A new improved maximal relevance and minimal redundancy method based on feature subset

Shanshan Xie¹ · Yan Zhang² · Danjv Lv¹ · Xu Chen¹ · Jing Lu¹ · Jiang Liu³

Accepted: 9 August 2022
© The Author(s) 2022

Abstract
Feature selection plays a very significant role for the success of pattern recognition and data mining. Based on the maximal relevance and minimal redundancy (mRMR) method, combined with feature subset, this paper proposes an improved maximal relevance and minimal redundancy (ImRMR) feature selection method based on feature subset. In ImRMR, the Pearson correlation coefficient and mutual information are first used to measure the relevance of a single feature to the sample category, and a factor is introduced to adjust the weights of the two measurement criteria. And an equal grouping method is exploited to generate candidate feature subsets according to the ranking features. Then, the relevance and redundancy of candidate feature subsets are calculated and the ordered sequence of these feature subsets is gained by incremental search method. Finally, the final optimal feature subset is obtained from these feature subsets by combining the sequence forward search method and the classification learning algorithm. Experiments are conducted on seven datasets. The results show that ImRMR can effectively remove irrelevant and redundant features, which can not only reduce the dimension of sample features and time of model training and prediction, but also improve the classification performance.

Keywords mRMR · Feature subset · Feature selection · ImRMR · Sequence forward search
1 Introduction

With the development of machine learning and artificial intelligence, more and more rich information is obtained from research objects, but some irrelevant and redundant features are added, which leads to higher feature dimensions. For the classification learning, the good learning samples are the key to train the classifier. The irrelevant or redundant information of the samples will increase the complexity of the classification algorithm as well as the time of training and prediction, and directly affect the classification performance. Therefore, it is necessary to select the optimal features that can best represent the characteristic of object, and eliminate irrelevant and redundant features to improve the efficiency of the classifier [1].

Feature selection is the process of selecting the optimal feature subset from the raw feature set to reduce the dimension of the feature space [2]. Its purpose is to simplify the data structure, interpret the data information, and improve the robustness, stability and identification of the model. According to whether the evaluation criteria of feature subset are related to classifier, feature selection methods are divided into three ways: filter, wrapper and embedded [3]. Filter method relies on the properties of the feature space itself, based on feature metrics to maximize data representation information. The feature selection and training classifier are two independent processes. For example, Zhao et al. [4] applied constraint score method to score audio features in environmental sound classification and select the optimal feature subset to improve the classification performance. Saqlain et al. [5] used fisher score method to select feature subsets, and transferred selected feature subsets to the RBF kernel-based SVM for diagnosing heart disease. Yong et al. [6] divided features into three levels: strong correlation, sub-strong correlation and other features, based on the measurement criterion of mutual information and the correlation between the feature and the label. Then the features are simplified. Wrapper method relies on preselected classifiers, takes the performance of the classifier as an evaluation criterion for feature subset, and removes features that negatively affect the classification performance. For example, Ye et al. [7] proposed a multimodal wrapper feature selection method based on effective distance, which considered the global relationship between samples. Mafarja et al. [8] wrapped the whale optimization algorithm into the classifier, and used the crossover and mutation characteristics of the optimization algorithm to enhance the speed and accuracy of feature search. In order to improve the classification performance of the financial credit scoring model, Jadhav et al. [9] proposed an information gain directed feature selection algorithm, and propagated the top \( m \) features through the genetic algorithm wrapper. The embedded method puts the feature selection process into the learning algorithm, which is a compromise between the filter method and the wrapper method. For example, Xu et al. [10] proposed a joint intra-class variation coefficient and inter-class mutual information, constructed a joint index to evaluate the feature contribution, and combined the embedded method for feature selection. Maldonado et al. penalized the cardinality of feature sets through the scale factor technology, and used SVM
as the embedded base classifier for feature selection. This method was applied to high-dimensional datasets with extremely uneven categories [11]. Liu et al. [12] proposed an unsupervised adaptive neighborhood embedded feature selection method, which mainly captures the internal geometric structure of the data based on the $K$-nearest neighbor method. Because both the wrapper and embedded methods have certain dependencies on the classification learning algorithm, and are prone to problems such as overfitting and low efficiency. The filter method has high efficiency and is suitable for various data types. The method has strong versatility and low algorithm complexity.

In the filter method, the feature metric is the key factor. That the maximal relevance and minimal redundancy (mRMR) criterion is a filter feature metric for pattern recognition [13], it aims to select the feature subset with the maximal dependency, maximal relevance and the minimal redundancy from the raw feature set. However, mRMR only considered the contribution of a single feature. The feature with the maximum contribution from the raw feature set are selected and added to form the optimal feature subset. In this way, the relevance between features will be ignored, and the joint contribution of multiple features to the classification will not be considered enough, resulting in the selected feature subset may not be optimal. There is still a certain room for improvement in the selection of the optimal feature subset. Therefore, based on the mRMR method, in this paper, the ImRMR (Improved mRMR) method is proposed. First, the relevance between each feature and the class is calculated through the Pearson correlation coefficient and mutual information, and tradeoff the two metrics to rank the features. Secondly, candidate feature subsets are generated with the proposed equal grouping method (EGM) according to the ranking features. The joint contribution of feature subset is evaluated by the relevance and redundancy of feature subset, and the ranking of feature subsets is found by incremental search method. Finally, the final optimal feature subset is obtained by combining the sequence forward search method (SFS). The experiments are conducted on seven datasets to verify the effectiveness of the ImRMR.

This paper adopts a supervised learning method to perform feature selection on these datasets through ImRMR, and uses advanced machine learning classification algorithm (random forest) to classify these features. In this study, 70% of each dataset was randomly selected for the training dataset, and the remaining 30% was used for the test dataset. We conduct extensive experiments and evaluate on various performance metrics (accuracy, dimensionality reduction rate, comprehensive rate, precision, recall and $F$-measure) to determine the effectiveness of the proposed method. Experiments show that the proposed feature selection method on random forest classifier outperforms the original mRMR and other methods.

The rest of this paper is organized as follows: Sect. 2 lists the related work done by various authors. Section 3 describes the mRMR method. Section 4 presents the proposed ImRMR method in detail. Then, the experiments and results are analyzed in Sect. 5. Finally, we give limitations and conclusion.
2 Related work

Based on the mRMR, Lu et al. extracted the key information most relevant to the fault location in traditional transmission line faults. By mining the implicit relationship between key features and fault location, the fault location result is obtained by synthesizing multiple feature information. This method has better fault location accuracy and better adaptability to transient components that appear after the fault [14]. Billah and Waheed [15] effectively reduced the dimensionality of the extracted features of endoscopic video frames with mRMR. Toaar et al. used the mRMR to select the features of the lung X-ray images for the diagnosis of pneumonia extracted by three deep models, and combined them to form an efficient feature set. The selected feature set provided robust and consistent features for pneumonia detection, and the mRMR can effectively reduce the dimension of the feature set [16]. Gu et al. calculated the relevance between power transformer fault feature quantities and the redundancy between feature quantities and fault types based on the mRMR, the optimal transformer fault feature set was obtained. Experiments show that the optimal feature set is more efficient than the traditional feature set in transformer fault diagnosis [17]. Erolu et al. [18] used the mRMR to select the features of breast ultrasound image extracted by the hybrid CNN structure, and achieved good results in classification and recognition. Fan et al. [19] used mRMR to select important lead-rhythm features extracted from electroencephalogram recordings to build a predictive model to predict the prognostic effect of acupuncture on depression. Tuncer et al. [20] used the discrete wavelet transform to decompose the electroencephalogram signal, and then used the mRMR to select the most discriminant feature from the texture features generated by the decomposed discrete wavelet transform subband, and used SVM classifier to classify the selected features. In the study of COVID-19 disease classification using supervised optimization machine learning technology, Sharma et al. [21] used the mRMR to remove irrelevant and misleading features in the high-dimensional data of COVID-19 to reduce the size of the search space of the cuckoo search algorithm and improve the learning efficiency. Baliarsingh et al. employed mRMR to select relevant subsets of genes from the microarray dataset. Then, simulated annealing is hybridized with Rao algorithm to improve the solution quality after each iteration of Rao algorithm. The discriminant genes selected on the SRBCT dataset have high classification accuracy [22].

In addition, many scholars have improved the mRMR to obtain better feature selection results. Feng and Zhang improved the mRMR by using conditional mutual information formula and three-dimensional calculation to determine the candidate connection of each attribute node in Bayesian network classifier. The method enhanced the reliability and robustness of small sample calculation [23]. Yao et al. introduced mRMR into the particle swarm optimization algorithm search process to select feature subset. The feasibility and effectiveness of the proposed algorithm were verified on the UCI dataset with SVM as the classifier [24]. Li and Wang [25] proposed a new mRMR method to use a variety of different evaluation criteria to measure the redundancy between features and the
relevance between features and categories, and an indicator vector $\lambda$ was introduced to describe the actual data dimension requirements of users. Wang et al. [26] improved the mRMR by merging relevance measurement coefficients to obtain a primary feature subset, and binary-coded the feature subset, and then combined the genetic algorithm to search for the optimal or suboptimal feature subset. Jo et al. [27] used the Pearson correlation coefficient as the redundancy metric and the $R$ value as the relevance metric to improve the mRMR. Ahmed et al. put forward an enhanced mRMR (EmRMR) filtering method to remove the noise features in ransomware, and selected the most relevant feature subset to characterize the real behavior of ransomware. EmRMR requires only a small amount of evaluation, avoiding unnecessary calculations inherent in the original mRMR [28]. Combines EmRMR with term frequency-inverse document frequency (TF-IDF), a weighted mRMR (WmRMR) technology was proposed to filter out runtime noise behavior according to the weight calculated by TF-IDF. Compared with the mRMR, WmRMR has low-dimensional complexity and less evaluation times, and better estimates the feature significance in the data captured in the early stage of ransomware attacks [29].

The existing improved methods for mRMR are based on the calculation of correlation and redundancy of a single feature, they ignore the joint contribution of multiple features, and do not take into account the calculation of redundancy and correlation between feature subsets. We propose an improve mRMR (ImRMR) based on feature subsets. The proposed method and performance evaluation will be discussed in the following sections.

3 mRMR feature selection

3.1 mRMR definition

Maximal relevance and minimal redundancy (mRMR) is a filter feature measurement criterion, which calculates the redundancy between features and the correlation between features and class based on mutual information $I(x;y)$. It selects the features that are most relevant to the category and have the least redundancy with other features from the raw feature set.

The mutual information $I(x;y)$ is defined as Eq. 1.

$$I(x;y) = \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \, dx \, dy$$  \hspace{1cm} (1)

where $p(x,y)$ is the joint probability density of random variables $x$ and $y$, and $p(x)$ and $p(y)$ are the marginal probability densities of $x$ and $y$, respectively.

Given a sample feature set $S = \{f_1, f_2, \ldots, f_n\}$ and a sample class $c$. The relevance between $S$ and $c$ is the mean of all mutual information between each feature $f_i$ and class $c$. It is shown in Eq. 2.
where \(|S|\) is the number of features in \(S\), and \(I(f_i; c)\) is the mutual information between the feature \(f_i\) and the class \(c\).

The redundancy of all features in \(S\) is the mean of all mutual information between feature \(f_i\) and feature \(f_j\). It is shown in Eq. 3.

\[
R(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j)
\]  

The mRMR method seeks the optimal features of the samples with maximal relevance \(D(S, c)\) and minimal redundancy \(R(S)\). So the criterion of feature measurement for mRMR can be as shown in Eq. 4.

\[
mRMR = \max_S (D(S, c) - R(S))
\]  

3.2 Incremental search method

Incremental search is used to quickly and efficiently select the optimal feature set. Given the raw feature set \(X\), if the optimal feature \(S_{m-1}\) has been selected, then it will continue to search for the optimal feature in the remaining feature space \(X - S_{m-1}\). Equation 4 can be illustrated as Eq. 5.

\[
\max_{f_j \in X - S_{m-1}} \left[ I(f_i; c) - \frac{1}{m-1} \sum_{f_j \in S_{m-1}} I(f_j; f_i) \right]
\]

4 Improved mRMR based on feature subset

The mRMR method measures the contribution of feature by calculating the relevance and redundancy of individual feature. The joint contribution of multiple features is ignored. Relevance and redundancy are only based on mutual information measures. Therefore, to obtain the optimal feature subset, from the perspective of feature subset, this paper uses two measurement criteria, Pearson correlation coefficient and mutual information to evaluate the relevance and redundancy of the feature subsets. A weight factor is introduced to tradeoff the two measurement criteria.

The whole process of proposed ImRMR is shown in Fig. 1. Firstly, we adopt equal grouping method (EGM) to initially divide the candidate feature subsets. Then the contribution of each candidate feature subset is calculated with the correlation and redundancy according to the Pearson correlation coefficient and mutual information. Finally, we convert the ranking of the feature subsets into the feature ranking and combine the
SFS search strategy to obtain the final preferred subset. Mapping the optimal feature subset to the raw sample set can carry out classification and recognition.

4.1 Initialize the candidate feature subsets with EGM

Given a sample set of $N \times M$, $N$ represents the number of samples, and $M$ refers to the feature dimension. The base decision tree in the random forest includes the feature number $\log_2 (M)$ as the candidate feature subsets [30]. The relevance between each feature and the class is calculated by the Pearson correlation coefficient and mutual information, and $\alpha$ is introduced to adjust the weight of the two indicators. The features are ranked on relevance. The relevance $I_{c}(f_i; c)$ between a single feature and the sample class is calculated as shown in Eqs. 6 and 7:

$$I_{c}(f_i; c) = \alpha \times I(f_i; c) + (1 - \alpha) \times \text{corr}(f_i; c)$$  \hspace{1cm} (6)$$

$$\text{corr}(f_i; c) = \frac{\sum f_i c - \frac{\sum f_i \sum c}{N}}{\sqrt{\left(\sum f_i^2 - \frac{(\sum f_i)^2}{N}\right) \left(\sum c^2 - \frac{(\sum c)^2}{N}\right)}}$$  \hspace{1cm} (7)$$
where $\alpha \in [0.1, 1]$, with a step size of 0.1, $I(f_i; c)$ represents the mutual information between each feature and class, and $\text{corr}(f_i; c)$ refers to the Pearson correlation coefficient between each feature and class.

The range of feature subsets $r$ is shown in Eqs. 8 and 9.

$$ r = \left[ 1, \text{round} \left( \frac{M}{F_c} \right) \right] $$  \hspace{1cm} (8)

$$ F_c = \text{round} \left( \log_2 (M) \right) $$  \hspace{1cm} (9)

where $F_c$ is the number of features in the feature subset.

This paper exploits an equal grouping method (EGM) to generate candidate feature subsets. According to the ranking features with relevance, each feature is assigned to a group one by one. When the first-round assignment has finished, the feature will be assigned to a group reversely till the end of the round, and keep going the assignation and swapping directions until all the features assigned to the groups.
For example, suppose the number of features $M = 20$, then $Fc = 4$, and the range of feature subsets $r = 5$, the feature subsets generation process is shown in Fig. 2.

### 4.2 Generate ranking feature subsets sequence

By calculating the contribution of each candidate feature subset, the ranking sequence of these feature subsets is obtained. It is mainly divided into the following three steps.

1. Calculate the relevance of all individual features in each candidate feature subset according to Eq. 6, and accumulate them to obtain the contribution of the candidate feature subset. The candidate feature subset with the largest contribution is taken as the first selected feature subset $F_y$;
2. Calculate the relevance $D(F_x; c)$ and redundancy $R(F_x; F_y)$ of the remaining candidate feature subsets according to Eqs. 10 and 12, respectively;
3. Combined with Eq. 14, the incremental search method is used to select feature subset. Then, contribution ranking of all feature subsets is obtained.

- Contribution of single feature
  
  $$\begin{align*}
  D(F_x; c) &= \beta \ast I(F_x; c) + (1 - \beta) \ast \text{corr}(F_x; c) \quad (10) \\
  I(F_x; c) &= \sum_{i=1}^{Fc} I(f_i; c), \quad \text{corr}(F_x; c) = \sum_{i=1}^{Fc} \text{corr}(f_i; c) \quad (11) \\
  \end{align*}$$

- Redundancy
  
  $$R(F_x; F_y) = \frac{1}{m-1} \sum_{F_y \in S_{m-1}} (\beta \ast I(F_x; F_y) + (1 - \beta) \ast \text{corr}(F_x; F_y) ) \quad (12)$$

- Improvement of mutual relevance
  
  $$I(F_x; F_y) = \sum_{i,j=1}^{Fc} I(f_i; f_j), \quad \text{corr}(F_x; F_y) = \sum_{i,j=1}^{Fc} \text{corr}(f_i; f_j) \quad (13)$$

- Maximal relevance and minimal redundancy
  
  $$\text{ImRMR} = \max_{F_x \in X - S_{m-1}} \left[ D(F_x; c) - R(F_x; F_y) \right] \quad (14)$$

where $\beta \in [0.1, 1]$, with a step size of 0.1.

### 4.3 Acquisition of optimal feature subset

Take out the features of the feature subsets sequence in turn to find features sequence. The optimal feature subset is obtained according to the ranking features sequence and the SFS search strategy.

The feature selection algorithm based on ImRMR and SFS is described as follows.
Algorithm 1 Feature selection algorithm based on ImRMR and SFS.

Syntax definitions: Let \( F_c \) be the number of features in feature subset, \( r \) be the range of feature subsets, \( RS \) be ranking feature subsets sequence, \( OS \) be ranking features sequence, \( CS \) be optimal feature subset, \( iacc \) and \( acc \) are accuracy of test set for \( CS \) and \( tempCS \), respectively.

Input: Training set, test set, and \( CS = \emptyset \).

Output: \( CS \)

Steps:
1. Calculate \( F_c \) and \( r \) according to Eq. 8 and Eq. 9;
2. Generate candidate feature subsets according to \( F_c \), \( r \), and EGM;
3. Calculate the contribution of each candidate feature subset according to Eq. 14 to obtain \( RS \);
4. Take out the features of \( RS \) in turn to obtain \( OS \), \( f_{max} \) is the first feature in \( OS \);
5. \( CS = CS \cup \{ f_{max} \} \), \( OS = OS - \{ f_{max} \} \) and calculate \( iacc \);
6. For each \( f_i \in OS \)
   6.1 \( tempCS = CS \cup \{ f_i \} \), and calculate \( acc \);
   6.2 If \( acc > iacc \)
   6.3 \( CS = CS \cup \{ f_i \} \);
   6.4 \( iacc = acc \);
6.3 End If
7. End for
8. Return \( CS \).

In addition, for the ImRMR algorithm, the same \( \alpha \) and different \( \beta \) settings make the sample set generate different feature subset when calculating the contribution of feature subset according to Eq. 14. Therefore, the different feature subsets \( b_1, b_2, \ldots, b_m \) can be obtained, and \( m \) represents the number of settings \( \beta \), where the feature subset \( b_{max} \) with the highest accuracy is selected as the optimal feature subset for the current \( \alpha \).

Different \( \alpha \) settings make the sample set generate different optimal feature subset such as \( CS_1, CS_2, \ldots, CS_n \), \( n \) represents the number of settings \( \alpha \). The feature subset \( CS^* \) with the highest accuracy among these \( n \) feature subsets is selected as the final optimal feature subset.

5 Experiment and result analysis

5.1 Experimental environment

The hardware platform is a laptop computer with 16G RAM and 11th Gen Intel(R) Core(TM) i7-11370H @ 3.30 GHz CPU. The operating system is Windows 10.
A new improved maximal relevance and minimal redundancy method…

64-bit. MATLAB 2018b is the data processing, programming and running platform as well WEKA 3.8.5.

5.2 Experimental datasets

Seven datasets were involved in experiments. Five of them are provided by the UCI Machine Learning Repository [31]: Musk dataset (Musk), Urban land cover dataset (Urban), Glass dataset (Glass), Libras Movement dataset (Movement) and Ionosphere dataset (Ionosphere). One dataset is the commonly used hyperspectral dataset Pavia University (PU) [32]. And the other one is crane songs dataset (Crane).

Each dataset is divided into training set and test set as the ratio of 7:3. The information of datasets are listed in Table 1.

5.3 The design of experiments

Two groups experimental scheme are designed. One group compared the proposed method of EGM with the method of randomly selecting features to generate candidate feature subsets. And the other compared the ImRMR with other feature selection methods, such as mRMR, InfoGain (IG) [33], Symmetrical Uncert (SU) [34], GainRatio (GR) [35] and ReliefF (RfF) [36].

The four feature selection methods of IG, SU, GR RfF and the classification method of random forest involved in the experiment are all built-in methods in WEKA 3.8.5, which are implemented by MATLAB 2018b calling the interface of WEKA 3.8.5, and the parameters of each feature selection method are default parameters. The mRMR method has no parameter settings, and $\alpha$ and $\beta$ in ImRMR are set in $[0.1, 1]$.

Random forest is an ensemble learning method based on bagging, which can handle classification problems and regression problems well, and is one of the most widely used machine learning methods at present. To verify effectiveness of the proposed ImRMR feature selection method, the random forest classifier is adopted to obtain classification results of the selected optimal feature subset. Each experiment is repeated 30 times independently, and the average of the experimental results is taken as the final result. The evaluation indicators include accuracy rate, dimensionality reduction rate, comprehensive rate, precision, recall and $F$-measure.

| Dataset    | Training set | Test set | Number of features | Number of categories |
|------------|--------------|----------|--------------------|----------------------|
| Musk       | 334          | 142      | 166                | 2                    |
| Urban      | 478          | 197      | 147                | 9                    |
| Ionosphere | 246          | 105      | 34                 | 2                    |
| Glass      | 150          | 64       | 9                  | 6                    |
| Movement   | 252          | 108      | 90                 | 15                   |
| PU         | 1800         | 900      | 103                | 9                    |
| Crane      | 243          | 100      | 75                 | 7                    |
The accuracy rate is used to evaluate the proportion of correctly identified samples to the total number of samples in the prediction results. Its calculation is shown in Eq. 15.

$$\text{Acc} = \frac{P_a}{N} \times 100\%$$  \hspace{1cm} (15)

where $P_a$ is the number of the correctly classified samples, and $N$ represents the number of all samples.

The dimensionality reduction rate as an evaluation indicator is introduced, as shown in Eq. 16.

$$Dr = \left(1 - \frac{Sc}{Oc}\right) \times 100\%$$  \hspace{1cm} (16)

where $Sc$ is the number of selected features, and $Oc$ represents the number of raw features. The larger the $Dr$ value, the stronger the ability to reduce dimensions.

The comprehensive rate considers the accuracy rate and the dimensionality reduction rate, as shown in Eq. 17.

$$Z = \theta \times \text{Acc} + (1 - \theta) \times Dr$$  \hspace{1cm} (17)

where $Z$ is the comprehensive rate, Acc is the accuracy rate, $Dr$ is the dimensionality reduction rate. The $\theta$ is the tradeoff factor. In experiments, the value of $\theta$ is 0.5.

The precision is used to evaluate the proportion of all predicted correct samples that contain actual correct samples. Its calculation is shown in Eq. 18.

$$\text{Precision} = \frac{TP}{TP + FP}$$  \hspace{1cm} (18)

where TP is the sum of the number of correctly classified samples and FP is the number of samples predicted to be correct that are actually wrong.

The recall rate is the percentage of in all samples where the correct sample is predicted to be correct. Its calculation is shown in Eq. 19.

$$\text{Recall} = \frac{TP}{TP + FN}$$  \hspace{1cm} (19)

where TP is the sum of the number of correctly classified samples and FN is the number of samples that are actually correct but predicted to be wrong.

$F$-measure is precision and recall weighted harmonic mean. $F$-measure calculation is shown in Eq. 20.

$$F\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$  \hspace{1cm} (20)
5.4 Result analysis

5.4.1 Comparison of EGM with random grouping

In this group experiments, EGM and the method of random grouping are, respectively, applied to the generating candidate feature subsets part of ImRMR, namely ImRMR-EGM and ImRMR-RS. The random grouping method is different from EGM only in that the features of its grouping are randomly assigned, and other parts are consistent with EGM.

Using SFS search strategy, by recording the dimensionality reduction rate, the accuracy rate, precision, recall, $F$-measure and the comprehensive rate when the classification accuracy obtained by the ImRMR-EGM and ImRMR-RS feature selection of the seven datasets reached the maximum value, the performances of the two methods are compared. The experimental results are shown in Table 2 and Fig. 3.

Seen from Table 2 and Fig. 3, ImRMR-EGM has good effects on the seven datasets. The dimensionality reduction rate reaches 55.56–90.96%, and the classification accuracy is 2.34–18.75% higher than that of the raw feature. ImRMR-EGM not only effectively reduces the dimension of the dataset and but also improves the classification accuracy.

In seven datasets, the comprehensive rate of ImRMR-EGM is higher than that of ImRMR-RS by 0.14–5.42%. And in most cases, the accuracy, precision, recall and $F$-measure of ImRMR-EGM are higher than ImRMR-RS. Therefore, the EGM has achieved good results in generating candidate feature subsets, which outperforms the random grouping method.

5.4.2 Comparison of ImRMR with other feature selection methods

Experiments with ImRMR and other five feature selection methods are conducted on seven datasets. Combine SFS search strategy, the results of various feature selection methods are classified by random forest. Their performances are compared with the evaluation indices including dimensionality reduction rate, accuracy rate, precision, recall, $F$-measure and comprehensive rate when the classification accuracy reaches the maximum.

The experimental results are shown in Table 3 and Fig. 4. And the abscissa labels 1, 2, 3, 4, 5, 6, and 7 in Fig. 4 represent raw feature and six feature selection method of ImRMR, mRMR, IG, SU, GR, and RfF, respectively.

Seen from Table 3 and Fig. 4, ImRMR outperforms five methods (mRMR, IG, SU, GR, and RfF) on seven datasets of Musk, Urban, Ionosphere, Glass, PU, Movement, and Crane. Among them, the comprehensive rate of ImRMR is 1.27–3.82%, 6.34–24.48%, 3.8–6.94%, 4.75–20.36% and 0.83–6% higher than the five methods on the five datasets of Urban, Glass, Movement, PU and Crane, respectively; for the Musk dataset, the comprehensive rate of ImRMR is only 0.6% lower than that of mRMR, and 2.67–4.18% higher than that of the other four methods; for the Ionosphere dataset, the comprehensive rate of ImRMR is only 0.99% and 0.04% lower than mRMR and SU, respectively, and 1.95–6.36% higher than the other three methods. The accuracy of ImRMR on six datasets (Urban,
Table 2 Experimental results of ImRMR-EGM and ImRMR-RS

| Dataset   | Performance (%) | Feature selection methods |   |   |   |
|-----------|----------------|---------------------------|---|---|---|
|           |                | Raw feature              | ImRMR-EGM | ImRMR-RS |   |
| Musk      |                | Acc 85.21                | 95.07 | 94.72 |   |
|           |                | Dr 0.00                  | 90.96 | 91.04 |   |
|           |                | Z 42.61                  | 93.02 | 92.88 |   |
|           |                | Recall 92.21             | 94.81 | 95.97 |   |
|           |                | F-measure 87.12          | 95.42 | 95.17 |   |
|           |                | Precision 82.56          | 96.05 | 94.60 |   |
| Urban     |                | Acc 82.23                | 87.82 | 87.02 |   |
|           |                | Dr 0.00                  | 90.48 | 88.41 |   |
|           |                | Z 41.12                  | 89.15 | 87.72 |   |
|           |                | Recall 82.40             | 87.93 | 87.43 |   |
|           |                | F-measure 81.81          | 88.03 | 87.02 |   |
|           |                | Precision 84.18          | 89.35 | 87.99 |   |
| Ionosphere|                | Acc 91.43                | 95.24 | 95.68 |   |
|           |                | Dr 0.00                  | 82.35 | 80.59 |   |
|           |                | Z 45.71                  | 88.80 | 88.14 |   |
|           |                | Recall 81.82             | 87.88 | 90.50 |   |
|           |                | F-measure 85.71          | 92.06 | 93.05 |   |
|           |                | Precision 90.00          | 96.67 | 95.80 |   |
| Glass     |                | Acc 65.63                | 84.38 | 82.81 |   |
|           |                | Dr 0.00                  | 55.56 | 46.30 |   |
|           |                | Z 32.81                  | 69.97 | 64.55 |   |
|           |                | Recall 64.62             | 78.08 | 76.06 |   |
|           |                | F-measure 63.36          | 81.15 | 77.18 |   |
|           |                | Precision 66.02          | 87.74 | 84.24 |   |
| Movement  |                | Acc 81.48                | 85.19 | 86.14 |   |
|           |                | Dr 0.00                  | 85.56 | 83.52 |   |
|           |                | Z 40.74                  | 85.37 | 84.83 |   |
|           |                | Recall 81.94             | 85.96 | 86.81 |   |
|           |                | F-measure 81.08          | 84.75 | 85.79 |   |
|           |                | Precision 83.03          | 87.17 | 87.08 |   |
| PU        |                | Acc 93.22                | 95.56 | 94.94 |   |
|           |                | Dr 0.00                  | 90.29 | 85.28 |   |
|           |                | Z 46.61                  | 92.92 | 90.11 |   |
|           |                | Recall 93.22             | 95.56 | 94.94 |   |
|           |                | F-measure 93.13          | 95.50 | 94.89 |   |
|           |                | Precision 93.33          | 95.56 | 94.98 |   |
| Crane     |                | Acc 90.00                | 93.00 | 92.47 |   |
|           |                | Dr 0.00                  | 84.00 | 81.16 |   |
|           |                | Z 45.00                  | 88.50 | 86.81 |   |
|           |                | Recall 90.32             | 94.02 | 92.70 |   |
|           |                | F-measure 90.24          | 93.73 | 92.72 |   |
|           |                | Precision 91.59          | 93.72 | 93.35 |   |

Best performance in bold
Ionosphere, Glass, PU, Movement and Crane) is higher than these five feature selection methods, and the accuracy on Musk datasets is the same as mRMR and better than the other four feature selection methods. In addition, in most cases, ImRMR is superior to the other five feature selection methods in terms of recall, $F$-measure and precision on seven sets of datasets. Therefore, from a comprehensive comparison, the ImRMR feature selection method is superior to the five feature selection methods.

From the above comparative experiments, experimental results show that the proposed ImRMR method can effectively remove irrelevant and redundant features, which can not only reduce the dimension of sample features, but also achieve better classification and recognition results.

5.4.3 Comparison with state-of-the-art methods

Recently, various researchers have analyzed various feature selection methods due to improved accuracy results. Table 4 shows the comparative analysis of the proposed method with other methods using the same dataset. It can be noticed that the proposed method shows higher accuracy (Acc) and comprehensive rate (Z) in feature selection compared to other methods using the same dataset.
| Dataset   | Performance (%) | Raw feature | Feature selection methods |
|-----------|-----------------|-------------|---------------------------|
|           |                 |             | ImRMR         | mRMR | IG | SU | GR | RfF |
| Musk      |                 |             | 85.21 95.07 | 95.07 | 83.80 | 87.32 | 86.62 | 91.55 |
| Dr        |                 | 0.00        | 90.96 92.17 | 94.58 | 90.36 | 92.17 | 89.16 |
| Z         |                 | 42.61       | 93.02 93.62 | 94.81 | 96.10 | 88.31 | 93.51 | 92.21 |
| Recall    |                 | 92.21       | 94.81 96.10 | 89.19 | 88.84 | 89.39 | 90.35 |
| F-measure |                 | 87.12       | 95.42 95.48 | 85.53 | 88.89 | 88.20 | 92.59 |
| Precision |                 | 82.56       | 96.05 94.87 | 82.93 | 84.71 | 84.52 | 88.24 |
| Urban     |                 | 82.23       | 87.82 82.74 | 84.26 | 85.28 | 84.26 | 82.74 |
| Dr        |                 | 0.00        | 90.48 90.48 | 91.16 | 90.47 | 86.39 | 89.80 |
| Z         |                 | 41.12       | 89.15 86.61 | 87.71 | 87.88 | 85.33 | 86.27 |
| Recall    |                 | 82.40       | 87.93 80.62 | 84.37 | 84.70 | 85.98 | 83.08 |
| F-measure |                 | 81.81       | 88.03 82.08 | 83.64 | 85.14 | 85.38 | 82.66 |
| Precision |                 | 84.18       | 89.35 86.48 | 84.61 | 86.99 | 85.96 | 84.22 |
| Ionosphere| Acc              | 91.43       | 95.24 94.29 | 93.33 | 92.38 | 94.29 | 94.29 |
| Dr        |                 | 0.00        | 82.35 85.29 | 79.41 | 85.29 | 79.41 | 70.59 |
| Z         |                 | 45.71       | 88.80 89.79 | 86.37 | 88.84 | 86.85 | 82.44 |
| Recall    |                 | 81.82       | 87.88 84.85 | 84.85 | 81.82 | 87.88 | 87.88 |
| F-measure |                 | 85.71       | 92.06 90.32 | 88.89 | 87.10 | 90.63 | 90.63 |
| Precision |                 | 90.00       | 96.67 96.55 | 93.33 | 93.10 | 93.55 | 93.55 |
| Glass     | Acc              | 65.63       | 84.38 82.81 | 70.31 | 78.13 | 68.75 | 71.88 |
| Dr        |                 | 0.00        | 55.56 44.44 | 33.33 | 33.33 | 22.22 | 33.33 |
| Z         |                 | 32.81       | 69.97 63.63 | 51.82 | 55.73 | 45.49 | 52.60 |
| Recall    |                 | 64.62       | 78.08 75.99 | 70.80 | 73.88 | 63.96 | 69.51 |
| F-measure |                 | 63.36       | 81.15 76.50 | 68.95 | 71.71 | 62.13 | 71.28 |
| Precision |                 | 66.02       | 87.74 81.23 | 70.47 | 72.28 | 72.44 | 80.02 |
| Movement  | Acc              | 81.48       | 85.19 81.48 | 82.41 | 83.33 | 84.26 | 82.41 |
| Dr        |                 | 0.00        | 85.56 75.56 | 74.44 | 77.78 | 78.89 | 80.00 |
| Z         |                 | 40.74       | 85.37 78.52 | 78.43 | 80.56 | 81.57 | 81.20 |
| Recall    |                 | 81.94       | 85.96 81.71 | 82.61 | 83.96 | 84.56 | 83.10 |
| F-measure |                 | 81.08       | 84.75 81.00 | 81.44 | 82.75 | 83.91 | 81.78 |
| Precision |                 | 83.03       | 87.17 83.42 | 84.17 | 83.90 | 85.45 | 84.00 |
| PU        | Acc              | 93.22       | 95.56 93.67 | 94.78 | 94.44 | 94.67 | 93.89 |
| Dr        |                 | 0.00        | 90.29 51.46 | 81.55 | 80.58 | 81.55 | 77.67 |
| Z         |                 | 46.61       | 92.92 72.56 | 88.17 | 87.51 | 88.11 | 85.78 |
| Recall    |                 | 93.22       | 95.56 93.67 | 94.78 | 94.44 | 94.67 | 93.89 |
| F-measure |                 | 93.13       | 95.50 93.59 | 94.71 | 94.37 | 94.62 | 93.84 |
| Precision |                 | 93.33       | 95.56 93.76 | 94.83 | 94.51 | 94.68 | 94.03 |
| Crane     | Acc              | 90.00       | 93.00 90.00 | 86.00 | 85.00 | 92.00 | 90.00 |
| Dr        |                 | 0.00        | 84.00 85.33 | 80.00 | 80.00 | 80.00 | 80.00 |
| Z         |                 | 45.00       | 88.50 87.67 | 83.00 | 82.50 | 86.00 | 85.00 |
| Recall    |                 | 90.32       | 94.02 89.73 | 86.72 | 86.00 | 92.06 | 91.80 |
| F-measure |                 | 90.24       | 93.73 90.93 | 86.63 | 85.44 | 92.43 | 90.90 |
| Precision |                 | 91.59       | 93.72 91.03 | 87.70 | 86.17 | 93.38 | 90.36 |

Best performance in bold
A new improved maximal relevance and minimal redundancy method…

5.5 Discussion of using SFSFs for ranking feature subsets sequence

To further explore the effectiveness of the feature subset selected by the ImRMR, the sequence forward selection feature subset method (SFSFs) is used for the ranking feature subsets sequence to verify the pros and cons of the feature subset.

The steps of the SFSFs are as below.

1. The optimal feature subset starts with an empty set;
2. The first feature subset in the ordered feature subset set is added for the first time, and then the feature subset is added iteratively according to the order, which is combined with the selected feature subset to form a new feature subset;
3. This process continues until the classification accuracy of the newly mapped training set is greater than or equal to the raw classification accuracy, and the corresponding feature subset is the selected optimal feature subset.

The feature selection method based on SFSFs is shown in Fig. 5. Combine the SFSFs search strategy to get the final preferred subset and mapping the optimal feature subset to the raw sample set can carry out classification and recognition. The effects of ImRMR and other five feature selection methods are compared by the dimensionality reduction rate, accuracy, comprehensive rate, recall, F-measure and precision when the classification accuracy with feature selection is greater than or equal to that of the raw datasets.

The experimental results are shown in Table 5 and Fig. 6. And the abscissa labels 1, 2, 3, 4, 5, 6, and 7 in Fig. 6 represent raw feature and six feature selection method of ImRMR, mRMR, IG, SU, GR, and RfF, respectively.
Table 4 Comparative analysis with other methods

| Study                  | Year | Method                                                                 | Dataset     | Acc (%) | Dr (%) | Z (%) |
|------------------------|------|------------------------------------------------------------------------|-------------|---------|--------|-------|
| Mafarja et al. [8]     | 2018 | Whale optimization approaches for wrapper feature selection with KNN classifier | Ionosphere  | 92.56   | 57.60  | 75.08 |
| Mafarja et al. [37]    | 2019 | Binary grasshopper optimization algorithm with mutation with KNN classifier | Ionosphere  | 96.59   | 79.41  | 88    |
| Du et al. [38]         | 2020 | Improved binary symbiotic organism search algorithm with transfer functions with KNN classifier | Ionosphere  | 92.96   | –      | –     |
| Xu et al. [10]         | 2020 | Maximum feature tree embedded with mutual information and coefficient of variation with random forest classifier | Ionosphere  | 94.32   | 64.71  | 79.52 |
| Ghosh et al. [39]      | 2020 | Binary social mimic optimization algorithm with x-shaped transfer function with KNN classifier | Ionosphere  | 95.71   | 76.47  | 86.09 |
| Han et al. [40]        | 2021 | Multi-objective particle swarm optimization with adaptive strategies with KNN classifier | Ionosphere  | 89.18   | –      | –     |
| Kang et al. [41]       | 2022 | Grey wolf improved flower pollination algorithm with KNN classifier     | Ionosphere  | 95.36   | 76.18  | 85.77 |
| Proposed method        | 2022 | ImRMR with random forest classifier                                     | Ionosphere  | 95.24   | 82.35  | 88.80 |
| Xu et al. [10]         | 2020 | Mutual information and coefficient of variation with random forest classifier | Crane      | 91.00   | 45.33  | 68.17 |
| Proposed method        | 2022 | ImRMR with random forest classifier                                     | Crane       | 93.00   | 84.00  | 88.50 |
| Zhang et al. [42]      | 2018 | Maximum joint mutual information algorithm with KNN classifier          | Movement    | 80.67   | 3.33   | 42.00 |
| Proposed method        | 2022 | ImRMR with random forest classifier                                     | Movement    | 85.19   | 85.56  | 85.37 |
| Zhang et al. [42]      | 2018 | Maximum joint mutual information algorithm with KNN classifier          | Musk        | 79.52   | 22.89  | 51.21 |
| Chen et al. [43]       | 2021 | Self-learning feature selection with random forest classifier           | Musk        | 88.63   | 53.73  | 71.18 |
| Han et al. [40]        | 2021 | Multi-objective particle swarm optimization with adaptive strategies with KNN classifier | Musk        | 83.80   | –      | –     |
| Proposed method        | 2022 | ImRMR with random forest classifier                                     | Musk        | 95.07   | 90.96  | 93.02 |

Best performance in bold
A new improved maximal relevance and minimal redundancy method…

From Table 5 and Fig. 6, the performance of ImRMR based on SFSFs is better than five feature selection methods including mRMR, IG, SU, GR, and RfF in seven datasets. Among them, the comprehensive rate of ImRMR is 5.32–26.46%, 15.64–24.39%, 21.57–31.02%, 9.39–30.21%, 11.33–29.33% higher than the five feature selection methods on the five datasets of Musk, Urban, Movement, PU and Crane, respectively; the comprehensive rate of ImRMR on Ionosphere was only 0.52% lower than mRMR, and 0.48–23.01% higher than the other four methods; the comprehensive rate of ImRMR on Glass is 3.3% and 8.08% lower than that of RfF and mRMR, respectively, and 1.47–12.58% higher than the other three methods. The accuracy of ImRMR on six datasets including Musk, Urban, Ionosphere, Glass, Movement and PU is higher than the five feature selection methods, and the accuracy on Crane dataset is the same as the five feature selection methods. Furthermore, ImRMR outperforms the other five feature selection methods in recall, F-measure and precision on the seven datasets in most cases. Therefore, in a comprehensive comparison, with SFSFs the ImRMR feature selection method is superior to the five feature selection methods such as mRMR, IG, SU, GR, and RfF, and can effectively measure the contribution of feature subset to obtain an effective ranking feature subsets sequence.

6 Limitations and future scope

This paper improves mRMR based on feature subsets, and proposes the ImRMR method, which is used to select efficient feature sets, reduce the dimension of feature sets, and improve the classification performance of samples. There still exists more issues to carry out in the future work, and some of the most important points are listed below:

- Extend the method to more datasets.
| Dataset   | Performance (%) | Raw feature | Feature selection methods |
|-----------|-----------------|-------------|--------------------------|
|           |                 |             | ImRMR | mRMR | IG   | SU  | GR   | RfF  |
| Musk      |                 |             |       |      |      |     |      |      |
| Acc       | 85.21           |             | 88.73 | 85.21 | 85.92 | 85.21 | 85.92 | 85.21 |
| Dr        | 0.00            |             | 92.77 | 43.37 | 72.89 | 81.93 | 84.94 | 77.71 |
| Z         | 42.61           |             | 90.75 | 64.29 | 79.40 | 83.57 | 85.43 | 81.46 |
| Recall    | 92.21           |             | 92.21 | 89.61 | 82.24 | 80.31 | 84.92 | 92.21 |
| $F$-measure | 87.12    |             | 89.87 | 86.79 | 87.65 | 86.79 | 87.50 | 87.12 |
| Precision | 82.56           |             | 87.65 | 84.15 | 83.53 | 84.15 | 84.34 | 82.56 |
| Urban     |                 |             |       |      |      |     |      |      |
| Acc       | 82.23           |             | 84.77 | 82.23 | 82.74 | 82.35 | 82.74 | 82.23 |
| Dr        | 0.00            |             | 90.48 | 51.02 | 54.42 | 51.02 | 61.22 | 44.22 |
| Z         | 41.12           |             | 87.62 | 66.63 | 68.58 | 67.13 | 71.98 | 63.23 |
| Recall    | 82.40           |             | 82.59 | 83.48 | 83.15 | 83.48 | 83.85 | 81.99 |
| $F$-measure | 81.81    |             | 83.38 | 83.12 | 82.26 | 83.39 | 82.96 | 81.73 |
| Precision | 84.18           |             | 86.31 | 84.75 | 84.33 | 84.63 | 84.20 | 83.98 |
| Ionosphere|                 |             |       |      |      |     |      |      |
| Acc       | 91.43           |             | 93.33 | 91.43 | 91.43 | 92.38 | 92.38 | 91.43 |
| Dr        | 0.00            |             | 70.59 | 73.53 | 66.67 | 69.08 | 70.59 | 70.59 |
| Z         | 45.71           |             | 81.96 | 82.48 | 58.95 | 81.48 | 81.48 | 81.01 |
| Recall    | 81.82           |             | 84.85 | 84.85 | 81.82 | 81.82 | 84.85 | 81.82 |
| $F$-measure | 85.71    |             | 88.89 | 86.15 | 85.71 | 87.10 | 87.50 | 85.71 |
| Precision | 90.00           |             | 93.33 | 87.50 | 90.00 | 93.10 | 90.32 | 90.00 |
| Glass     |                 |             |       |      |      |     |      |      |
| Acc       | 65.63           |             | 82.81 | 65.63 | 68.75 | 68.75 | 68.75 | 67.19 |
| Dr        | 0.00            |             | 33.33 | 66.67 | 44.44 | 33.33 | 22.22 | 55.56 |
| Z         | 32.81           |             | 58.07 | 66.15 | 56.60 | 51.04 | 45.49 | 61.37 |
| Recall    | 64.62           |             | 73.83 | 54.49 | 68.02 | 66.12 | 63.96 | 67.14 |
| $F$-measure | 63.36    |             | 75.88 | 61.47 | 66.69 | 64.75 | 62.13 | 65.95 |
| Precision | 66.02           |             | 84.26 | 70.51 | 68.77 | 67.60 | 72.44 | 67.94 |
| Movement  |                 |             |       |      |      |     |      |      |
| Acc       | 81.48           |             | 83.33 | 82.41 | 82.41 | 81.48 | 82.41 | 82.41 |
| Dr        | 0.00            |             | 86.67 | 25.56 | 31.11 | 31.11 | 44.44 | 34.44 |
| Z         | 40.74           |             | 85.00 | 53.98 | 56.76 | 56.30 | 63.43 | 58.43 |
| Recall    | 81.94           |             | 83.72 | 83.67 | 83.24 | 82.48 | 83.34 | 82.28 |
| $F$-measure | 81.08    |             | 82.83 | 82.35 | 81.70 | 81.16 | 81.87 | 81.22 |
| Precision | 83.03           |             | 84.11 | 83.45 | 82.70 | 83.19 | 82.33 | 83.57 |
| PU        |                 |             |       |      |      |     |      |      |
| Acc       | 93.22           |             | 93.89 | 93.22 | 93.67 | 93.44 | 93.56 | 93.22 |
| Dr        | 0.00            |             | 88.35 | 38.83 | 28.16 | 49.51 | 69.90 | 29.13 |
| Z         | 46.61           |             | 91.12 | 60.03 | 60.91 | 71.48 | 81.73 | 61.17 |
| Recall    | 93.22           |             | 93.89 | 93.22 | 93.67 | 93.44 | 93.56 | 93.22 |
| $F$-measure | 93.13    |             | 93.80 | 93.16 | 93.59 | 93.38 | 93.51 | 93.19 |
| Precision | 93.33           |             | 93.95 | 93.28 | 93.81 | 93.51 | 93.63 | 93.33 |
| Crane     |                 |             |       |      |      |     |      |      |
| Acc       | **90.00**       |             | **90.00** | **90.00** | **90.00** | **90.00** | **90.00** | **90.00** |
| Dr        | 0.00            |             | 80.00 | 53.33 | 57.33 | 49.33 | 34.67 | 21.33 |
| Z         | 45.00           |             | 85.00 | 71.67 | 73.67 | 69.67 | 62.33 | 55.67 |
| Recall    | 90.32           |             | 91.88 | 89.73 | 91.80 | 89.81 | 91.80 | 89.81 |
| $F$-measure | 90.24    |             | 90.02 | 90.08 | **90.84** | 89.24 | 90.51 | 89.71 |
| Precision | **91.59**       |             | 89.23 | 91.29 | 90.40 | 89.10 | 89.90 | 90.56 |

Best performance in bold
This study currently proposes EGM to divide candidate feature subsets. Other more feature subset division methods are needed to be explored to improve the quality of candidate feature subsets.

This study only uses Pearson correlation coefficient and mutual information to measure correlation and redundancy. In the next work, we will try to explore other measurement criteria to improve the effectiveness of the selected features.

7 Conclusion

In this paper, considering the joint contribution between multiple features, the proposed ImRMR method expands the feature selection process based on feature subset.

EGM is used to divide the raw feature set into multiple different candidate feature subsets. Two criteria of Pearson’s correlation coefficient and mutual information are used to calculate the correlation and redundancy of candidate feature subsets, and weights are introduced to tradeoff both criteria. Then the SFS search strategy is combined for the ranking feature subsets sequence to obtain the optimal feature subset. Compared with five methods, including mRMR, InfoGain, Symmetrical Uncert, GainRatio, and ReliefF, experimental results verified the effectiveness on seven datasets. In most cases, ImRMR outperforms the other methods, and can effectively obtain the optimal feature subset and improve the performance of the classification.

Acknowledgements This study was supported by the National Natural Science Foundation of China under Grant No. 61462078 and under Grant No. 31860332, the Yunnan Provincial Science and
Technology Department under Grant No. 202002AA10007, the Yunnan Provincial Department of Education under Grant No. 2022Y558.

**Author contributions** Conceptualization, SSX and YZ; Data curation, JL and JL; Funding acquisition, YZ, DJL and SSX; Investigation, JL and JL; Methodology, SSX and YZ; Resources, JL and JL; Supervision, YZ and XC; Validation, SSX, DJL and XC; Writing—original draft, SSX; Writing—review & editing, YZ, DJL and XC. All authors have read and agreed to the published version of the manuscript.

**Data availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

**Declarations**

**Conflict of interest** The author(s) declare no competing interests.

**Ethical statement** In this paper, the experiments did not use live cranes.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

**References**

1. Hall MA (1999) Correlation-based feature selection for machine learning. Doctoral Dissertation, Univ Waikato
2. Liu H, Motoda H (1999) Feature extraction, construction and selection: a data mining perspective. J Am Stat Assoc. https://doi.org/10.2307/2669967
3. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182. https://doi.org/10.1162/153244303322753616
4. Zhao S, Zhang Y, Xu HF, Han T (2019) Ensemble classification based on feature selection for environmental sound recognition. Math Probl Eng. https://doi.org/10.1155/2019/4318463
5. Saqlain SM, Sher M, Shah FA, Khan I, Ashraf MU, Awais M, Ghani A (2019) Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. Knowl Inf Syst 58:139–167. https://doi.org/10.1007/s10115-018-1185-y
6. Yong JY, Zhou ZM (2020) Multi-level feature selection algorithm based on mutual information. J Comput Appl 40:3478–3484
7. Ye TT, Liu MX, Zhang DQ (2016) Effective distance based multi-modality feature selection. Pattern Recognit Artif Intell 29:658–664
8. Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. Appl Soft Comput 62:441–453. https://doi.org/10.1016/j.asoc.2017.11.006
9. Jadhav S, He H, Jenkins K (2018) Information gain directed genetic algorithm wrapper feature selection for credit rating. Appl Soft Comput 69:541–553. https://doi.org/10.1016/j.asoc.2018.04.033
10. Xu HF, Zhang Y, Liu J, Lv DJ (2021) Feature selection method based on coefficient of variation and maximum feature tree. J Nanjing Norm Univ (Nat Sci Ed) 44:111–118
11. Maldonado S, López J (2018) Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification. Appl Soft Comput 67:94–105. https://doi.org/10.1016/j.asoc.2018.02.051
12. Liu YF, Li WB, Gao Y (2020) Adaptive neighborhood embedding based unsupervised feature selection. J Comput Res Dev 57:1639–1649. https://doi.org/10.7544/issn1000-1239.2020.20200219
13. Peng H, Long F, Ding C (2005) Feature selection based on mutual information criteria of maximandependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 27:1226–1238. https://doi.org/10.1109/tpami.2005.159
14. Lu SH, Sun M, Xie JH, Guo J, Jia WK, Su DY (2020) Fault location method of transmission line based on the maximum correlation-minimum redundancy algorithm. Electr Meas Instrum 57:79–85
15. Billah M, Waheed S (2020) Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection. Multimed Tools Appl 79:23633–23643. https://doi.org/10.1007/s11042-020-09151-7
16. Toaar M, Ergen B, Cmer Z (2019) A deep feature learning model for pneumonia detection applying combination of mRMR feature selection and machine learning models. IRBM 41:212–222. https://doi.org/10.1016/j.irbm.2019.10.006
17. Gu C, Yang Y, Zhang XX, Jin M, Zhou SY (2018) Feature selection for transformer fault diagnosis based on maximal relevance and minimal redundancy criterion. Adv Technol Electr Eng Energy 37:6
18. Erolu Y, Yildirim M, Inar A (2021) Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. Comput Biol Med 133:104407. https://doi.org/10.1016/j.compbiomed.2021.104407
19. Fan X, Huang X, Zhao Y, Wang L, Yu H, Zhao G (2022) Predicting prognostic effects of acupuncture for depression using the electroencephalogram. Evid Based Complement Alternat Med 2022:1381683. https://doi.org/10.1155/2022/1381683
20. Tuncer T, Dogan S, Baygin M, Rajendra Acharya U (2022) Tetromino pattern based accurate EEG emotion classification model. Artif Intell Med 123:102210. https://doi.org/10.1016/j.artmed.2021.102210
21. Sharma DK, Subramanian M, Malyadri P, Reddy BS, Sharma M, Tahreem M (2022) Classification of COVID-19 by using supervised optimized machine learning technique. Mater Today Proc 56:2058–2062. https://doi.org/10.1016/j.matpr.2021.11.388
22. Baliarsingh SK, Muhammad K, Bakshi S (2021) SARA: a memetic algorithm for high-dimensional biomedical data. Appl Soft Comput 101:107009. https://doi.org/10.1016/j.asoc.2020.107009
23. Feng YJ, Zhang FB (2014) Max-relevance min-redundancy restrictive BAN classifier learning algorithm. J Chongqing Univ (Nat Sci Ed) 37:71–77
24. Yao X, Wang XD, Zhang YX, Quan W (2013) A maximum relevance minimum redundancy hybrid feature selection algorithm based on particle swarm optimization. Control Desicion 28:413-417+423
25. Li SY, Wang GB (2021) New MRMR feature selection algorithm. CAAI Trans Intell Syst 16:649–661
26. Wang HH, Huang L, Zhou YW, Zhao YK (2019) Application of improved mRMR feature selection in human activity recognition. J Chongqing Univ Posts Tel-e kommun (Nat Sci Ed) 31:261–269
27. Jo I, Lee S, Oh S (2019) Improved measures of redundancy and relevance for mRMR feature selection. Compututers 8:42. https://doi.org/10.3390/computers8020042
28. Ahmed YA, Koer B, Huda S, Al-rimy B (2020) A system call refinement-based enhanced Minimum Redundancy Maximum Relevance method for ransomware early detection. J Netw Comput Appl 167:102753. https://doi.org/10.1016/j.jnca.2020.102753
29. Ahmed YA, Huda S, Al-Rimy B, Alharbi N, Saeed F, Ghaleb FA, Ali IM (2022) A weighted minimum redundancy maximum relevance technique for ransomware early detection in industrial IoT. Sustainability (Basel) 14:1–15
30. Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
31. UCI Machine Learning Repository. https://archive-beta.ics.uci.edu/ml/datasets. Accessed 1 Nov 2021
32. Grupo De Inteligencia Computacional. http://www.ehu.eus/ccwintco/index.php/. Accessed 1 Nov 2021
33. Liu QH, Liang ZY (2011) Optimized approach of feature selection based on information gain. Comput Eng Appl 47:130-132+136
34. Kannan SS, Ramaraj N (2010) A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm. Knowl-Based Syst 23:580–585. https://doi.org/10.1016/j.knosys.2010.03.016
35. Modinat M, Abimbola A, Abdullah ef B, Opeyemi A (2015) Gain ratio and decision tree classifier for intrusion detection. Int J Comput Appl 126:975–8887. https://doi.org/10.5120/ijca2015905983
36. Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: Proceedings of European Conference on Machine Learning Catania, Italy, April 6–8. https://doi.org/10.1007/3-540-57868-4_57
37. Mafarja M, Aljarah I, Faris H, Hammouri AI, Al-Zoubi AM, Mirjalili S (2019) Binary grasshopper optimisation algorithm approaches for feature selection problems. Expert Syst Appl 117:267–286. https://doi.org/10.1016/j.eswa.2018.09.015
38. Du ZG, Pan JS, Chu SC, Chiu YJ (2020) Improved binary symbiotic organism search algorithm with transfer functions for feature selection. IEEE Access 8:225730–225744. https://doi.org/10.1109/ACCESS.2020.3045043
39. Ghosh KK, Singh PK, Hong J, Geem ZW, Sarkar R (2020) Binary social mimic optimization algorithm with x-shaped transfer function for feature selection. IEEE Access 8:97890–97906. https://doi.org/10.1109/ACCESS.2020.2996611
40. Han F, Chen WT, Ling QH, Han H (2021) Multi-objective particle swarm optimization with adaptive strategies for feature selection. Swarm Evol Comput 62:100847. https://doi.org/10.1016/j.swevo.2021.100847
41. Kang Y, Wang HN, Tao L, Yang HX, Yang XK, Wang F, Li H (2022) Hybrid improved flower pollination algorithm and gray wolf algorithm for feature selection. Comput Sci 49:125–132. https://doi.org/10.11896/jsjknx.210600135
42. Zhang L, Wang C (2018) Multi-label feature selection algorithm based on joint mutual information of max-relevance and min-redundancy. J Commun 39:111–122. https://doi.org/10.11959/j.issn.1000-436x.2018082
43. Chen JT, Yuan SH, Lv DD, Xiang Y (2021) A novel self-learning feature selection approach based on feature attributions. Expert Syst Appl 183:115219. https://doi.org/10.1016/j.eswa.2021.115219

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.