

Development of Data Mining Models Based on Features Ranks Voting (FRV)

Mofreh A. Hogo*

Department of Electrical Engineering, Benha Faculty of Engineering, Benha University, Benha, 13518, Egypt

*Corresponding Author: Mofreh A. Hogo. Email: mhogo@ut.edu.sa

Received: 14 January 2022; Accepted: 08 March 2022

Abstract: Data size plays a significant role in the design and the performance of data mining models. A good feature selection algorithm reduces the problems of big data size and noise due to data redundancy. Features selection algorithms aim at selecting the best features and eliminating unnecessary ones, which in turn simplifies the structure of the data mining model as well as increases its performance. This paper introduces a robust features selection algorithm, named Features Ranking Voting Algorithm FRV. It merges the benefits of the different features selection algorithms to specify the features ranks in the dataset correctly and robustly; based on the feature ranks and voting algorithm. The FRV comprises of three different proposed techniques to select the minimum best feature set, the forward voting technique to select the best high ranks features, the backward voting technique, which drops the low ranks features (low importance feature), and the third technique merges the outputs from the forward and backward techniques to maximize the robustness of the selected features set. Different data mining models were built using obtained selected features sets from applying the proposed FRV on different datasets; to evaluate the success behavior of the proposed FRV. The high performance of these data mining models reflects the success of the proposed FRV algorithm. The FRV performance is compared with other features selection algorithms. It successes to develop data mining models for the Hungarian CAD dataset with Acc. of 96.8%, and with Acc. of 96% for the Z-Alizadeh Sani CAD dataset compared with 83.94% and 92.56% respectively in [48].

Keywords: Evaluator; features selection; data mining; forward; backward; voting; feature rank

1 Introduction

The performance of the data mining model (degree of manipulating and/or discovering useful information) is highly affected by the dataset size (the number of instances and the number of features within). The development of data mining models faces different challenges related to datasets as, the size of the collected dataset from different sources of information, the existence of noisy attributes (e.g., features redundancies), and the missed data (values of some features due to the possible faults



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in the real data acquisition systems). On the other side; the performance of the data mining model can be significantly increased by eliminating these challenges to minimize their effects. The feature selection technique is one of the most important methods which discover and select the most important features. The necessity and utilization of feature selection algorithms and techniques have been exhaustively studied and presented in many types of research. The feature selection process utilizes different algorithms comprises Binary Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Integrated Square Error (ISE), Hybrid Feature Selection Methods (HFSM), Efficient Semi-Supervised Feature Selection (ESFS), Ranking Obtained from Aggregating Several Features Selection Methods, Meaning Based Feature Selection (MBFS) approaches, Sequential Forward and Sequential Backward Optimization Techniques, and many others. All feature selection algorithms' goal is to reduce the size of the dataset as well as to eliminate unwanted features.

The problems of selecting the best feature selection algorithm can be summarized as: No one feature selection algorithm is suitable for all datasets types, a specific feature in a dataset may be ranked differently (degree of its importance compared with other features) by different feature algorithms, and the lack of robustness of the feature selection algorithm. This paper introduces a good feature selection algorithm that avoids these problems, the paper's idea is to develop a hybrid technique that utilizes different feature selection algorithms to increase the performance of selecting the best feature set. The hybrid technique merges the benefits of the different algorithms to maximize the robustness as well as to select the most important feature set. Another issue is to specify the feature rank in the dataset correctly and consistently.

The idea of the presented work is to find the most robust rank of each feature in a dataset, which can be obtained by merging the different ranks of the feature from using different feature ranking algorithms and to find the most correct rank of the feature using the voting technique. The paper proposes the Feature Ranking Voting Algorithm (FRV).

The goal of this work is to introduce a robust feature selection algorithm with high performance, by hybrid and merging the advantages of different feature selection algorithms using the feature ranks (feature ranks will be obtained from a specific number of different feature ranking algorithms) and voting algorithm which will be utilized to specify the features ranks correctly.

The rest of the paper is organized as Section 2 is reserved for the literature review. The theories overview and the proposed FRV algorithm are presented in Section 3. Section 4 is reserved for the description of the experimental design. Results from analysis and interpretation are presented in Section 5, and Section 6 is reserved for the conclusion and future extension.

2 Literature Review

The review of different feature selection algorithms and techniques will be presented and discussed in this section. The review presentation will be divided into six different categories of feature selection algorithms.

2.1 Binary Particle Swarm Optimization Methods (PSO)

A novel binary PSO with the integration of a new neighbor selection strategy to solve the feature selection problem in text clustering as presented in [1]. It introduced a new updating strategy to learn from the neighbor's best position to enhance the performance of the text clustering algorithm, but its performance was not significantly higher than other algorithms. An unsupervised learning algorithm is presented in [2] to solve the text feature selection problem using particle swarm optimization (PSO) and

the term frequency-inverse document frequency (TF-IDF) as an objective function to evaluate each text feature at the level of the document, the performance was relatively as same as other algorithms. A multiple swarm heterogeneous BPSO based on BPSO and three modified BPSO algorithms named BoPSO, SAHBPSO, and CSAHBPSO, is presented in [3] the performance is improved with a very small percentage when utilized to build real models. A binary PSO feature selection algorithm was implemented in [4], to select the most important features from high dimensional gene expression data, aiming at removing the redundant features and keeping the distinctive capability using the K-NN classifier to compare the results obtained from different datasets, the algorithm was not robust. Several feature selection techniques as genetic algorithm, greedy algorithms, best-first search algorithms, and PSO algorithm were utilized to achieve effective user authentications as presented in [5] to select the most important features to build the user's pattern in minimal time, but their performance was insignificantly improved. A modified particle swarm optimization (MPSO) feature selection algorithm was used to build a face recognition system as presented in [6] based on the coefficients extracted by Discrete Wavelet Transform (DWT). The implementation of a faster and more accurate model for predicting Parkinson's disease early using the filter and Wrapper technique to reduce the number of features is presented in [7] for assisting doctors to focus on the most important medical measurements.

2.2 Genetic Algorithms Methods (GA)

A robust genetic algorithm (GA)-based feature selection methodology to achieve high classification accuracy for predicting the quality of life for lung transplant patients is presented in [8]. Using three decision-analytic models, GA-KNN, GA-ANN, and GA-SVM, to solve the problem of feature selection for lung transplants. The obtained results were good but were not enough to build a robust model. The Automatic feature group combination selection method based on GA for the functional regions clustering in DBS is presented in [9]. Adaptive intrusion detection via GA-GOGMM-based pattern learning with fuzzy rough set-based attribute selection is implemented in [10]. The implementation of feature selection and analysis algorithms were utilized to build a classifier for the Android malware detection system was presented in [11] to defend the mobile user from malicious threats and to build a high-performance classifier model. The development of a face recognition system based on local feature selection and Naïve Bayesian classification was presented in [12]. A feature selection algorithm based on correlations of gene-expression (SVMRFE) with minimum Pearson correlation for predicting cancer is implemented in [13] to select genes that are both relevant for classification and fewer attributes to produce a system with high prediction accuracy. Features selection methods for analyzing an essay of Chinese, English learners, and text features such as words, phrases, paragraphs, and chapters are presented in [14] to extract important features and build a model, for distinguishing high-quality essays from others, it effectively solves the problem of low differentiation scores in automated essay scoring system. Genetic programming (GP) algorithm is developed for selecting fewer features and optimizing features space as presented in [15] determined the proper patterns of intracranial electroencephalogram (IEEG) and functional magnetic resonance imaging (fMRI) signals, which was more effective than conventional feature selection algorithms. Several feature selection methods for biodegradation process evaluation are presented in [16] selected the optimal features using Genetic Algorithms and a weighted Euclidean distance. A random five features with the Genetic concept are applied in the field of text summarization and classification as presented in [17,18]. A feature selection algorithm for profiling the statistical features of qualitative features through an ECG waveform is implemented in [19] to differentiate abnormal heartbeats and normal (NORM) for the detection of worms in networking applications for finding the frequently repeated strings in a packet stream. A feature selection algorithm called Meaning Based Feature

Selection (MBFS) based on the Helmholtz principle of the Gestalt theory of human perception was applied in the area of image processing and presented in [20] is used in image processing and text mining systems to improve their accuracy and simplicity.

2.3 Hybrid Features Selection Methods

A system for Credit scoring IGDFS was implemented in [21] based on the Information Gain and Wrapper technique, using three different classical decision-making models of KNN, NB, and SVM to select the best features for the credit scoring problem, but the classification accuracy was highly sensitive to the dataset type and size, the rate of positive and negative samples in the dataset, and its robustness was very low. A hybrid feature selection method as presented in [22] two features subsets are obtained by applying two different filters, then a feature weight-based union operation is introduced to merge the obtained feature subsets, and finally, the hierarchical agglomerative clustering algorithm produces clusters, the performance was low. A hybrid feature selection algorithm was implemented in [23] using a re-sampling method that combines an oversampling and an under-sampling technique for binary classification problems. A novel hybrid feature selection strategy in quantitative analysis of laser-induced breakdown spectroscopy is implemented in [24] constructed different credit scoring combinations by selecting features with three approaches are, linear discriminate analysis (LDA), rough sets (RST), and F-score approaches) optimized features space by removing both irrelevant and redundant features. A hybrid feature selection technique (HBFS) as presented in [25] selected the important features according to their labels and eliminated the redundancy between relevant features to reduce the complexity. A hybrid comprised of four feature selection algorithms was applied in the area of credit scoring to improve the banks' performance in credit scoring problems as presented in [26] utilized four feature selection algorithms including PCA, GA, information gain ratio, and relief attribute evaluation function.

2.4 Semi-Supervised Feature Selection Methods (ESFS)

Feature selection technique using an efficient semi-supervised feature selection (ESFS) is presented in [27] applied on VHR remote sensing images, using a two-stage feature selection method; the first is weight matrix fully integrates local geometrical structure and discriminative information and the second is weight matrix incorporated into a norm minimization optimization problem of data reconstruction, to objectively measure the effectiveness of features for a thematic class. The use of the feature selection algorithm to find the correct pattern recognition models by reducing the uninformative features in portraits of biomedical signals was presented in [28] used supervised learning and the technology of revealing and the implementation of reserves to increase the classification system accuracy. The application of a feature selection algorithm for detecting microchips with faults to solve the semiconductor industry problem of detecting microchips is introduced in [29]. Profiling of a statistical feature selection to improve the accuracy of the automatic arrhythmia-diagnosis system was presented in [30] selected qualitative features to classify abnormal and normal heartbeats, left bundle branch block, right bundle branch block, ventricular premature contractions, and atrial premature contractions come under abnormal heartbeats. The utilization of a feature selection method to build a model for optimal predicting of the electricity price is presented in [31] to help in the decision-making of future maintenance scheduling of the power plants, risk management, plan future contracts, purchasing raw materials, and determine the market pricing. The selection of the important features from images of multi-camera tracking in public places is presented in [32] tracked moving objects across distributed cameras that provide the most optimal trade-off accuracy and speed based on color, texture, and edge features. The identification of mammogram images using the fuzzy feature

selection approach is implemented in [33] utilized fuzzy curve and fuzzy surface selected feature from mammogram images that isolated the significant features from the original features according to their significance, eliminated unwanted features, reduced the feature space dimension, and built a simplified classification scheme. The feature selection approach for fall detection is introduced in [34] reduced the effect of the fall' problem for old people which results in dangerous consequences even death.

2.5 Ranking Features Selection Methods

A generic framework for matching and ranking trustworthy context-dependent services is presented in [35] used logic and set theory, the performance was better compared with other research, but with a small percentage. Two different criteria for evaluating the ranks of features were implemented in [36] a single feature and synergistic feature technique to measure the difference of densities between two datasets. Hybrid feature selection methods with voting feature intervals and statistical feature selection methods based on feature variance were implemented in [37]. The ranking obtained from aggregating several feature selection methods as a filter-wrapper algorithm is implemented in [38] called FSSO-BOP, which used a filter-based consensus ranking with ties to guide the posterior wrapper phase. Two feature selection methods with separability and thresholds algorithms, classification, and regression tree for developing urban impervious surface extraction are implemented in [39].

2.6 Sequential Forward and Sequential Backward Methods

The sequential forward feature selection algorithm is implemented in [40] proposed a new set of muscle activity based on features for facial expression recognition, to choose the most important set of muscles for recognition of basic facial expressions. An optimal feature selected sequence is implemented in [41] used Markov decision process and dynamic programming (DP), the sequence described the order that features selected for classification to learn a strategy for generating the orders only with the feedback of circumstance. Gender classification from NIR Images by using Quadrature Encoding Filters of the most relevant features is developed in [42] extracted texture-based features comprising Local Binary Patterns (LBP) and Histogram of Oriented Gradient (HOG) of face images from the FERET face database. An efficient feature selection-based Bayesian and Rough set approach for intrusion detection presented in [43] aimed at improving the performance of the intrusion detection system.

From the listed survey of the related work we can conclude the problem, the goal, and the proposed solution of this work as follows:

The problems related to the feature selection algorithms can be summarized with the following 3 points as follows:

- The different feature selection algorithms select a different set of features for the same dataset and different ranks the same feature in the same dataset with different ranks.
- No one optimal feature selection algorithm can select the best features from all different datasets.
- The single feature selection algorithm has a lack of robustness.

The proposed voting algorithm comprises of three different voting techniques to select the minimum best feature set, the first is the forward voting technique selects the best features according to their high voted ranks, the second is the backward voting technique which drops the worst feature with low voted ranks, and the third voting technique merges the outputs from the forward and backward techniques to maximize the robustness of the selected features set.

The phases of the proposed work are listed below:

- The extraction of the feature' ranks using different feature ranking algorithms.
- The applying the voting technique to categorize the features according to their voting into high rank and low ranks features.
- The applying of the forward voting technique to select the best features according to their higher voted ranks.
- The applying of the backward voting technique to drop the worst feature with low voted ranks.
- The merging technique merges the outputs from the forward and backward techniques to maximize the robustness of the selected feature.
- The testing of the selected set of features by utilizing them to build a data mining model and measure its performance.

3 The Proposed FRV Algorithm

This section presents the main phases of the proposed Feature Ranking Voting algorithm FRV. The FRV of three different proposed techniques to select the minimum best feature set, the forward voting technique selects the best features according to their high voted ranks, the backward voting technique drops the worst feature with low voted ranks, and the third technique merges between the outputs from the forward and backward techniques to maximize the robustness of the selected features. This section presents the forward voting technique, the backward voting technique, and the merges technique between the outputs from the forward and backward techniques to maximize the robustness of the selected features.

The proposed Feature Rank Voting algorithm (FRV) depends mainly on the computation of the feature ranks using different feature ranking algorithms, the application of the voting techniques which categorize the feature set into the High-rank feature, and Low-rank feature. After these two phases, the 3 different techniques will be applied forward, backward, and merge techniques.

The proposed Feature Ranking Voting algorithm FRV consists of the main five steps:

Using various feature classification algorithms, compute the feature ranks. The highest and lowest functions are chosen depending on their ranks. The use of forward-feature filtering on features with the highest voting rankings Backward function filtering algorithms are used on features with the lowest voting ranks. The combination of the two subsets of features is chosen in measures 3 and 4. The phases of the proposed FRV are shown in Fig. 1. The steps of FRV are summarized as follows:

Let $D = \{x_1, x_2, \dots, x_n\}$ be a dataset consisting of n examples or instances. Each instance $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ is a vector of m values, where each value x_{ij} of this vector represents a feature of that instance. The vector $f_j = (x_{1j}, x_{2j}, \dots, x_{nj})^t$ is the vector of values of a feature f_j .

Let F is the specified number of feature selection algorithms from 1 to k (for example if five algorithms will be used, then $k = 5$ and $F:1$ to 5)

Step 1, 2: apply the algorithms for the selection of features from 1 to 5 on the D -dataset; each algorithm will output various ranges of features ranks as f_{Rjk} (each algorithm for the features selection provides different weights on a given basis) Where R is the rank of each feature obtained from the feature selection algorithm, i is feature order in D , k is the k^{th} feature selection algorithm $k:1$ to 5.

Whenever the D is made up of 20 different features with 5 different algorithms for feature selection, the results of this stage are shown by a 2-D feature array expressed by $R = [20 \times 5]$.

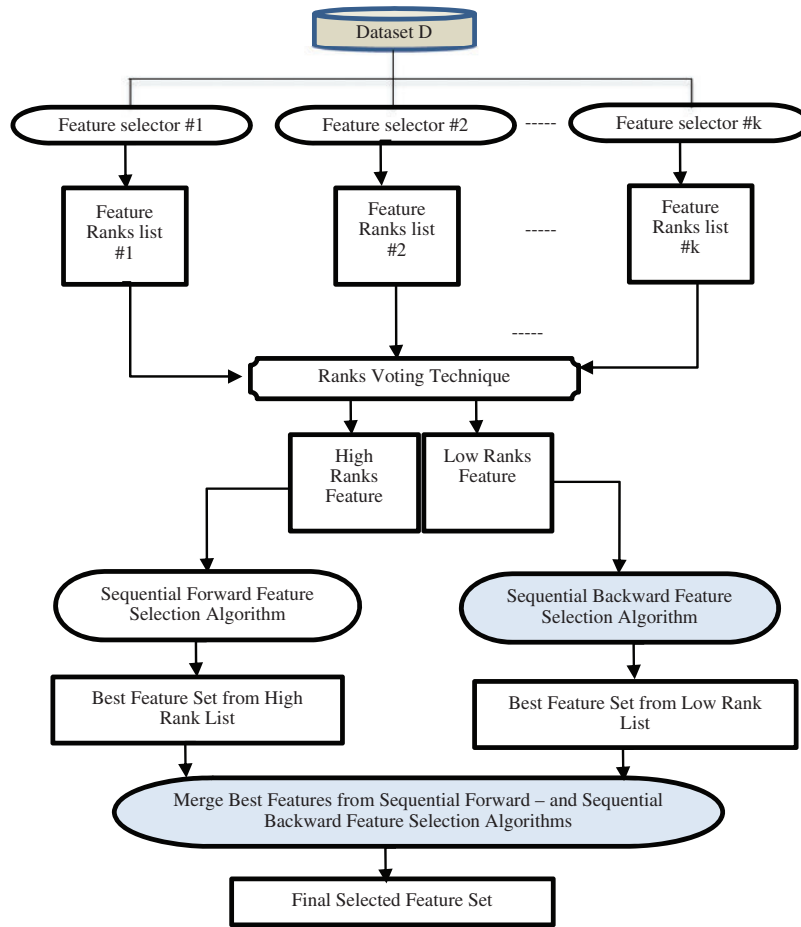


Figure 1: The phases of the proposed FRV algorithm

Rows \mathbf{R} in the matrix are sorted from the top (highest) to lower (lowest) ranks where the first row includes the features set with the highest ranks, the second row has the second-highest features set, and so on; the last row of the \mathbf{R} matrix contains the features set with lowest-ranks.

Step 3: Elect the highest-rank feature set and lowest-rank feature set by applying the Voting technique as follows: Start with an empty list \mathbf{V} , with \mathbf{n} -dimension features voting: \mathbf{n} is the number of the features, \mathbf{R} is a ranking matrix, $\mathbf{H} = \mathbf{n}/2$, and $\mathbf{L} = \mathbf{n}/2$ (number of features is even), and $\mathbf{H} = (\mathbf{n}/2) + 1$ and $\mathbf{L} = \mathbf{n}-\mathbf{H}$ (number of features is odd)

$$\begin{aligned} &\text{for } \mathbf{R} = 1 \rightarrow \mathbf{H} \\ &f_{Hi} = \max \left(\sum_{F=1}^k \text{frequency}(f_i)_R \right) \\ &\mathbf{R} = \mathbf{R} + 1 \\ &\text{end for} \\ &\text{for } \mathbf{R} = \mathbf{H} + 1 \rightarrow \mathbf{n} \\ &f_{Li} = \max \left(\sum_{F=1}^k \text{frequency}(f_i)_R \right) \end{aligned}$$

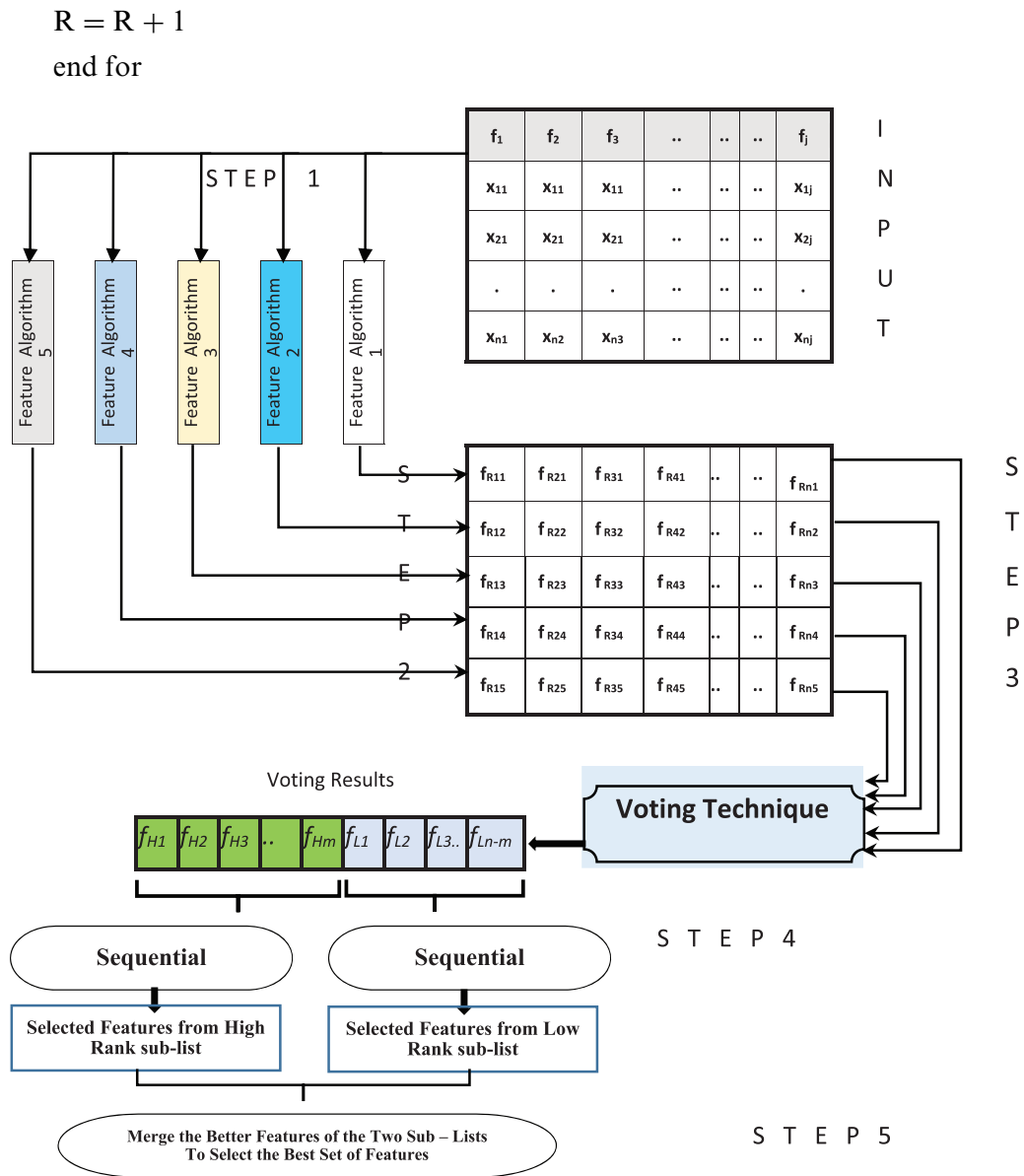


Figure 2: Detailed steps of the proposed (FRV)

Step 4. a: Apply the Forward Voting Selection based on Ranks Voting Algorithm: Start with the input of the highest rank list set of features obtained in step 3:

```

YH_subopt ← ∅
for i = 1 → H do
    fi ← max (J (YH_subopt ∪ {fi})) | fi ∈ fHi and fi ∈ / YH_subopt
    YH_subopt ← YH_subopt ∪ {fi}
end for
    
```


Step 4. b: Selection of the feature backward based on the algorithm of the Ranks voting: Start by entering the lowest list of features in step 3:

```

YL_ support ← all features
for i = n → H do
fi ← min (J (YL_ subopt ∪ {fi})) | fi ∈ fLi and fi ∉ YL_ subopt
YL_ subopt ← YL_ subopt-{fi}
end for

```

Step 5: Combine the voted highest-rank features set and voted lowest-rank set from steps 4. a and 4. b to find the best set of YH_ subopt and YL_ subopt selected features.

Fig. 2, illustrates a detailed example for FRV.

4 Experimental Design

This section presents the experiments and the workflow of building the structure of the data mining model applying the proposed FRV algorithm using real datasets, the workflow is divided into two phases:

- a. Phase 1: The datasets description and preprocessing; to solve the problem of the missed features' values (if exist).
Phase 2: Apply the Proposed FRV Algorithm. As well as the description of the experiments' setup including:
- b. The application of different five features ranking algorithms on these datasets to find the features' ranks,
- c. The voting lists extraction (the highest ranks feature list and the lowest ranks features list),
- d. Finally, the application of the proposed FRV algorithm on the extracted voting list to find the optimal features,
- e. The building of the data mining model using the selected features set from real-world data sets to evaluate the performance of the selected feature set from the different datasets using five different classifiers.

4.1 Datasets Description and Preprocessing

The datasets that were used in this study are described here. To find the feature ranks, the different five feature ranking algorithms will be applied to these datasets. Calculation of voting lists (the list of features with the highest ranks and the list of features with the lowest ranks). The proposed FRV algorithm was used to find the optimal features on a voting list, and the output of the selected features set from different datasets was evaluated using five different classifiers to demonstrate the success and applicability of the proposed FRV algorithm in real-world applications. Ten different datasets were used in this paper, three of them are used to illustrate how the proposed FRV works in detail (different in the number of attributes and the number of instances); the smallest dataset is the Iris dataset (4 attributes and 150 instances), the second dataset is the Diabetes dataset (9 attributes and 768 instances), and the third dataset is the CAD Hungarian (14 attributes and 294 instances). The other 7 datasets are presented with their results and analysis in Section 5. More details about the data description can be found at [44,45].

4.2 Features Ranking by Different Features Evaluators

Five attribute evaluators were applied in this paper to identify the ranks of the attributes in each dataset, these attribute evaluators are implemented in [46] and shortly described as follows:

Information Gain Ranking Filter: Determines the value of an attribute by calculating the class's information gain: $H(\text{Class}) - H(\text{Class} | \text{Attribute}) = \text{InfoGain}(\text{Class}, \text{Attribute})$. One Rule (short for OneR) Attribute Evaluation: Uses the OneR classifier to determine the value of an attribute. Gain Ratio Attribute Evaluation: Determines the value of an attribute by calculating the class gain ratio. $(H(\text{Class}) - H(\text{Class} | \text{Attribute})) / H(\text{Class})$ GainR (Class, Attribute) = $(H(\text{Class}) - H(\text{Class} | \text{Attribute})) / \text{GainR}(\text{Class}, \text{Attribute})$. The Correlation Attribute Evaluation (CA) tests the target class's attributes. The correlation between each attribute and the target class attribute is calculated using Pearson's correlation process. It takes nominal attributes into account on a value basis, with each value serving as an indicator. Attribute Evaluation with Relief: The Relief algorithm is a simple, efficient, and effective method of weighing attributes. The Relief algorithm produces a weight for each attribute ranging from 1 to 1, with higher positive weights suggesting more predictive attributes. These five attribute evaluators are used to illustrate how the proposed algorithm works moreover each application may choose its attribute evaluators with a variable number of evaluators (more or less).

4.3 Experiments Steps and Design Using the Proposed FRV Algorithm

The experiments were designed with the same methodology (set of steps) using the same evaluation methods for all datasets and were grouped as follows:

Experiment 1: Make use of all of the features in the data sets.

Experiment 2: uses the chosen features obtained from the proposed FRV algorithm: apply the feature selection phase for (forward voting). Use the Feature Selection Phase for Backward Voting. Combine the best features from stages b and c into a single function package. Some of the tasks were carried out using the WEKA software package implemented in [47] for the evaluation process of the proposed FRV algorithm.

5 Results Interpretation and Analysis

5.1 Results Analysis and Interpretation for the Iris Dataset

The Feature Ranks obtained by applying the Five Feature Evaluators on the Iris dataset are presented in [Tabs. 1a & 1b](#) illustrates the interpretation of how the voting technique works. While [Tab. 1c](#) illustrates the performance results of the different five classifiers using the obtained selected features for the Iris dataset. From the features ranks obtained in [Tab. 1](#), voting of the features in the highest rank list is as follows: Three attribute evaluators voted Petal Width as the largest, and this voting is described as 4 in the features range (3). Two attribute evaluators voted Petal Length as the largest, and this voting is described as 3 in the features range (2). Three attribute evaluators voted Petal Length, which has the third order in the features range, as the second-highest, and this voting is presented as 3 (3) Three attribute evaluators voted Petal Width as the second highest in the features range, and this voting is presented as 4.

Table 1a: The feature ranks obtained by applying the five feature evaluators on the iris dataset

Information gain			OneR			Gain ratio			Correlation			Relief-F			Voting	
Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	VOTE	
3	1.418	petal length	4	95.333	petal width	4	0.871	petal width	3	0.615	petal length	4	0.376	petal width	4(3), 3(2)	H
4	1.378	petal width	3	92.667	petal length	3	0.734	petal length	4	0.592	petal width	3	0.359	petal length	3(3), 4(2)	H
1	0.698	sepal length	1	72	sepal length	1	0.381	sepal length	1	0.478	sepal length	1	0.14	sepal length	1(5)	L
2	0.376	sepal width	2	55.333	sepal width	2	0.242	sepal width	2	0.397	sepal width	2	0.122	sepal width	2(5)	L

Table 1b: Features ranks obtained and voting technique on the iris dataset

Feature rank evaluator	Highest ranks		Lowest ranks	
	1 st Highest rank	2 nd Highest rank	2 nd Lowest	1 st Lowest
Information gain	petal length	petal width	sepal length	sepal width
OneR	petal width	petal length	sepal length	sepal width
Gain ratio	petal width	petal length	sepal length	sepal width
Correlation	petal length	petal width	sepal length	sepal width
Relief-F	petal width	petal length	sepal length	sepal width
Voting	petal length (2), petal width (3)	petal length (3), petal width (2)	sepal length (5)	sepal width (5)
Features order after voting	petal width	petal length	sepal length	sepal widths

Table 1c: Performance of the five classifiers for the iris dataset using the results of FRV algorithm

Exp. group	Attributes	Multi-perceptron	Naïve bayesian	Naïve bayes	TREES.J48	RULES.PART
		ACC%	ACC%	ACC%	ACC%	ACC%
Experiment 1	1, 2, 3, 4	98.7	96	96	98	97.3
Experiment 2.1	4, 3	96.7	96	96	98	97.3
	4, 3, 1	98.7	96.7	96.7	98	97.3
Experiment 2.2	4, 3, 1	98.7	96.7	96.7	98	97.3
	Highest ACC %	98.7	96.7	96.7	98	97.3
The success of the Proposed		✓	✓	✓	✓	✓

Therefore, the highest voted rank list of features will be as Petal-Width, Petal Length, and the lowest voted rank list of features will be as Sepal Width, Sepal Length. The forward feature selection step based on the voting techniques works on the Petal Width and Petal Length as inputs, while the backward voting feature selection step based on the voting techniques works on the Sepal Width and Sepal Length as inputs.

5.2 Results Analysis and Interpretation for the CAD Hungarian Dataset

The Feature Ranks obtained by applying the Five Feature Evaluators on the CAD Hungarian dataset are presented in Tab. 3a, while Tab. 3b illustrates the performance of applying the different five classifiers on the selected features for the CAD Hungarian dataset. Tab. 4 illustrates the results of applying the proposed features rank voting algorithm FRV on different real datasets and using its results to build different data mining’ models. The results illustrate the number of features in each dataset, the number of instances, the selected features using the FRV algorithm, the performance percentage of the data mining model based on the selected features, and the features reduction percentage.

5.3 General Remarks from the Results Analysis of the Proposed FRV

For the same dataset, the attribute rankings obtained by applying the five different attribute evaluators are not the same, proving that “no two-attribute evaluator gives the same ranks for features in the same dataset.” Via the voting technique, the proposed FRV was able to solve this problem. The effectiveness of the proposed FRV algorithm in building data mining models with higher efficiency and fewer features was shown by the performance results obtained by implementing the five different classifiers using all attributes and selected attributes in the proposed FRV algorithm. Finally, Tab. 4 shows the effects of using the proposed FRV algorithm to build data mining models with a high-performing percentage.

5.4 Results Analysis and Interpretation for the Diabetes Dataset

The Feature Ranks obtained by applying the Five Feature Evaluators on the Diabetes dataset are illustrated in Tab. 2a, while Tab. 2b illustrates the performance of applying the different five classifiers on the selected features for the diabetes dataset.

Table 2a: The feature ranks obtained by applying the five feature evaluators on the diabetes dataset

Information gain			OneR			Gain ratio			Correlation			Relief-F			VOTE		
Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	order	Rank	Name	order	Rank	Name			
2	0.19	plas	2	71.5	Plas	2	0.09	plas	2	0.46	plas	2	0.027	plas	<u>2(5)</u>	2	H
6	0.07	mass	1	67.4	Preg	6	0.08	mass	6	0.29	mass	6	0.015	mass	<u>6(4)</u>	6	H
8	0.07	age	8	66.9	Age	8	0.07	age	8	0.23	age	4	0.012	skin	<u>1(1)</u>	8	H
5	0.05	insu	5	64.9	Insu	1	0.05	preg	1	0.22	preg	1	0.011	preg	<u>8(4)</u>	1	H
															<u>4(1)</u>		
															<u>1(3)</u>		
															<u>5(2)</u>		

(Continued)

Table 2a: Continued

Information gain			OneR			Gain ratio			Correlation			Relief-F			VOTE	
Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	order	Rank	Name	order	Rank	Name		
4	0.04	skin	6	64.1	Mass	5	0.04	insu	7	0.17	pedi	8	0.009	age	4(1) 6(1) 5(1) 7(1) 8(1)	5 L
1	0.03	preg	4	63.3	Skin	7	0.022	pedi	5	0.13	insu	7	0.007	pedi	7(2) 5(1) 4(1) 1(1)	4, 7 L
7	0.02	pedi	3	62.5	Pres	4	0.02	skin	4	0.07	skin	3	0.005	pres	4(2) 3(2) 7(1)	4.3 L
3	0.01	pres	7	58.2	pedi	3	0.015	pres	3	0.06	pres	5	0.004	insu	4(2) 3(2) 7(1)	4.3 L

Table 2b: Performance of the five classifiers for the diabetes dataset using the results of FRV

Exp. group	Attributes	Multi-perceptron	Naïve bayesian	TREES.J48	RULES.PART
		ACC%	ACC%	ACC%	ACC%
Experiment 1	F1..f8	80.6	76.3	84.1	81.3
Experiment 2.1	2, 6, 8, 1	74.5	75.2	78.9	76.8
	2, 6, 8, 1, 5	75.9	75	78.9	77.7
	2, 6, 8, 1, 4	75.4	75.5	79.4	77.6
	2, 6, 8, 1, 4, 7	78.1	76.3	72.4	77.6
	2, 6, 8, 1, 7	77.6	76.5	83.2	76.9
	2, 6, 8, 1, 7, 3	78.6	76.9	84.1	76.9
Experiment 2.2	2, 6, 8, 1, 5, 3, 7	79.4	76.8	84.1	78.3
	2, 6, 8, 1, 5, 7	76.8	76.4	83.2	76.4
	2, 6, 8, 1, 3, 7	78.7	<u>77</u>	84.1	79.9
	2, 6, 8, 1, 3, 7	78.4	75.8	81	76.9
Max. ACC %, Min ERR%		80.6	77	84.1	81.3
The success of the FRV		≈	✓	✓	≈

Table 3a: Feature ranks obtained by applying the five feature evaluators on CAD hungarian dataset

Information gain			OneR			Gainratio			Correlation			ReliefF			VOTE:3, 9, 10, 8, 2, 11, 6			
Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	Order	Rank	Name	
3	0.25	chest_pain	11	81.9	slope	9	0.27909	exang	9	0.584	exang	3	0.29	chest_pain	3(2), 9(2), 11(1)	3, 9, 11	H	
9	0.24	Exang	9	81.2	exang	10	0.24103	oldpeak	10	0.545	oldpeak	2	0.155	sex	9(2), 10(2), 2(1)	9, 10	H	
10	0.22	Oldpeak	10	79.2	oldpeak	3	0.15088	chest_pain	3	0.421	chest_pain	10	0.054	oldpeak	10(3), 3(2)	10	H	
8	0.07	Thalach	3	78.5	chest_pain	8	0.07626	thalach	8	0.331	thalach	7	0.052	restecg	8(3), 3(1), 7(1)	8	H	
2	0.06	Sex	6	65.9	fbs	2	0.06962	sex	2	0.272	sex	9	0.037	exang	2(3), 6(1), 9(1)	2	H	
11	0.03	Slope	13	64.9	thal	11	0.05013	slope	5	0.202	chol	11	0.035	slope	11(3), 13(1), 5(1)	11	H	
6	0.017	Fbs	4	64.3	trestbps	6	0.04751	fbs	6	0.163	fbs	1	0.018	age	6(3), 4(1)	1(1)	6	H
7	0.002	Restecg	2	63.9	sex	7	0.00331	restecg	1	0.156	age	4	0.015	trestbps	7(2), 2(1), 7(1), 4(1)	7	L	
13	0.00207	Thal	7	63.9	restecg	13	0.00133	thal	4	0.139	trestbps	6	0.013	fbs	13(2), 4(1), 7(1), 6(1)	13	L	
4	0	Trestbps	12	63.9	ca	4	0	trestbps	11	0.126	slope	8	0.011	thalach	4(2), 1(1), 12(1), 8(1)	4	L	
5	0	Chol	5	62.2	chol	5	0	chol	7	0.015	restecg	5	0.005	chol	5(4), 7(1)	5	L	
12	0	Ca	8	61.9	thalach	12	0	ca	13	0.009	thal	13	0.005	thal	13(2), 12(2), 8(1)	13, 12	L	
1	0	Age	1	0	age	1	0	age	12	0	ca	12	0	ca	1(3), 12(2)	1, 12	L	

Table 3b: Performance of the five classifiers for the CAD hungarian dataset using the results of FRV

Exp. group	Attributes	Multi-perceptron	Naïve bayesian	TREES.J48	RULES.PART
		ACC%	ACC%	ACC%	ACC%
Experiment 1	All attributes F1..F13	96.9	85	84.1	86.4
Experiment 2.1	3, 9, 10, 8, 2, 11, 6	90.4	85.1	86.8	88.7
	2, 3, 6, 7, 8, 9, 10, 11	92.5	85.1	86.8	87.1
	2, 3, 6, 8, 9, 10, 11, 13	88.4	84.7	81.3	83.4
	2, 3, 4, 6, 8, 9, 10, 11	92.5	85	87.5	87.8
	2, 3, 5, 6, 8, 9, 10, 11	93.8	86.1	87.5	88.8
	2, 3, 5, 6, 8, 9, 10, 11, 12	92.8	<u>86.1</u>	87.5	<u>88.8</u>
	2, 3, 5, 6, 8, 9, 10, 11, 13	87.4	85.7	81.3	83.7
	2, 3, 5, 6, 8, 9, 10, 11, 1	93.5	86.1	88.1	88.5
Experiment 2.2	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	95.9	85.7	84.1	86.4
	2, 3, 4, 5, 6, 8, 9, 10, 11, 13	<u>96.8</u>	85.7	84.1	86.4
	2, 3, 4, 5, 6, 7, 8, 9, 10, 11	94.6	85.7	<u>88.4</u>	90.8
	2, 3, 4, 6, 7, 8, 9, 10, 11	94.9	85	84.7	88.1
	2, 3, 5, 6, 7, 8, 9, 10, 11	95.9	86.1	87.5	86.4
	2, 3, 5, 6, 8, 9, 10, 11	93.9	86.1	87.5	88.8
	2, 3, 5, 8, 9, 10, 11	92.5	85.8	87.5	87.5
	2, 3, 5, 6, 8, 9, 10	91.8	83.3	87.5	88.1
	2, 3, 5, 6, 8, 9, 10, 11	94.9	86.1	87.5	88.8
	Highest ACC % & Min ERR%	96.9	86.1	88.4	90.8
The success of the proposed	≈	✓	✓	✓	

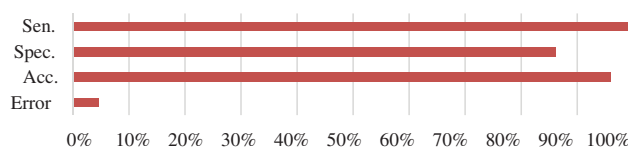
Table 4: Performance results using the proposed FRV for nine different datasets

Dataset	# Features	# Instances	# Selected features	ACC%	Features reduction %
Breast cancer	10	699	7	97.3%	30%
Diabetes	9	768	6	84.1%	33.3%
Ionosphere	34	351	22	96.8%	35.29%
Liver disorder	7	345	4	98.5%	42.85%
APS failure at Scania trucks	171	60000	106	97.9%	38.01%
Heart disease	75	303	47	96.8%	37.3%
Letter recognition	16	20000	10	94.7%	37.5%
CAD hungarian	14	294	10	96.8%	28.5%
Iris dataset	4	150	3	98.7%	25%
Z-Alizadeh Sani dataset	56	303	13	96%	77%

5.5 Performance Metrics

This section presents the performance measurement for the developed data mining models, using the obtained features selected from the proposed FRV. This step will be carried out by applying the proposed FRV on the dataset of CAD dataset (this dataset was chosen to prove the possibility for applying the proposed technique in critical and real applications) presented in [45] represent CAD patients' records which consists of 303 patients' records, each with 56 attributes, categorized into four categories; demographic features set. The patients' classes are normal (N) or CAD (C). The 56 features are reduced to 13 features using the proposed FRV, these 13 features are used to build a data mining model.

The performance of the built data mining model is tested using 10 folds cross-validation technique, the sensitivity, specificity, and accuracy evaluation–metrics [46], are used as measures for the performance analysis of the built data mining model, where: $(TN + TP) = (TN + TP + FN + FP)$ Accuracy $TP / (TP + FN)$ sensitivity (The percentage of actual positives, which are correctly identified) (Percentage of negatives correctly defined,) Specificity = $TN / (TN + FP)$. The description of these terms is as follows: TP: True positive, TN: True negative, FN: False-negative, and FP: False positive. These terms are obtained from the classification confusion matrix. Positive = Summation of the positive column, Negative = Summation of the negative column, and N = total number of instances. These performance metrics results are illustrated in Fig. 3. The calculation of the different performance metrics is presented below:

**Figure 3:** Performance measures for the obtained data mining model using FVR

Positive = TP + FN = 214 + 2 = 216, Negative = TN + FP = 75 + 12 = 87, TP = 214, FN = 2, FP = 12, and TN = 75 Exactly: $\frac{TP + TN}{TP + TN + TP + FP + FN} = \frac{(214 + 75)}{(214 + 75 + 2 + 12)} = 0.96$, Error = FN + FP/TP + TN + FP = 2 + 12, respectively). $\frac{TP}{(TP + FN)} = \frac{214}{(214 + 2)} = 0.9907$, $\frac{TN}{(TN + FP)}$ Specificity = $\frac{0.75}{(75 + 12)} = 0.862$.

Tab. 5 illustrates the accuracy comparison of the proposed FRV with other studies using the same datasets.

Table 5: Classification accuracies obtained by other studies using the same datasets

Study	Author (Year)	Focus area	Dataset	Accuracy %	Our proposed algorithm	
					Acc. %	Reduction %
[48]	Nasarian et al. (2020)	CAD	Hungarian	83.94%	96.8%	28.57%
			Long-beach-va	81.58%	82.31%	57.14%
			Z-Alizadeh Sani	92.59%	96%	76.78%
[49]	Abdar et al. (2019)	CAD	Z-Alizadeh Sani	93.08%	96%	53.7%
[50]	Abdar et al. (2018)	Diagnosis of breast cancer	Breast cancer wisconsin diagnost	98.07%	98.24%	65.62%
[51]	Comak et al. (2007)	Liver disorder data	Liver-disorders	94.29%	98.26%	14.28%

6 Conclusion and Future Extension

The work presented in this paper proposed a new feature selection algorithm FRV, The FRV comprises of three different techniques to select the minimum best feature set, the forward voting technique selects the best features according to their high voted ranks, the backward voting technique drops the worst feature with low voted ranks, and the third technique merges between the outputs from the forward and backward techniques to maximize the robustness of the selected features. Based on the feature ranks and the voting technique, the proposed FRV performance, and its robustness are evaluated by building different data mining models using different datasets. The performance of the built data mining model reflects the performance of the FRV algorithm. The FRV proved its success to select the best feature set with high robustness, this feature set helps in building intelligible and simpler data mining models through different stages as firstly, the identification of the features' ranks by applying five different attribute evaluators, secondly, the voting technique was applied on the obtained ranks of the features to construct the highest and lowest two lists of features, thirdly, the proposed FRV technique was applied using the ordered voted subsets of features to select the most suitable one, fourthly, the results of the proposed FRV technique were evaluated using five different classification models and ten datasets, and finally, the obtained results were interpreted and analyzed. The obtained results showed the success of the proposed FRV algorithm to a more concise data mining model for real-world applications. The proposed FRV proved its robustness, the extension of the proposed work will be the involvement of parallel processing to increase its performance.

The future extension for this work will be the comparison study between the performance of data mining building using deep learning and the presented FRV algorithm. Another extension of this work will be the application of the FRV algorithm in high dimension data as marketing datasets to select the most important features related to increasing the customers' loyalty applications, third extension will be the development of preprocessing tools that can be used in data mining development. Finally, the utilizing of the GA as an optimizing technique to select the higher and lower features ranks (the threshold level distinguish between the higher and lower features ranks).

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of the paper.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] N. Kushwaha and M. Pant, "Link-based BPSO for feature selection in big data text clustering," *Future Generation Computer Systems*, vol. 82, pp. 190–199, 2018.
- [2] L. M. Abualigah, A. T. Khader and E. S. Hanandeh, "A new feature selection method to improve the document clustering using particle swarm optimization algorithm," *Journal of Computational Science*, vol. 25, pp. 456–466, 2018.
- [3] S. Gunasundari, S. Janakiraman and S. Meenambal, "Multiswarm heterogeneous binary PSO using win-win approach for improved feature selection in liver and kidney disease diagnosis," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 135–154, 2018.
- [4] S. Dara and H. Banka, "A binary PSO feature selection algorithm for gene expression data," in *Proc. ICACACT*, IEEE, pp. 1–6, 2014.
- [5] A. Darabseh and A. S. Namin, "Effective user authentications using keystroke dynamics based on feature selections," in *Proc. ICMLA*, IEEE, pp. 307–312, 2015.
- [6] T. Khadhraoui, S. Ktata, F. Benzarti and H. Amiri, "Features selection based on modified pso algorithm for 2d face recognition," in *Proc. CGiV*, IEEE, pp. 99–104, 2016.
- [7] A. B. Soliman, M. Fares, M. M. Elhefnawy and M. Al-Hefnawy, "Features selection for building an early diagnosis machine learning model for Parkinson's disease," in *Proc. AIPR*, IEEE, pp. 1–4, 2016.
- [8] A. Oztekin, L. Al-Ebbini, Z. Sevkli and D. Delen, "A Decision-analytic approach to predicting quality of life for lung transplant recipients: A hybrid genetic algorithm-based methodology," *European Journal of Operational Research*, vol. 266, no. 2, pp. 639–651, 2018.
- [9] L. Cao, J. Li, Y. Zhou, Y. Liu and Liu, H. Cao, "Automatic feature group combination selection method based on GA for the functional regions clustering in DBS," *Computer Methods and Programs in Biomedicine*, vol. 183, pp. 105091, 2020.
- [10] J. Liu, W. Zhang, Z. Tang, Y. Xie, T. Ma, J. Zhang *et al.* "Adaptive intrusion detection via GA-GOGMM-based pattern learning with fuzzy rough set-based attribute selection," *Expert Systems with Applications*, vol. 139, pp. 112845, 2020.
- [11] M. Z. Mas'ud, S. Sahib, M. F. Abdollah, S. R. Selamat and R. Yusof, "Analysis of features selection and machine learning classifier in android malware detection," in *Proc. ICISA*, IEEE, pp. 1–5, 2014.
- [12] W. Ouarda, H. Trichili, A. M. Alimi and B. Solaiman, "Combined local features selection for face recognition based on naïve Bayesian classification," in *Proc. HIS*, IEEE, pp. 240–245, 2013.
- [13] A. Pattanatepapon, W. Suwansantisuk and P. Kumhom, "Feature selection based on correlations of gene-expression values for cancer prediction," in *Proc. IECBES*, IEEE, pp. 716–721, 2016.

- [14] M. Wang, Y. Tan and C. Li, "Features selection of high-quality essays in automated essay scoring system," in *Proc. ICECE*, IEEE, pp. 3027–3030, 2011.
- [15] O. Smart and L. Burrell, "Genetic programming and frequent itemset mining to identify feature selection patterns of IEEG and FMRI epilepsy data," *Engineering Applications of Artificial Intelligence*, vol. 39, pp. 198–214, 2015.
- [16] A. Rammal, H. Fenniri, A. Goupil, B. B. Chabbert, I. Bertrand *et al.*, "Features' selection based on weighted distance minimization application to biodegradation process evaluation," in *Proc. EUSIPCO*, IEEE, pp. 2651–2655, 2015.
- [17] A. Abuobieda, N. Salim, A. T. Albaham, A. H. Osman and Y. J. Kumar, "Text summarization features selection method using a pseudo-genetic-based model," in *Proc. CAMP*, IEEE, pp. 193–197, 2012.
- [18] Y. Yang and Y. Wu, "The improved features selection for text classification," in *Proc. ICCET*, IEEE, pp. V6–268, 2010.
- [19] P. Chandrakar and S. G. Qureshi, "A DWT based video watermarking using random frame selection," *International Journal of Research in Advent Technology*, vol. 3, no. 6, pp. 115–128, 2015.
- [20] M. Tutkan, M. C. Ganiz and S. Akyokuş, "Helmholtz principle-based supervised and unsupervised feature selection methods for text mining," *Information Processing & Management*, vol. 52, no. 5, pp. 885–910, 2016.
- [21] S. Jadhav, H. He and K. Jenkins, "Information gain directed genetic algorithm wrapper feature selection for credit rating," *Applied Soft Computing*, vol. 69, pp. 541–553, 2018.
- [22] Y. Wang and L. Feng, "Hybrid feature selection using component co-occurrence-based feature relevance measurement," *Expert Systems with Applications*, vol. 102, pp. 83–99, 2018.
- [23] S. Cateni, V. Colla and M. Vannucci, "A hybrid feature selection method for classification purposes," in *Proc. EMS*, IEEE, pp. 39–44, 2014.
- [24] C. Yan, J. Liang, M. Zhao, X. Zhang, T. Zhang *et al.*, "A novel hybrid feature selection strategy in quantitative analysis of laser-induced breakdown spectroscopy," *Analytica Chimica Acta*, vol. 1080, pp. 35–42, 2019.
- [25] S. Nisar and M. Tariq, "Intelligent feature selection using hybrid-based feature selection method," in *Proc. INTECH*, IEEE, pp. 168–172, 2016.
- [26] F. N. Koutanaei, H. Sajedi and M. Khanbabaei, "A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring," *Journal of Retailing and Consumer Services*, vol. 27, pp. 11–23, 2015.
- [27] X. Chen, L. Song, Y. Hou and G. Shao, "Efficient semi-supervised feature selection for VHR remote sensing images," in *Proc. IGARSS*, IEEE, pp. 1500–1503, 2016.
- [28] A. P. Shulyak and A. D. Shachykov, "About the impact of informative features selection in the mutually orthogonal decompositions of biomedical signals for their recognition," in *Proc. ELNANO*, IEEE, pp. 228–231, 2016.
- [29] D. Bertonecelli and P. Caianiello, "Customer return detection with features selection," in *Proc. Design and Diagnostics of Electronic Circuits & Systems*, IEEE, pp. 268–269, 2014.
- [30] C. Ch and M. Sharma, "Qualitative features selection techniques by profiling the statistical features of ECG for classification of heartbeats," *Biomedical Research*, vol. 28, no. 2, pp. 230–236, 2017.
- [31] A. Mohamed and M. E. El-Hawary, "On the optimization of SVMs kernels and parameters for electricity price forecasting," in *Proc. EPEC*, IEEE, pp. 1–6, 2016.
- [32] N. N. A. Aziz, Y. M. Mustafah, A. W. Azman, N. A. Zainuddin and M. A. Rashidan, "Features selection for multi-camera tracking," in *Proc. ICCCE*, pp. 243–246, 2014.
- [33] R. B. Dubey, M. Hanmandlu and S. Vasikarla, "Features selection in mammograms using fuzzy approach," in *Proc. ITNG*, IEEE, pp. 123–128, 2014.
- [34] M. Hagui and M. A. Mahjoub, "Features selection in video fall detection," in *Proc. IPAS*, IEEE, pp. 1–5, 2014.
- [35] A. Bawazir, W. Alhalabi, M. Mohamed, A. Sarirete and A. Alsaig, "A formal approach for matching and ranking trustworthy context-dependent services," *Applied Soft Computing*, vol. 73, pp. 306–315, 2018.

- [36] J. Zhang and S. Wang, "A novel single-feature and synergetic-features selection method by using ise-based kde and random permutation," *Chinese Journal of Electronics*, vol. 25, no. 1, pp. 114–120, 2016.
- [37] T. Handhayani, "Batik lasem images classification using voting feature intervals 5 and statistical features selection approach," in *Proc. ISITIA*, IEEE, pp. 13–16, 2016.
- [38] J. A. Aledo, J. A. Gá, D. Molina and A. Rosete "FSS-OBOP: Feature subset selection guided by a bucket order consensus ranking," in *Proc. SSCI*, IEEE, Vol. 2 pp. 1–8, 2016.
- [39] H. Fu, Z. Shao, C. Tu and Q. Zhang, "Impacts of feature selection for urban impervious surface extraction using the optical image and SAR data," in *Proc. EORSA*, IEEE, pp. 419–423, 2016.
- [40] K. S. Benli and M. T. Eskil, "Extraction and selection of muscle-based features for facial expression recognition," in *Proc. ICPR*, IEEE, pp. 1651–1656, 2014.
- [41] B. S. Liu, L. C. Han and D. Sun, "Finding the optimal sequence of features selection based on reinforcement learning," in *Proc. CCIS*, IEEE, pp. 347–350, 2014.
- [42] T. E. Juan and C. A. Perez. "Gender classification from NIR images by using quadrature encoding filters of the most relevant features," *IEEE Access*, vol. 7, pp. 29114–29127, 2019.
- [43] P. Mahendra, S. Tripathi and K. Dahal, "An efficient feature selection-based Bayesian and rough set approach for intrusion detection," *Applied Soft Computing*, vol. 87, pp. 105980, 2020.
- [44] D. Dua and C. Graff, *UCI Machine Learning Repository*, Irvine, CA: The university of California, School of Information and Computer Science, 2004. [Online]. Available: <http://www.archive.ics.uci.edu/ml>.
- [45] R. Alizadehsani, J. Habibi, M. J. Hosseini, H. Mashayekhi, R. Boghrati *et al.*, "A data mining approach for the diagnosis of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 1, pp. 53–61, 2013.
- [46] J. Dziak, D. Coffman, S. Lanza, R. Li and L. Jermiin, "Sensitivity and specificity of information criteria," *bioRxiv*, 449751, 2019.
- [47] I. Fatma A. and O. A. Shiba. "Data mining: WEKA software (an overview)," *Journal of Pure and Applied Science*, vol. 18, no. 3, pp. 113–120, 2019.
- [48] E. Nasarian, M. Abdar, M. A. Fahami, R. Alizadehsani, S. Hussain *et al.* "Association between work-related features and coronary artery disease: A heterogeneous hybrid feature selection integrated with balancing approach," *Pattern Recognition Letters*, vol. 133, pp. 33–40, 2020.
- [49] M. Abdar, W. Książek, U. R. Acharya, R. S. Tan, V. Makarenkov and P. Pławiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 179, pp. 104992, 2019.
- [50] M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao *et al.* "A new nested ensemble technique for automated diagnosis of breast cancer," *Pattern Recognition Letters*, vol. 132, pp. 123–131, 2020.
- [51] E. Comak, K. Polat, S. Güneş and A. Arslan, "A new medical decision-making system: Least square support vector machine (LSSVM) with fuzzy weighting pre-processing," *Expert Systems with Applications*, vol. 32, no. 2, pp. 409–414, 2007.