * R)/(P + R)) \quad (5)$$

such that;

$$P = TP/(TP + FP) \quad (6)$$

and;

$$R = TP/(TP + FN) \quad (7)$$

For more information, the reader may refer to [17].

4. The Proposed Approach

This section discusses the three main pillars of the proposed method, namely, the weighted vector, forward selection/backward elimination, and custom threshold. The step-by-step algorithm for the proposed method is presented below.

4.1. Weighted Vector

The chi-square statistic calculates the relationships among attributes (features) concerning the class attribute. For example, a chi-square test can be used to test whether age and education level are related to all people in Kuwait. Aslam and Smarandache [8] reported that the chi-square test determines and tests a significant association between two categorical variables. Lu *et al.* [31] studied how to fuse heterogeneous local test statistics with linear weights for chi-square in a distributed radar system. Chi-square statistics use nominal data, so they use frequencies rather than variances or means. The chi-square test computes the sum of the squared differences between the observed and expected values, as shown in Equation (8). The feature with a higher X^2 is more important for the classification decision.

$$X^2 = \sum_{i=0}^{r} \sum_{j=0}^{c} \frac{(Oij - Eij)^2}{Eij} \quad (8)$$

where Oij and Eij are the observed frequency and expected frequency, respectively, c is the class number, whereas r is the number of bins used for the discretization of numerical features.

To understand and easily use the results of the chi-square algorithm, each calculated chi-square value is normalized to a value between 0 and 1. This normalization is known as the weight of the chi-square value W(X^2). The higher the weight calculated for a feature is, the more relevant it is judged. In this case, the observed results fit well. All the weights are listed in the weighted vector.

4.2. Forward Selection and Backward Elimination

The forward selection method is used in machine learning to find the optimal reduct by selecting the most relevant features from the given dataset. This technique starts with an empty set and then adds the most relevant features that have a large correlation with the dependent feature one at a time. At each iteration, the subset is trained, and the decision is made to keep or remove that attribute from the subset. If the model is not improved, then the attribute will be removed.

The backward elimination method is used in machine learning to find the best reduct by removing the most irrelevant features from a given dataset. This technique starts with all features of the original dataset and removes the least significant features one at a time on the basis of the conclusions drawn from the training process.

Frederick compared the performance of forward selection and backward elimination methods [21]. Chowdhury and Turin highlighted the importance of forward selection and backward elimination in clinical prediction [14]. The forward-backward selection algorithm applies to many types of data. It starts with a forward phase and then a backward phase on the selected features [12]. The concepts of forward selection and backward elimination are used in this research without the training process, which is iterated in wrapper methods. This gives the proposed method preference over the wrapper methods, which are computationally more expensive. Moreover, the proposed method can eliminate a subset of features at once rather than the backward elimination method, which removes one feature at a time.

4.3. Custom Threshold Value

The goal is to find the best threshold value that helps in producing a smaller number of features while maintaining high performance. Threshold values of 1%, 5%, and 10% were extensively tested. The 1% value produced many features (similar to the original features) that violated the goal, so it was ignored. The 10% value produces a very short, reduced subset, but the performance was low and violated the goal stated, so it was also ignored. Finally, the 5% threshold value was tested, and it produced reasonable features while maintaining high performance.

Selecting 0.5 as the threshold best value was the finest choice for choosing the most informative features that have a high influence on the predictions and improve the performance measures. Moreover, using cross-validation to tune the threshold value justifies that 0.5 is the best value. The metrics, including accuracy, precision, recall, F-measure, and AUC, demonstrate that 0.5 is the best threshold value. In addition, this value performs well across the different classification methods used, achieving consistent results that confirm its fitness to the context of the datasets used.

4.4. The Proposed Algorithm

The proposed algorithm for feature selection consists of three stages. In stage 1, the chi-square method is applied to find the relevancy between every independent attribute in a dataset and the dependent attribute. A higher value of a chi-square represents greater appropriateness of that attribute to be part of the reduced subset. The results of the chi-square algorithm are represented by the chi-square vector. In stage 2, the normalization process is applied to each value in a chi-square vector to convert it to a value between zero and one. This normalization increases the understanding of the result of the chi-square values since it has no upper limit. Normalized values are represented in the so-called weighted vector. In stage 3, the forward selection and backward elimination processes are iteratively applied to the weighted vector. In each iteration, the algorithm searches for weights equal to 1 (the forward selection part) and then selects the corresponding attributes and stores them in the reduced subset (*S1*). After that, the algorithm eliminates the attributes of custom weights less than or equal to 0.05 (the backward elimination part) and stores them in the rejected subset (*S2*). The corresponding chi-square values for all eliminated features were also eliminated from the chi-square vector, and the normalization process was repeated for the new chi-square vector. The process is repeated until no more attributes remain. Finally, *S1* is chosen as the best-reduced subset that has the most informative features. Several tests were conducted to find the most likely threshold value (the custom value), which reflects the test of eliminating the unimportant attributes. The best value is 0.05. This value increased the accuracy of the reduced subset generated by Forward selection and Backward Elimination (FBWV).

Given an original dataset $ODS=(R, A, V, f)$, where R is a set of rows, A is a set of attributes, $V=\cup a\in A$, V_n is a domain of attribute a, and $f{:}R\times A\rightarrow V$ is a function such that $f(r,a)\in V_n$ for every $a\in A$, $r\in R$. Let ₳(*Red*) be an approximation accuracy of the reduced subset (*Red*) generated by the FBWV. For the O*DS*, the minimal subset generated is called *MS*, where $MS\subseteq A$ is called the reduct of the *ODS* and is denoted by REDUCT(*ODS*).

To specify the FBWV in detail, the following variables are used: *ODS*, *IDA* (the list of dependent attributes), *D* (the dependent attribute), *CS* (the list of chi-square values or the chi-square vector), *SA* (the list of selected attributes), *EA* (the list of eliminated attributes), *WV* (the list of the weighted vector), *W* (the weights value), *Ac* (the classification accuracy), *C* (the classifier), and *α* (the custom threshold value).

The structure of the proposed method is shown in Figure 1 and is explained as follows:

- *Step* 1: Start with the original dataset (*ODS*), *initialize IDA* with all the attributes in the *ODS*, initialize *SA* with the blank, initialize *EA* with the blank, and initialize *WV* with the blank.
- *Step* 2: For each independent attribute in *ODS*, the chi-square (X^2) value was calculated concerning the dependent attribute, and all values were stored in a chi-vector.
- *Step* 3: The chi-vector is normalized to values between [0, 1], and the weighted vector ($WV(X^2)$) is generated.
- *Step 4*: While *IDA* is not empty, it moves attributes with W=1 to *SA* and attributes with W<=0.05 to *EA*. The corresponding chi-square values are removed from the *CS,* and the weights are recalculated through the normalization process.
- *Step* 5: *SA* is the list with the most appropriate attributes.
- *Step* 6: Construct the reduced subset from *ODS* and *SA*.
- *Step* 7: Evaluate the reduced subset via different classifiers.
- *Step* 8: The classifier that gives a higher accuracy value is chosen.

The proposed algorithm was built as shown below:

Algorithm 1: FBWV Feature Selection Method.

```
Input: ODS, IDA
Output: Selected Features SA
SA ← Ø  //List of selected attributes, initially empty.
EA ← Ø  //List of eliminated attributes, initially empty.
CS ← Ø  //List of chi-square values, initially empty.
α=0.05  //The threshold value.
//Loop until all independent attributes in ODS are chosen.
//Calculate the chi-square value for each Attribute A_i
//concerning the dependent attribute D.

for each attribute A in IDA (A_i∈ODS)
      X^2_AixD ← ∑((O-E)^2/ E)
      CS_i← X^2_AixD
Call the Weight function (algorithm 2).
while IDA is not empty do for each corresponding weight
      for values in WV
if W_i =1 then
      SA_i←IDA_i
      Delete the corresponding attribute from IDA
      Delete the corresponding attribute from CS
else
   if W_i <= α then
EA_i←IDA_i
delete the corresponding attribute from IDA
      Delete the corresponding attribute from CS
```

end if
end if
end for
Call the Weight function (Algorithm (2)).
end while
return SA // the list of selected features.

Algorithm 2: Weighted Function.

Input: CS list
Output: WV list
//Empty the WV list and calculate the weight W for each
//value in CS
$W \leftarrow \emptyset$
for each value in CS
$CS_{i(scaled)} \leftarrow CS_i - min(CS)/max(CS)-min(CS)$
$WV_i \leftarrow SC_{i(scaled)}$

Algorithm 3: The Best Classifier BC is Based on Ac.

Input: SA, the classifiers $(C_{i=1\ to\ n})$
Output: the best BC

Call the Feature selection Algorithm (1).
//for all classifiers $C_{i=1\ to\ n}$ *Loop*
for each $C_{i=1\ to\ n}$
Execute C_i
$Ac\text{-}List_i \leftarrow Ac(C_i)$
end for
$BC = max(Ac\text{-}List)$

4.5. Complexity of the Proposed Model

Big-O notation was used to evaluate the complexity of the proposed algorithm. For more details about big-O, the reader may refer to [19]. Algorithm (1) calculates the chi-square value for each attribute concerning the classification attribute, and one loop needs to be used to pass through all the attributes of the dataset. Let N be the number of conditional attributes and T_1 be the time complexity for the loop; then, the worst case is $O(N)$. Algorithm (1) subsequently generates the reduction via two nested loops; the outer loop iterates n times, and for each iteration of the inner loop, the list (in its worst case) is shortened by one element, meaning that the inner loop will need to run m times. Let T_2 be the time complexity for the reduction nested loops, including the calling of Algorithm (2) (A2); then, the worst case for T_2 is $O(n)*(O(m)+T(A2))$.

Algorithm (2) is called by Algorithm (1). Let N be the number of chi-square values processed and T_3 be the time complexity, which is O(N). In total, the time complexity for Algorithms (1) and (2) together is $T1+T2+T(A2)$, which yields $O(N)+O(n)*(O(m)+T(A2))$. The result will be $O(N)+O(n)*(O(m)+O(N))+O(N)=O(N)+O(n)*O(N)+O(N)=O(N)+O(nN)+O(N)=O(N)+O(N^2)+O(N)$, which gives $O(N^2)$. In conclusion, the complexity of the FBWV is $O(N^2)$.

The complexity of RS is $O(n^2)$, where n is the input number of points from a given universe, and complex calculations are needed within each iteration [43]. The complexity of ScC and weight guided via KNN is $O(NM)$, where N is the number of input features and M is the number of input values for the mode function [3]. Both require complex calculations to achieve the partial result of each iteration. The proposed method requires fewer intensive calculations. Compared with that of the other methods, the complexity of the proposed method is greater.

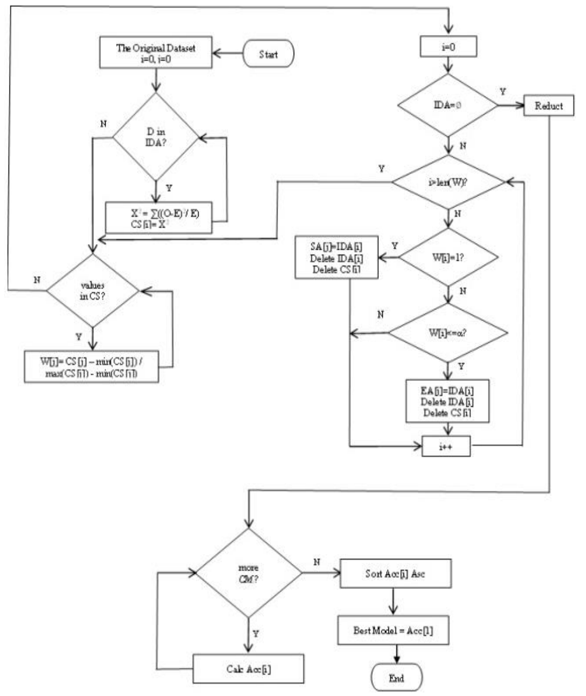

Figure 1. The flowchart of the proposed model.

5. Comparison of Results

In this section, the results are presented, and experiments are conducted to determine the performance of the proposed algorithm FBWV and compare its performance to that of other state-of-the-art reduction algorithms mentioned earlier. To do so, the nine previously explained datasets were used. The efficiency of the proposed FBWV algorithm and the other state-of-the-art algorithms was evaluated via LR, FLM, RF, and GBT. The evaluation was conducted concerning accuracy, reduction amount, F-measure, and AUC.

The results showed that the FBWV algorithm is very promising for selecting an efficient reduct that increases the accuracy. The FBWV algorithm outperforms the other tested algorithms in terms of accuracy in most cases, as shown in Tables 5 to 9.

The major performance metrics used during the development of the FBWV algorithm are accuracy and reduction amount. This study tries to choose the best minimal set with the highest accuracy.

Tables 4 to 12 present the overall results of this study. Table 4 shows the reduction amounts of all the reduction algorithms used, including the FBWV. It competes with other methods in terms of the reduction rate for many of the tested datasets. Tables 5 to 9 show the accuracies of the LR, FLM, RF, and GBT methods when the reducts generated by the RS, WG, ScC, and FBWV are used for the nine benchmark datasets used in this study. Tables 11 and 12 show the performance of the F-measure and AUC, respectively. As explained by Pirgazi *et al*. [41], the reduction amount could be high, but the accuracy is low; such a reduction should be rejected. The phishing dataset is an example where the reduction amount of the WG (96.7%) is greater than the reduction amount of the FBWV (86.7%), but the accuracy rate of the FBWV using the LR (91.36%) was greater than that of the WG (55.70%), with a large difference. These results reflect the importance of the reduct produced by the FBWV. The reader may refer to similar cases shown in the results of this work. Al-Shalabi [3] reported that both the reduction amount and accuracy are important for a good classification system. The best choice is when both are high.

Table 4. The deduction rate of each reduction algorithm.

		RS		WG		ScC		FBWV	
Dataset	**Original**	**#Reduced**	**rate**	**#Reduced**	**rate**	**#Reduced**	**rate**	**#Reduced**	**rate**
Iono	34	18	0.529	31	0.912	26	0.765	20	0.588
SGC	20	18	0.90	17	0.85	19	0.95	13	0.65
SpamBase	57	40	0.702	45	0.79	51	0.895	47	0.825
Messidor	19	15	0.79	15	0.79	16	0.842	12	0.632
Pop-Failure	20	18	0.90	15	0.75	18	0.90	14	0.70
Austra	14	11	0.786	12	0.857	9	0.643	10	0.714
Heart Disease	13	10	0.769	7	0.539	8	0.615	7	0.539
Phishing	30	7	0.233	29	0.967	25	0.833	26	0.867
Sonar	60	38	0.633	50	0.833	39	0.65	46	0.767

Table 5. The accuracy of the austra and heart disease datasets.

	Austra					Heart disease				
CM/FSA	**Original**	**RS**	**WG**	**ScC**	**FBWV**	**Original**	**RS**	**WG**	**ScC**	**FBWV**
Logistic regression	0.8531	0.8579	0.6953	0.8678	0.8528	0.7942	0.6758	0.7650	0.9092	0.8317
Fast large margin	0.5556	0.6396	0.6396	0.7819	**0.8528**	0.8567	0.6542	0.6900	0.8700	0.8833
Random forest	0.8782	0.8579	0.7010	0.8733	0.8579	0.7675	0.6883	0.7167	0.8442	0.8481
Gradient boosted trees	0.8326	0.8477	0.6701	0.8586	0.8579	0.8050	0.7033	0.7517	0.8208	0.8025

Table 6. The accuracy of the phishing and sonar datasets.

	Phishing					Sonar				
CM/FSA	**Original**	**RS**	**WG**	**ScC**	**FBWV**	**Original**	**RS**	**WG**	**ScC**	**FBWV**
Logistic regression	0.9307	0.9313	0.5570	0.9047	0.9241	0.7288	0.5258	0.6621	0.7000	0.7167
Fast large margin	0.9307	0.9310	0.5570	0.9126	0.9218	0.7288	0.5409	0.5500	0.7333	0.7636
Random forest	0.9253	0.9237	0.5570	0.9256	0.9269	0.6833	0.7136	0.5758	0.7167	0.7333
Gradient boosted trees	0.9405	0.9332	0.5570	0.9246	0.9626	0.7470	0.6833	0.6773	0.6773	0.8136

Table 7. The accuracy of the iono and SGC datasets.

	Iono					SGC				
CM/FSA	**Original**	**RS**	**WG**	**ScC**	**FBWV**	**Original**	**RS**	**WG**	**ScC**	**FBWV**
Logistic regression	0.8500	0.7800	0.6500	0.8600	0.7800	0.7343	0.6644	0.6958	0.7018	0.7439
Fast large margin	0.8400	0.7700	0.6500	0.8400	0.8100	0.7018	0.6574	0.6888	0.7018	0.7018
Random forest	0.9500	0.9400	0.8500	0.9100	0.9500	0.7204	0.7018	0.7018	0.7018	0.7158
Gradient boosted trees	0.9000	0.9200	0.8300	0.9300	0.9600	0.7474	0.7018	0.7028	0.7018	0.7368

The analytical results of this research prove the importance of the proposed model. The FBWV achieves a high reduction rate and the highest accuracy among all feature selection methods for specific classifiers tested on all datasets. The highest accuracy was for the Iono dataset, with 96% accuracy when the GBT classifier was used. All classifiers produced the highest accuracy for the SGC, sonar, and pop-failure datasets, indicating that these datasets reduced by the FBWV are appropriate for achieving the highest accuracy. The heart disease, phishing, iono, and messidor datasets achieved the highest accuracy when using two classifiers, whereas the austra and spamBase datasets achieved the highest accuracy when using only one classifier. At least one classifier can produce the highest accuracy for the reduct produced by FBWV.

Table 8. The accuracy of the spamBase and messidor datasets.

	SpamBase					Messidor				
CM/FSA	**Original**	**RS**	**WG**	**ScC**	**FBWV**	**Original**	**RS**	**WG**	**ScC**	**FBWV**
Logistic regression	0.6164	0.5548	0.6768	0.8851	0.7960	0.6231	0.6708	0.5394	0.5303	0.5958
Fast large margin	0.7146	0.4247	0.7688	0.8227	0.8044	0.7561	0.6522	0.6442	0.5410	0.6526
Random forest	0.6621	0.7329	0.7123	0.8813	0.8455	0.5303	0.5793	0.5714	0.5915	0.6462
Gradient boosted trees	0.6271	0.6058	0.6167	0.6903	0.8189	0.5305	0.7121	0.6707	0.6191	0.6262

Table 9. The accuracy of the pop-failure dataset.

CM/FSM	Original	RS	WG	ScC	FBWV
Logistic regression	0.9157	0.9157	0.9157	0.9157	0.9611
Fast large margin	0.9222	0.9157	0.9157	0.9290	0.9869
Random forest	0.9157	0.9157	0.9157	0.9157	0.9157
Gradient boosted trees	0.9157	0.9157	0.9157	0.9157	0.9546

Table 10. The increase in accuracy (%) achieved by the FBWV proposed method for all datasets.

CM/FSM	Austra	Heart Disease	Phishing	Sonar	Iono	SGC	SpamBase	Messidor	Pop-Failure
Logistic regression				1.21		0.96			4.54
Fast large margin	29.72	2.66		3.48		0		-0.1	6.47
Random forest		7.66	1/0.16	5	0	-0.46		0.11	0
Gradient boosted trees			1/2.21	6.66	6	-0.0088	19.18		3.89

Table 10 shows the increase in accuracy achieved by the FBWV proposed method over the original dataset. The largest increase occurred when the FLM classified the austra dataset, and the prediction accuracy increased by 29.72%. RF also achieved the same accuracy as the original dataset for the iono and pop-failure datasets. In some cases where the value is negative, the accuracy of the reduced dataset was less than that of the original dataset. The negative rate is small such that it approaches zero in some cases, so it is not countable.

Tables 11 and 12 show the performance results of the F-measure and AUC, respectively, generated by the four classifiers explained earlier. Each classifier trains the reducts generated by each of the four feature selection methods used in this research. The results highlight the high performance of the FBWV, which highlights the importance of its reduction. The ranking of the importance of the four classifiers was calculated and presented to simplify the reading of the results. F-measure evaluation puts FLM, RF, and GBT in the first rank (priority), as they are applied to the reduct generated by the FBWV, whereas LR ranks second. Similarly, the AUC measurement puts RF and GBT in the first rank, whereas LR and FLM are in the second rank.

The results once more (as previously explained, the importance of the reduct where classifiers gave high accuracy) show the importance of the reduct generated by the FBWV since it mostly gives high F-measure and AUC values compared with those of other reducts.

Table 11. F-measure rates for the reducts of the nine datasets given by the four classifiers.

6	FSM	Austra	Heart Disease	Phishing	Sonar	Iono	SGC	SpamBase	Messidor	Pop-Failure	Priority
LR	**FBWV**	0.8649	0.8629	0.9340	0.7376	0.8491	0.8276	0.8520	0.3809	**0.9794**	2
	RS	0.8624	0.7402	0.9393	0.6400	0.8304	0.7918	0.4199	0.6815	0.9559	3
	WGFS	0.7615	0.8052	0.7154	0.6401	0.7879	0.8127	0.7698	0.6833	0.9559	4
	ScC	0.8896	0.9244	0.9186	0.7167	0.8546	0.7886	0.9056	0.6931	0.9559	1
FLM	**FBWV**	0.8649	0.9004	0.9319	0.7742	0.8575	0.8247	0.8573	0.5102	0.9929	1
	RS	0.7188	0.7123	0.9387	0.6167	0.8246	0.7858	0.0957	0.6978	0.9559	2
	WGFS	0.6949	0.7494	0.7154	0.4189	0.7879	0.8118	0.7741	0.7218	0.9559	2
	ScC	0.8029	0.8938	0.9247	0.7493	0.8390	0.7886	0.8647	0.6974	0.9623	2
RF	**FBWV**	0.8624	0.8613	0.9350	0.7409	0.9650	0.8188	0.8837	0.5749	0.9559	1
	RS	0.8624	0.7531	0.9321	0.7629	0.9549	0.8247	0.8124	0.5621	0.9559	2
	WGFS	0.7224	0.7665	0.7154	0.4871	0.8902	0.8247	0.7874	0.6313	0.9559	3
	ScC	0.8425	0.8712	0.9295	0.7376	0.9275	0.8247	0.9048	0.4442	0.9559	2
GBT	**FBWV**	0.8729	0.8390	0.9667	0.8267	0.9707	0.8358	0.8678	0.7310	0.9752	1
	RS	0.8543	0.7613	0.9737	0.7200	0.9412	0.8247	0.7545	0.7473	0.9559	2
	WGFS	0.7605	0.7826	0.7154	0.6834	0.8775	0.8218	0.7632	0.7162	0.9559	3
	ScC	0.8659	0.8339	0.9414	0.6993	0.9071	0.7886	0.6580	0.7152	0.9559	3

Table 12. 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	Priority
LR	**FBWV**	0.9159	0.9032	0.9791	0.8135	0.8144	0.7612	0.9070	0.8196	0.9161	2
	RS	0.8514	0.7429	0.9804	0.7183	0.7824	0.4911	0.9358	0.7471	0.3879	3
	WGFS	0.7633	0.8735	0.5015	0.6954	0.5582	0.6146	0.9209	0.7871	0.4640	4
	ScC	0.9231	0.9232	0.9684	0.7689	0.8220	0.6974	0.9482	0.8341	0.8668	1
FLM	**FBWV**	0.9143	0.8845	0.9794	0.8276	0.8198	0.6079	0.90.34	0.8067	0.9333	2
	RS	0.6828	0.7074	0.9802	0.6297	0.7714	0.4945	0.9180	0.7402	0.3618	3
	WGFS	0.6739	0.7939	0.5015	0.6617	0.3780	0.6247	0.8868	0.7984	0.3266	4
	ScC	0.8594	0.9039	0.9689	0.7917	0.8418	0.6974	0.9247	0.8321	0.8673	1
RF	**FBWV**	0.9056	0.9238	0.9845	0.8197	0.9692	0.7335	0.9078	0.7146	0.8466	1
	RS	0.8282	0.7181	0.9818	0.7649	0.9626	0.4903	0.8261	0.6060	0.4776	3
	WGFS	0.7483	0.8165	0.5015	0.7311	0.9165	0.6190	0.8134	0.6429	0.5984	3
	ScC	0.8992	0.8911	0.9743	0.7971	0.8866	0.6784	0.9431	0.7143	0.8760	2
GBT	**FBWV**	0.9128	0.8782	0.9937	0.8603	0.9692	0.7315	0.9034	0.7814	0.8981	1
	RS	0.8696	0.7063	0.9959	0.7871	0.9637	0.5490	0.8114	0.7759	0.5967	2
	WGFS	0.6882	0.8266	0.4985	0.7419	0.8741	0.6643	0.7610	0.7176	0.5722	3
	ScC	0.9084	0.8289	0.9782	0.8335	0.9078	0.6974	0.9406	0.7650	0.8879	2

Figure 2 shows the accuracy comparison of the four classifiers for each of the nine reducts generated by the four reduction algorithms. For each FSA, four columns are shown that represent the accuracy of the classifiers (LR in blue, FLM in red, RF in green, and GBT in purple) applied to the reduced dataset. The same accuracies were found for the original dataset to justify the comparisons. The F-measure and AUC results are graphically presented in Figures 3 and 4, respectively. For each FSA, nine columns were drawn that represent the F-measure (Figure 3) and AUC (Figure 4), where each column represents one dataset. The columns starting from the left of each FSA represent Austra, Heart Disease, Phishing, Sonar, Iono, SGC, SpamBase, Messidor, and Pop-Failure, respectively.

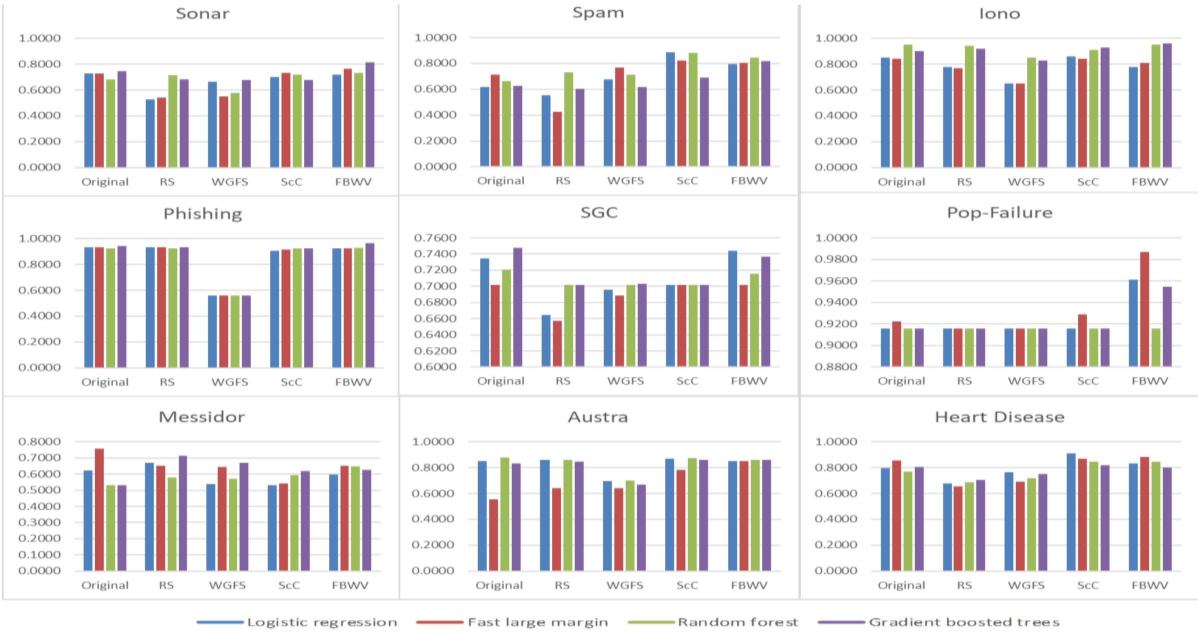

Figure 2. The accuracy rate results.

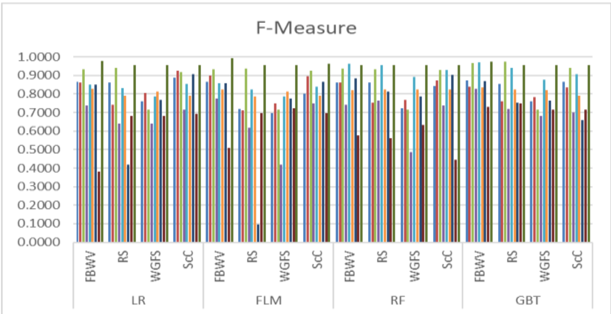

Figure 3. F-measure results.

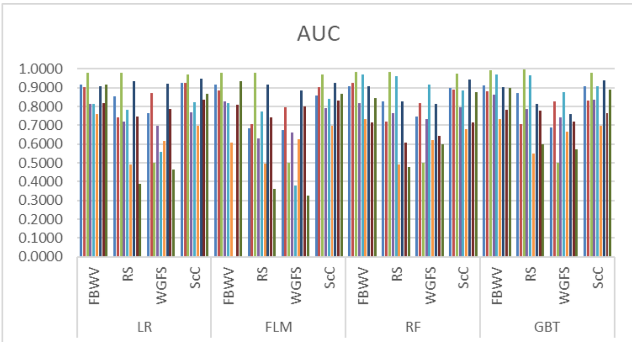

Figure 4. AUC results.

6. Conclusions and Future Work

This research introduces a new novel feature selection algorithm named the FBWV. It is a successive forward selection and backward elimination algorithm based on a chi-square weighted vector and a custom threshold value. The result represents the minimal set of the most important features. It achieves high performance in terms of accuracy, reduction rate, F-measure, and AUC measures. The FBWV was compared with three state-of-the-art feature selection algorithms, namely, ScC, WG, and RS. Nine datasets were used, and reducts were generated and trained by LR, FLM, RF, and GBT. The performance results of all of them were reported. A comparison of all the results shows the merit of the proposed algorithm over the others.

The findings of this work are represented by the following points:

- The FBWV is a pioneer feature selection algorithm that introduces the best minimal reduct with high accuracy.
- The accuracy of FBWV reduct is highly competitive with that of the other reducts generated by the other three feature reduction algorithms.
- Rather than the accuracy, the F-measure and AUC performed better for the reduct generated by the FBWV than for the other reducts.

The overall results emphasize the importance of the FBWV and its superiority over the other algorithms.

Finally, the FBWV is promising for feature selection, and it provides good support for other tracks of science, including AI, data science, machine learning, and the IoT.

The scope of this research is binary classification, so the FBWV was tested on datasets of two. This research can be extended by the development of algorithms that are able to work with multiple classes and unstructured datasets. Moreover, future work could include another list of feature selection methods for comparison.

Acknowledgment

The author gratefully acknowledges the financial assistance provided by the Research Sector, The Arab

Open University-Kuwait Branch under Decision No. 24045.

References

[1] Aggarwal N., Shukla U., Saxena G., Rawat M., Bafila A., Singh S., and Pundir A., "Mean Based Relief: An Improved Feature Selection Method Based on ReliefF," *Applied Intelligence*, vol. 53, pp. 23004-23028, 2023. DOI:10.1007/s10489-023-04662-w

[2] Ali W. and Saeed F., "Hybrid Filter and Genetic Algorithm-Based Feature Selection for Improving Cancer Classification in High-Dimensional Microarray Data," *Processes*, vol. 11, no. 2, pp. 1-22, 2023. https://doi.org/10.3390/pr11020562

[3] Al-Shalabi L., "Hybrid Feature Selection based ScC and Forward Selection Methods," *International Journal of Data and Network Science*, vol. 8, no. 2, pp. 1117-1128, 2024. DOI:10.5267/j.ijdns.2023.11.022

[4] Al-Shalabi L., "New Feature Selection Algorithm Based on Feature Stability and Correlation," *IEEE Access*, vol. 10, pp. 4699-4713, 2022. DOI:10.1109/ACCESS.2022.3140209

[5] Al-Shalabi L., "Rough Set-Based Reduction of Incomplete Medical Datasets by Reducing the Number of Missing Values," *The International Arab Journal of Information Technology*, vol. 16, no. 2, pp. 203-210, 2019. https://iajit.org/portal/PDF/March%202019,%20No.%202/11460.pdf

[6] Al-Shalabi L., Shaaban Z., and Kasasbeh B., "Data mining: A Preprocessing Engine," *Journal of Computer Science*, vol. 2, no. 9, pp. 735-739, 2006. DOI:10.3844/jcssp.2006.735.739

[7] Alsini R., Naz A., Khan H., Bokhari A., Daud A., and Ramzan M., "Using Deep Learning and Word Embeddings for Predicting Human Agreeableness Behavior," *Scientific Reports*, vol. 14, no. 29875, pp. 1-19, 2024. https://doi.org/10.1038/s41598-024-81506-8

[8] Aslam M. and Smarandache F., "Chi-Square Test for Imprecise Data in Consistency Table," *Frontiers in Applied Mathematics and Statistics*, vol. 9, pp. 1-5, 2023. DOI:10.3389/fams.2023.1279638

[9] Bermejo P., Gamez J., and Puerta J., "Adapting the CMIM Algorithm for Multilabel Feature Selection. A Comparison with Existing Methods," *Expert Systems*, vol. 35, no. 1, pp. e12230, 2017. https://doi.org/10.1111/exsy.12230

[10] Bolon-Canedo V., Sanchez-Marono N., and Alonso-Betanzos A., "A Review of Feature Selection Methods on Synthetic Data," *Knowledge Information Systems*, vol. 34, no. 3, pp. 483-519, 2013. https://doi.org/10.1007/s10115-012-0487-8

[11] Bommert A., Sun X., Bischl B., Rahnenfuhrer J., and Lang M., "Benchmark for Filter Methods for Feature Selection in High-Dimensional Classification Data," *Computational Statistics and Data Analysis*, vol. 143, pp. 106839, 2020. https://doi.org/10.1016/j.csda.2019.106839

[12] Borboudakis G. and Tsamardinos I., "Forward-Backward Selection with Early Dropping," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1-39, 2019. https://jmlr.org/papers/volume20/17-334/17-334.pdf

[13] Breiman L., "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324

[14] Chaudhary A., Kolhe S., and Rajkamal, "Performance Evaluation of Feature Selection Methods for Mobile Devices," International *Journal of Engineering Research and Applications*, vol. 3, no. 6, pp. 587-594, 2013. https://www.ijera.com/papers/Vol3_issue6/CW36587594.pdf

[15] Cherrington M., Thabtah F., Lu J., and Xu Q., "Feature Selection: Filter Methods Performance Challenges," *in Proceedings of the International Conference on Computer and Information Sciences*, Sakaka, pp. 1-4, 2019. DOI:10.1109/ICCISci.2019.8716478

[16] Chowdhury M. and Turin T., "Variable Selection Strategies and its Importance in Clinical Prediction Modelling," *Fam Med Com Health*, vol. 8, no.1, pp. 1-7, 2020. DOI:10.1136/fmch-2019-000262

[17] Christen P., Hand D., and Kirielle N., "A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1-24, 2023. https://doi.org/10.1145/3606367

[18] Cilia N., D'Alessandro T., Stefano C., Fontanella F., and Di Freca A., "Comparing Filter and Wrapper Approaches for Feature Selection in Handwritten Character Recognition," *Pattern Recognition Letters*, vol. 168, pp. 39-46, 2023. https://doi.org/10.1016/j.patrec.2023.02.028

[19] Cormen T., Leiserson C., Rivest R., and Stein C., *Introduction to Algorithms*, MIT Press, 2009. https://enos.itcollege.ee/~japoia/algorithms/GT/Introduction_to_algorithms-3rd%20Edition.pdf

[20] Das A., Sengupta S., and Bhattacharyya S., "A Group Incremental Feature Selection for Classification Using Rough Set Theory based Genetic Algorithm," *Applied Soft Computing*, vol. 65, pp. 400-411, 2018.

[21] Frederick O., Maxwell O., Ifunanya O., Udochukwu E., Kelechi O., Ngonadi L., and Idris H., "Comparison of Some Variable Selection Techniques in Regression Analysis," *American Journal of Biomedical Science and Research*, vol.

6, no. 4, pp. 281-293, 2019. https://doi.org/10.3389/fams.2023.1279638

[22] Guyon I. and Elisseeff A., "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.

[23] Hall M., Feature Selection for Discrete and Numeric Class Machine Learning, Technical Report, 1999. https://ml.cms.waikato.ac.nz/publications/1999/99MH-Feature-Select.pdf

[24] Haq A., Zeb A., Lei Z., and Zhang D., "Forecasting Daily Stock Trend Using Multi-Filter Feature Selection and Deep Learning," *Expert Systems with Applications*, vol. 168, pp. 114444, 2021. https://doi.org/10.1016/j.eswa.2020.114444

[25] Hoque N., Singh M., and Bhattacharyya D., "EFS-MI: An Ensemble Feature Selection Method for Classification," *Complex Intelligent System*, vol. 4, no. 2, pp. 105-118, 2017. https://doi.org/10.1007/s40747-017-0060-x

[26] Hu M., Tsang E., Guo Y., and Xu W., "Fast and Robust Attribute Reduction Based on the Separability in Fuzzy Decision Systems," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 5559-5572, 2022. DOI:10.1109/TCYB.2020.3040803

[27] Huanhuan G., Yanying L., Jiaoni Z., Baoshuang Z., and Xialin W., "A New Filter Feature Selection Algorithm for Classification Task by Ensembling Pearson Correlation Coefficient and Mutual Information," *Engineering Applications of Artificial Intelligence*, vol. 131, pp. 107865, 2024. https://doi.org/10.1016/j.engappai.2024.107865

[28] Jenifa G., Padmapriya K., Sevanthi P., Karthika K., Pandi V., and Arumugam D., "An Effective Personality Recognition Model Design using Generative Artificial Intelligence based Learning Principles," *in Proceedings of the International Conference on Computing and Data Science*, Chennai, pp. 1-6, 2024. DOI:10.1109/ICCDS60734.2024.10560368

[29] Ke W., Wu C., Wu Y., and Xiong N., "A New Filter Feature Selection Based on Criteria Fusion for Gene Microarray Data," *IEEE Access*, vol. 6, pp. 61065-61076, 2018. DOI:10.1109/ACCESS.2018.2873634

[30] Lazar C., Taminau J., Meganck S., Steenhoff D., Coletta A., Molter C., De Schaetzen V., Duque R., Bersini H., and Nowe A., "A Survey on Filter Techniques for Feature Selection in Gene Eexpression Microarray Analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 4, pp. 1106-1119, 2012. DOI:10.1109/TCBB.2012.33

[31] Lu J., Zhou S., Feng T., and Tong Z., "A Chi-Squared Approximation-based Weighting Method for Distributed Detection," *in Proceedings of the IEEE Radar Conference*, New York, pp. 1-6, 2022. DOI:10.1109/RadarConf2248738.2022.9764160

[32] Maseno E. and Wang Z., "Hybrid Wrapper Feature Selection Method based on Genetic Algorithm and Extreme Learning Machine for Intrusion Detection," *Journal of Big Data*, vol. 11, pp. 1-25, 2024. https://doi.org/10.1186/s40537-024-00887-9

[33] Mohtasham F., Pourhoseingholi M., Hashemi Nazari S., Kavousi K., and Reza Zali M., "Comparative Analysis of Feature Selection Techniques for COVID-19 Dataset," *Scientific Reports*, vol. 14, pp. 1-20, 2024. DOI:10.1038/s41598-024-69209-6

[34] Mohtashami M. and Eftekhari M., "A Hybrid Filter-Based Feature Selection Method via Hesitant Fuzzy and Rough Sets Concepts," *Iranian Journal of Fuzzy Systems*, vol. 16, no. 2, pp. 165-182, 2019. DOI: 10.22111/ijfs.2019.4550

[35] Natekin A. and Knoll A., "Gradient boosting Machines, A Tutorial," *Frontiers in Neurorobotics*, vol. 7, pp. 1-21, 2013. DOI:10.3389/fnbot.2013.00021

[36] Ngo N., Michel P., and Giorgi R., "Multivariate Filter Methods for Feature Selection with the γ-Metric," *BMC Medical Research Methodology*, vol. 24, no. 307, pp. 1-22, 2024. DOI:10.1186/s12874-024-02426-9

[37] Parmar M., Sonker A., and Sejwar V., "A Comparative Analysis for Filter-Based Feature Selection Techniques with Tree-based Classification," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 10, pp. 360-369, 2023. https://doi.org/10.17762/ijritcc.v11i10s.7643

[38] Parodi S., Verda D., Bagnasco F., and Muselli M., "The Clinical Meaning of the Area Under a Receiver Operating Characteristic Curve for the Evaluation of the Performance of Disease Markers," *Epidemiology Health*, vol. 44, pp. 1-10, 2022. https://doi.org/10.4178/epih.e2022088

[39] Patel D., Saxena A., and Wang J., "A Machine Learning-Based Wrapper Method for Feature Selection," *International Journal of Data Warehousing and Mining*, vol. 20, no. 1, pp. 1-33, 2024. https://doi.org/10.4018/IJDWM.352041

[40] Peng C., Lee K., and Ingersoll G., "An Introduction to Logistic Regression Analysis and Reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3-14, 2002. https://doi.org/10.1080/00220670209598786

[41] Pirgazi J., Kallehbasti M., Sorkhi A., and Kermani A., "An Efficient Hybrid Filter-Wrapper Method based on Improved Harris Hawks Optimization for Feature Selection," *BioImpacts*, vol. 15, pp. 30340, 2025. DOI:10.34172/bi.30340

[42] Pourpanah F., Shi Y., Lim C., Hao Q., and Tan C., “Feature Selection Based on Brain Storm Optimization for Data Classification,” *Applied Soft Computing Journal*, vol. 80, pp. 761-775, 2019. https://doi.org/10.1016/j.asoc.2019.04.037

[43] Qian Y., Liang X., Wang Q., Liang J., Liu B., Skowron A., Yao Y., Ma J., and Dang C., “Local Rough Set: A Solution to Rough Data Analysis in Big Data,” *International Journal of Approximate Reasoning*, vol. 97, pp. 38-63, 2018. https://doi.org/10.1016/j.ijar.2018.01.008

[44] Quinlan J., “Unknown Attribute Values in Induction,” *in Proceedings of the 6th International Workshop on Machine Learning*, New York, pp. 164-168, 1989. https://doi.org/10.1016/B978-1-55860-036-2.50048-5

[45] Rebbah F., Chamlal H., and Ouaderhman T., “Accurate Analysis for Univariate-Based Filter Methods for Microarray Data Classification,” *Journal of Algorithms and Computational Technology*, vol. 18, pp. 1-11, 2024. DOI:10.1177/17483026241232295

[46] Ringsquandl M., Lamparter S., Brandt S., Hubauer T., and Lepratti R., “Semantic-Guided Feature Selection for Industrial Automation Systems,” *in Proceedings of the the Semantic Web-ISWC, Lecture Notes in Computer Science*, Bethlehem, pp. 225-240, 2015. https://doi.org/10.1007/978-3-319-25010-6_13

[47] Saeidi S., “Identifying Personality Traits of WhatsApp Users Based on Frequently Used Emojis Using Deep Learning,” *Multimedia Tools and Applications*, vol. 83, pp. 13873-13886, 2024. DOI:10.1007/s11042-023-15209-z

[48] Sinayobye J., Kaawaase K., Kiwanuka F., and Musabe R., “Hybrid Model of Correlation Based Filter Feature Selection and Machine Learning Classifiers Applied on Smart Meter Data Set,” *in Proceedings of the IEEE/ACM Symposium on Software Engineering in Africa*, Montreal, pp. 1-10, 2019. DOI:10.1109/SEiA.2019.00009

[49] Soheili M. and Moghadam A., “DQPFS: Distributed Quadratic Programming-Based Feature Selection for Big Data,” *Journal of Parallel and Distributed Computing*, vol. 138, pp. 1-14, 2020. https://doi.org/10.1016/j.jpdc.2019.12.001

[50] Velayutham C. and Thangavel K., “Unsupervised Quick Reduct Algorithm Rough Set Theory,” *Journal of Electronic Science and Technology*, vol. 9, no. 3, pp. 193-201, 2011.

[51] Velusamy K. and Manavalan R., “Performance Analysis of Unsupervised Classification based on Optimization,” *International Journal of Computer Applications*, vol. 42, no. 19, pp. 22-27, 2012. DOI:10.5120/5801-8090

[52] Wah Y., Ibrahim N., Hamid H., Abdul-Rahman S., and Fong S., “Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximizing Classification Accuracy,” *Pertanika Journal of Science and Technology*, vol. 26, no. 1, pp. 329-340, 2018. https://www.researchgate.net/publication/322920304_Feature_selection_methods_Case_of_filter_and_wrapper_approaches_for_maximising_classification_accuracy

[53] Wisniewski G. and Yvon F., “Fast Large-Margin Learning for Statistical Machine Translation,” *International Journal of Computational Linguistics and Applications*, vol. 4, no. 2, pp. 45-62, 2013.

[54] Xue B., Zhang M., and Browne W., “A Comprehensive Comparison on Evolutionary Feature Selection Approaches to Classification,” *International Journal of Computational Intelligence and Applications*, vol. 14, no. 2, pp. 1-22, 2015. https://doi.org/10.1142/S146902681550008X

[55] Zhang Y., Gong D., Hu Y., and Zhang W., “Feature Selection Algorithm Based on Bare Bones Particle Swarm Optimization,” *Neurocomputing*, vol.148, pp. 150-157, 2015. https://doi.org/10.1016/j.neucom.2012.09.049

[56] Zhu Z., Ong Y., and Dash M., “Wrapper-Filter Feature Selection Algorithm Using a Memetic Framework,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37, no. 1, pp. 70-76, 2007. DOI:10.1109/TSMCB.2006.883267

Luai Al-Shalabi is an Associate Professor of data mining at Arab Open University, Kuwait Branch. He completed his PhD in computer science in 2000 with a focus on data mining. His areas of interest include data mining, data science, knowledge discovery, and machine learning. He has many publications in reputable local and international conferences and journals, mostly on data mining and machine learning.