IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 October 2022, accepted 7 November 2022, date of publication 17 November 2022, date of current version 22 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3222812

RESEARCH ARTICLE

Hybrid Feature Selection Method Based on Feature Subset and Factor Analysis

LIZENG GONG[1], SHANSHAN XIE[2], YAN ZHANG[1], MENGYAO WANG[1], AND XIAOYAN WANG[1]
[1]College of Mathematics and Physics, Southwest Forestry University, Kunming 650224, China
[2]College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China

Corresponding authors: Yan Zhang (zhangyan@swfu.edu.cn) and Xiaoyan Wang (849612371@swfu.edu.cn)

This work was supported in part by the Yunnan Provincial Science and Technology Department under Grant 202002AA10007, in part by the National Natural Science Foundation of China under Grant 61462078 and Grant 31860332, and in part by the Yunnan Provincial Department of Education under Grant 2022Y558.

ABSTRACT With the advent of big data era and the rapid improvement of raw data scale, feature selection, as the basis and critical technologies for data mining, plays an increasingly important role. However, most studies on feature selection methods, mainly directed to treat the single feature or overall feature subset, while the influence of the correlation and redundancy of features in the feature subset on the classification results is ignored. In this paper, a hybrid feature selection method based on feature subsets generated by factor analysis (FAFS_HFS) is proposed. Firstly, this method generates feature subsets from the maximum load (maximum explanatory power) of each feature through factor analysis. Then, minimal redundancy and maximal relevance (mRMR) and sequential forward selection (SFS) are used to remove the redundancy of each feature subset. Finally, fisher score based on feature subset (FSF-score) is utilized to evaluate and obtain the optimal feature subsets. Experiments are conducted on 14 datasets, the results show that FAFS_HFS method has higher classification accuracy and lower dimension on almost all datasets, especially in high-dimensional datasets, and it has competitive efficiency and classification performance compared with other methods.

INDEX TERMS Hybrid feature selection, factor analysis, mRMR, fisher score, sequential forward selection.

I. INTRODUCTION

Along with the amount of data generated by contemporary applications has grown dramatically in terms of the number of instances and features, and the ever-increasing sizes in actual data, feature selection has become an indispensable machine learning process in data pre-processing [1]. As one of the commonly used techniques in pattern recognition and data mining, especially in terms of the practicality of dealing with high-dimensional data, feature selection can efficiently eliminate weak correlation, noisy, redundant features and avoid dimensional disasters, thus boosting the accuracy of model learning and enhancing the capacity of classification prediction [2].

According to the relationship with the classifier, feature selection methods can be divided into four types: filter, wrapper, embedded and hybrid [3]. The filter method mainly defines the contribution of each feature and sorts them according to the correlation between features and classification labels. It is necessary to select an appropriate stop criterion due to the filter method only selects features with high ranking. The wrapper method is more dependent on the learning process than the filter method. It retrieves the most appropriate feature set by applying the greedy search strategy [4], [5]. But the wrapper method requires high computation and time cost. The embedded method is a built-in feature selection method. It monitors feature evaluation by repeatedly executing the learning algorithm, and optimizes the objective function in the process of training the classifier to realize feature selection. The hybrid method utilizes the combination of filter and wrapper or embedded and wrapper methods to complete feature selection.

Up to now, the process of feature selection is still a hotspot in the field of machine learning. How to reduce the dimension

The associate editor coordinating the review of this manuscript and approving it for publication was Long Xu.

of the dataset while ensuring high prediction accuracy is the major research issue. In recent years, many scholars have proposed different methods for the feature selection process of filter, wrapper and embedded. For example, Kaya and Fidan proposed a new filter feature selection method by using Pearson as the parametric correlation coefficient and Kendall as the nonparametric correlation coefficient [6]. Eftekhari et al. introduced a filter-based feature selection method to improve the classification performance of microarray datasets by means of selecting the significant features [7], and proposed two methods for unsupervised feature selection based on the spectral clustering [8]. In addition, Gonzalez-Lopez proposed mutual information maximization [9], Euclidean norm maximization and geometric mean maximization based on mutual information [10]. Xie et al. introduced an improved maximal relevance and minimal redundancy feature selection method based on feature subset to reduce the dimension of sample features and the training time of the model, and improve the classification performance [11].

Recently, hybrid feature selection method has gradually become a focus of research, and methods on hybrid feature selection were proposed. For example, scholars Jain and Singh introduced a two-phase hybrid feature selection method based on principal component analysis (PCA), reliefF and adaptive support vector machine [12]. Abasabadi et al. proposed a hybrid feature selection method based on SLI and genetic algorithm for microarray datasets [13]. Uzer et al. proposed a hybrid feature selection method based on sequential forward selection, sequential backward selection and PCA [14]. Besides, Alzaqebah et al. proposed a hybrid feature selection method based on particle swarm optimization and adaptive local search method [15]. Aziz proposed a hybrid machine learning framework based on a nature-inspired cuckoo search (CS) algorithm to minimize the number of selected genes, and maximize the classification accuracy of the used classifier [16], and a modified artificial bee colony metaheuristics optimization technique based on CS, naïve bayes and independent component analysis [17]. Ghosh et al. selected features by using the Relief, least absolute shrinkage and selection operator, and integrating the traditional classifiers with bagging and boosting methods in training process [18]. Tiwari developed a new hybrid feature selection method, namely, the iterative feature selection using dynamic butterfly optimization algorithm based on interaction maximization [19]. These hybrid feature selection methods can yield better classification ability than uses either method alone.

However, all the above hybrid feature selection methods do not fully consider the influence of synergies between features on the contribution of features, and ignore or not adequately evaluate the impact of correlation between features on classification [20]. And the coexistence of relevant features usually leads to information redundancy. Many feature selection methods can remove irrelevant features to gain "the m best features", but there may still be a lot of redundancy between them. To figure out this problem, methods based on feature clustering or feature grouping are proposed. Li et al. put forward a fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction [21]. Song et al. proposed a fast clustering-based feature selection algorithm, which used graph-theoretic clustering method to divide feature subsets, and selected the most representative feature strongly related to the target classes from each cluster [22]. Dehghan proposed an agglomerative hierarchical clustering-based method, and used mutual information to select the typical features in each cluster [23]. García-Torres et al. proposed a novel scatter search strategy, which used feature grouping to generate a population of diverse and high-quality solutions [24]. However, these methods mainly filter out irrelevant features or information on the overall feature subset, or do not fully consider the correlation and redundancy among features. Therefore, the most representative features in each feature subset can be selected by effective redundant filtering to obtain high classification accuracy and effectively reduce the raw feature set dimension.

In this paper, we propose a hybrid feature selection method FAFS_HFS based on feature subsets generated through factor analysis (FA). It uses FA to generate multiple feature subsets with strong internal correlation and weak external correlation. At the same time, since there may be some relatively irrelevant features that will be integrated into the same feature subset, and to better consider the correlation and redundancy among features of each feature subset, minimal redundancy maximal relevance based on feature subset (FSmRMR), which combining mRMR and SFS, is utilized to remove the redundancy of each feature subset. And fisher score based on feature subset (FSF-score) is utilized to evaluate and select the feature subsets after removing redundancy. The proposed FAFS_HFS method effectively considers the correlation and coordination between features on classification.

The main contributions of this paper are as follows:

(1) Compared with traditional feature selection methods for single feature selection, this paper explores a hybrid feature selection method based on factor analysis from the perspective of feature subset.
(2) Compared with the traditional mRMR and F-score methods, the proposed method fully considers the influence of correlation, redundancy and joint influence between features on feature selection results.
(3) The experimental results of 14 datasets illustrate that the proposed method has superior performance than other feature selection methods involved. And the dimension reduction rate achieved on the 4702-dimensional dataset is about 98.5%.

The rest of this paper is organized as follows. Section 2 introduces the methods used in this paper. Section 3 constructs the model of proposed method FAFS_HFS in detail. And the experiments and results are analyzed in Section 4. Then, the discussion about the comparison of FAFS_HFS with other feature selection methods and the sensitive analysis are conducted in Section 5. Finally, we give conclusion and the works will be done in the future.

II. METHODS

A. FACTOR ANALYSIS

Factor analysis (FA) as a popular multivariate statistical technique transforms some dependent features into some other features called factors, so that the first factor of this transformation has the main information of the first dataset [25], [26]. Generally, FA has three main functions: (1) reduce factor dimension; (2) calculate factor weight; (3) calculate the summary score of the weighting factor. Under the condition of not or less losing the raw data information as much as possible, it aggregates many complicated variables into a few independent common factors, which can reflect the main information of the raw many variables. It reduces the number of variables, and reflects the internal relationship between variables.

In other words, FA is the transformation of a high-dimensional dataset into a low-dimensional dataset by considering the minimum factor to reduce the dimension of the transformed dataset. That is, FA is performed by examining patterns of correlation (or covariance) between observed measures. Highly correlated (positive or negative) measures may be affected by the same factors, while those relatively unrelated measures may be influenced by different factors [27].

According to the above idea, we can construct feature subsets with strong internal correlation and weak external correlation through FA to minimize the influence of redundant features on feature selection. Reduce the dimension of high-dimensional data as much as possible in the form of feature subset constructed by FA, so as to achieve the goal of efficient feature selection. It is the key point of this paper to generate feature subsets through FA.

B. MINIMAL REDUNDANCY AND MAXIMAL RELEVANCE

Minimal redundancy and maximal relevance (mRMR) is a method of finding a feature set that meets the characteristics of maximal relevance between feature set and categories and minimal redundancy among features within the feature set [28], [29]. The mRMR approximates the theoretically optimal maximum-dependency feature selection algorithm, which maximizes the mutual information between the selected feature and the distribution of classification variables, that is, the feature set has the maximum dependence on categories. The calculation method is as follows:

Given two random variables X and Y, their probability density functions (corresponding to the continuous variable) are $p\,(x)\,, p\,(y)\,, p\,(x, y)\,(x\epsilon X, y\epsilon Y)$, then the mutual information is shown in equation (1):

$$MI\,(X; Y) = \int\int p(x, y)log\frac{p(x, y)}{p(x)p(y)}dxdy \tag{1}$$

If X and Y are a series of discrete sequences, mutual information can be expressed as equation (2):

$$MI\,(X; Y) = \sum\nolimits_{x\epsilon X}\sum\nolimits_{y\epsilon Y} p(x, y)log\frac{p(x, y)}{p(x)p(y)} \tag{2}$$

The correlation between feature set S and class C is defined by the average of all mutual information values between each feature f_m and class C (output). The definition is shown in equation (3):

$$D\,(S) = \frac{1}{|S|}\sum\nolimits_{f_m\epsilon S} MI(f_m; C) \tag{3}$$

The redundancy of all features in set S is the average of all mutual information values between feature f_m and feature f_n as equation (4):

$$R\,(S) = \frac{1}{|S|^2}\sum\nolimits_{f_m,f_n\epsilon S} MI(f_m; f_n) \tag{4}$$

mRMR is the combination of the maximum difference between the above correlation and redundancy. It's defined as equation (5):

$$mRMR = max\phi(D\,(S)\,, R(S)) \tag{5}$$

where $\phi = (D\,(S) - R(S))$.

C. FISHER SCORE

Fisher score (F-score) method is a general method to calculate maximum likelihood estimate (MLE), and its essence approximates Newton's method. The key idea of F-score is to find a subset of features, such that the distance between data points of different classes is as large as possible, and the distance between data points of the same class is as small as possible [30], [31]. The features are sorted by the score of each feature, and finally the optimal feature subset is obtained [32].

The F-score of mth feature is shown in equation (6):

$$Fscore\,(f_m) = \frac{\sum_{c=1}^{N} n_c(\mu_c - \mu_m)^2}{\sum_{c=1}^{N} n_c\sigma_c^2} \tag{6}$$

where N represents the number of categories, n_c is the number of $c(c = 1, 2\ldots, N)$ class samples; μ_c and σ_c represent the mean and variance of mth feature in class c samples. μ_m is the global mean of mth feature. The higher the F-score, the better the feature.

D. SEQUENTIAL FORWARD SELECTION

Sequential forward selection (SFS), a feature selection method proposed by Whitney, is one of the simplest greedy algorithms [14]. SFS begins with zero attributes, evaluates all feature subsets with one feature, and selects the one with the best performance [33]. Then the selected features are added to the subset to generate a better subset. This process repeats until there is no improvement in the subset. However, SFS is generally suboptimal and suffers from the so-called "nesting effect" [34]. The features at the bottom of the ranking will be no chance to be selected. Therefore, SFS needs to be combined with feature evaluation method, and it is particularly important to select a suitable feature evaluation method.

III. FAFS_HFS MODEL CONSTRUCTION

This paper proposes the hybrid feature selection method FAFS_HFS based on feature subset and FA. This method comprises three phases, which are shown as follows:

(1) Phase I. Generate feature subsets through FA: the features of the raw datasets are divided into multiple feature subsets by FA, and construct strong correlation between features within subsets and weak correlation between subsets.

(2) Phase II. Remove redundant features of each subset by FSmRMR: FSmRMR method is explored to remove the redundant features from each subset, with adjusting the accuracy and dimension reduction rate of each subset.

(3) Phase III. Evaluate feature subsets through FSF-score: FSF-score is used to measure the contribution of each feature subset after redundancy removal, then feature subsets are selected according to FSF-score.

The specific framework of FAFS_HFS is as shown in Figure 1.

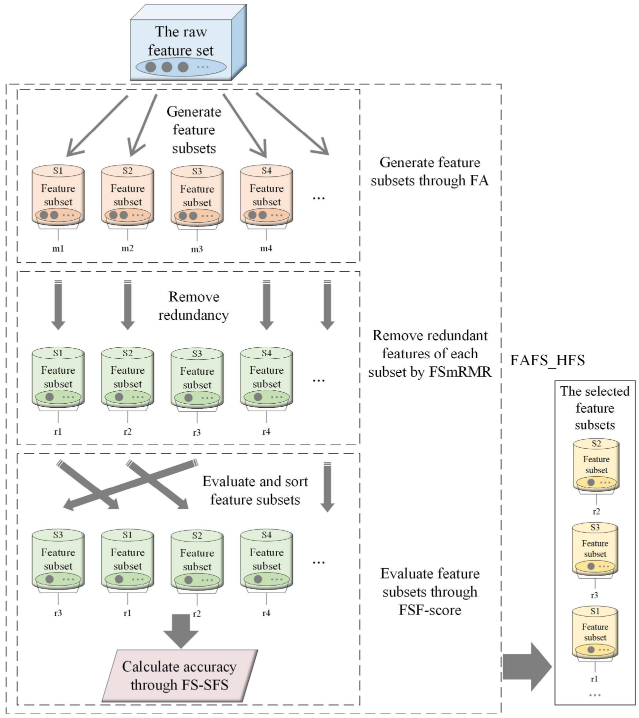

FIGURE 1. The phase diagram of proposed FAFS_HFS method.

A. GENERATE FEATURE SUBSETS THROUGH FATOR ANALYSIS

The feature selection aims to reduce or filter out redundant and irrelevant features, improve the overall correlation between features, so as to reduce the complexity of learning tasks, and enhance the classification accuracy and dimensionality reduction rate.

FA can drop dimensions according to the dependence between variables and features, and the raw features will be divided as feature subsets by the maximum explanatory power, to construct multiple feature subsets with strong internal correlation and weak external correlation. This method can effectively avoid the influence of features with weak correlation with the classification labels, and further prevent the elimination of non-redundant features during dimensionality reduction. Figure 2 shows the process of generating feature subsets through FA.

The steps of generating feature subsets are as follows:

1) Calculate the correlation coefficient matrix R of features and produce the elementary load matrix A, and calculate the variance contribution rate and cumulative variance contribution rate of each factor.
2) The load matrix A will be rotated depending on the number of feature subsets, which is determined according to the contribution rate of cumulative variance.
3) According to the maximum load of each feature, that is, the maximum explanatory power, determine the subset to which the feature belongs, and ultimately generate the feature subset S_i with m_i features.

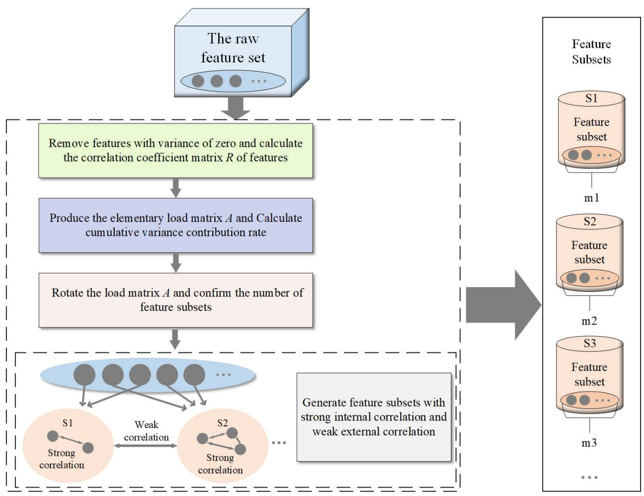

FIGURE 2. Generate feature subsets through factor analysis (FA).

B. REMOVE REDUNDANT FEATURES OF EACH SUBSET BY FSMRMR

This phase purposes to adjust the accuracy and dimension reduction rate of each feature subset through redundant removal. In FSmRMR method, first calculate the accuracy of each feature subset. Next, the features of the number of m_i in the feature subset S_i are ranked according to the correlation between features and categories C and the redundancy among features. Then, select features in feature subset S_i through SFS to remove the redundant features. Finally, the number of remaining features in feature subset S_i is r_i.

The calculation of the FSmRMR method is based on equations (3), (4) and (5). The process of removing redundant features of each subset by FSmRMR is shown in Figure 3.

C. EVALUATE AND SELECT FEATURE SUBSETS THROUGH FSF-SCORE

The fisher score based on feature subset (FSF-score) is designed to evaluate and select feature subsets. Firstly, the F-score of features of the number of r_i in feature subset S_i is calculated. Second, according to the F-score of all features in feature subset S_i, the average F-score of feature subset S_i is obtained, that is, the contribution degree Cb_i of feature subset S_i. Equation (7) shows the calculation of Cb_i.

$$Cb_i = \frac{1}{r_i} \sum_{m=1}^{r_i} \frac{\sum_{c=1}^{N} n_c(\mu_c - \mu_m)^2}{\sum_{c=1}^{N} n_c \sigma_c^2} \tag{7}$$

where m represents the mth feature. Third, rank feature subsets according to the contribution. Finally, select feature subsets using SFS based on feature subset (FS_SFS) method. Therefore, the $f_m(m = 1, 2, \ldots, r_i)$ in equation (6) is the features in the feature subset S_i.

Figure 4 produces the process of evaluating and selecting feature subsets through FSF-score. In sum, the detailed description of the FAFS_HFS is shown in Algorithm 1.

Algorithm 1 The Proposed FAFS_HFS Method

Input: $D(F_1, F_2, F_3, \ldots, F_m, C)$ – the given data set, *Max_cumC* – maximum cumulative variance contribution rate
Output: S – selected feature subsets.
//==Phase 1: Generate feature subsets through FA ==
R = correlation coefficient matrix (D)
[*COEFF*, *Contribution*] = FA (R)
// COEFF: coefficient matrix, Contribution: variance contribution rate
Subsets_num = cumulative contribution rate < *Max_cumC*
A = elementary load matrix (R)
Rotated_matrix = Rotate (A(1: *Subsets_num*, :)) ∪A(*Subsets_num*+1: end, :)
S = generate empty feature subsets in number of *Subsets_num*
for *each* $F_j \in D$ **do**
  max = Max (*Rotated_matrix* (j, :))
  $F_j \rightarrow S_{max}$
end
//==Phase 2: Remove redundant features of each subset by FSmRMR ==
for i = *1 to Subsets_num* **do**
  $D(S_i)$ = correlation between features ($F_j \in S_i$) and classification tags
  $R(S_i)$ = redundancy between features ($F_j \in S_i$)
  $S_i = SFS(\max \phi_i (D(S_i), R(S_i)))$*// SFS: Sequential forward selection*
end
//==Phase 3: Evaluate and select feature subsets through FSF-score ==
for i = *1 to Subsets_num* **do**
  Cb_i = average F-score of features ($F_j \in S_i$)
end
temp_S = sort (Cb_i)*// Feature subsets are sorted according to* Cb_i
S = FS_SFS (*temp_S*) *// FS_SFS: Sequential forward selection based on feature subset*
return S

IV. EXPERIMENTS AND RESULTS ANALYSIS

A. EXPEIMENT DATASETS

In this paper, fourteen datasets are selected. Twelve of the datasets are selected from UCI Machine Learning Repository [35] publicly available datasets: Vehicle silhouettes (Vehicle), Ionosphere (Iop), Dermatology (Dml), Connectionist bench (CB), Synthetic control chart time series (SC), Libras movement (LM), Musk (Musk), Urban land cover (ULC), Cardiac arrhythmia (CA), CNAE-9 (CNAE), MicroMass (MM), DBWorld e-mails (DBW). One dataset is crane songs dataset (Crane), and the other one is the hyperspectral dataset of Pavia University (PU) [36].

The dimension of the datasets ranges from 18 to 4702, and the number of instances ranges from 64 to 2700. Each dataset is divided into training set and test set by 7:3. The description of datasets is shown in Table 1.

TABLE 1. The description of datasets.

Dataset	Instances	Training set	Test set	Features	Classes
Vehicle	946	662	284	18	4
Iop	351	246	105	34	2
Dml	366	257	109	34	6
CB	208	146	62	60	2
SC	600	420	180	60	6
Crane	343	243	100	75	7
LM	360	252	108	90	15
PU	2700	1800	900	103	9
ULC	675	507	168	147	9
Musk	476	334	142	166	2
CA	452	317	135	279	13
CNAE	1080	756	324	856	9
MM	931	401	170	1300	9
DBW	64	46	18	4702	2

B. EXPERIMENTAL DESIGN AND SETUP

To verify the effectiveness of the proposed method, three groups of experiments are designed for comparison and analysis.

The design of the experiments is shown in Figure 5.

In the experiments, all related numerical calculation and model training are based on MATLAB and WEKA software. The operating condition is Intel(R) Core (TM) i5-6300U CPU @ 2.40GHz/16384MB RAM/Windows 10/MATLAB R2018a/WEKA 3.8.6. This paper uses accuracy (Acc), dimension reduction (DR) and overall evaluation (OE, considering both accuracy and dimension reduction) indexes to evaluate the performance of feature selection methods.

The Acc index represents the proportion of the sample number accurately classified to the total sample number after model training, and the formula of Acc is shown in equation (8). Where T_c and F_c represent samples with correct and wrong classification respectively.

$$Acc = \frac{T_c}{T_c + F_c} \tag{8}$$

The DR represents the feature dimension reduction rate The calculation of DR is shown in equation (9):

$$DR = 1 - \frac{D_{selected}}{D_{raw}} \tag{9}$$

where $D_{selected}$ is the number of selected features, and D_{raw} is the number of raw features.

In feature selection, not only the accuracy of model classification, but the dimension reduction rate should be considered. Therefore, in this study, we consider the classification

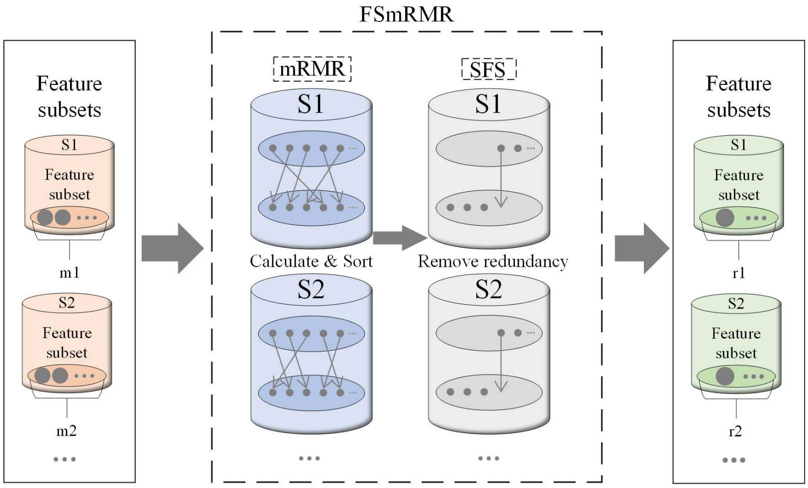

FIGURE 3. Remove redundant features of each subset by FSmRMR.

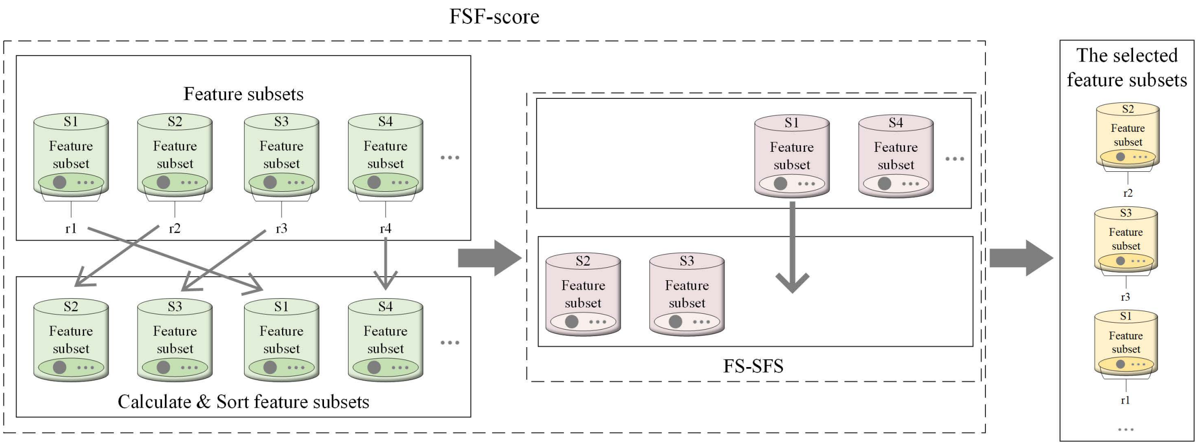

FIGURE 4. Evaluate and select feature subsets through FSF-score.

accuracy and dimension reduction rate of the model at the same time, the OE index is defined in equation (10):

$$OE = \alpha_w * Acc + (1 - \alpha_w) * DR \tag{10}$$

where coefficient α_w represents the weight. Its value is between 0 and 1. Here, $\alpha_w = 0.5$.

The REP-Tree is used for model training and classification.

C. EXPERIMENT I: HYBRID OF MRMR AND F-SCORE (MF)

For the single feature, both mRMR and F-score are effective feature selection methods. The former one specializes on dimensionality reduction while the later one with accuracy improving. Therefore, we combine them to a new method, mRMR F-score (MF) and compare performance with the traditional mRMR and F-score methods.

In MF method, mRMR is used to reduce the dimension of the raw feature set. Then, feature selection is carried out in the remaining feature set by F-score method. Figure 6 shows the result of the Experiment I. Seen from Figure 6, in terms of Acc index, MF method reaches

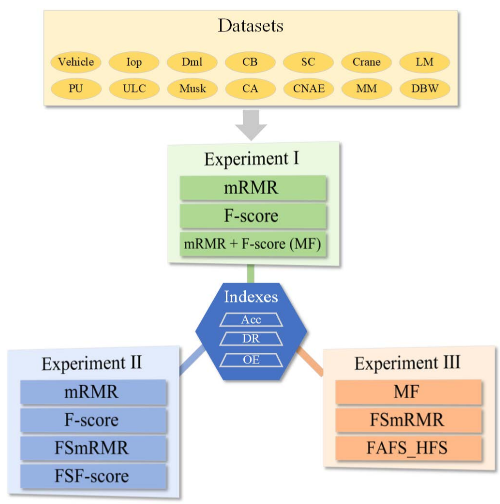

FIGURE 5. The design of experiments.

52.00-93.58%, same as mRMR, while F-score reaches 46.30-96.33%. In terms of DR index, it can be clearly

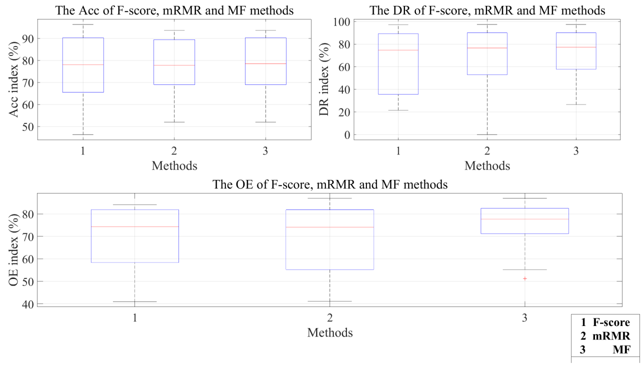

FIGURE 6. Comparison of F-score and mRMR with MF method.

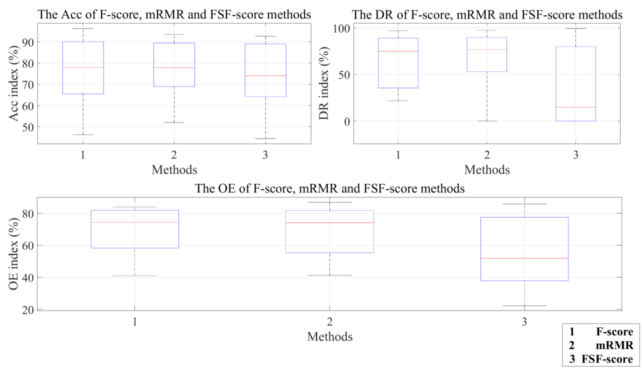

FIGURE 7. Comparison of F-score and mRMR with FSF-score method.

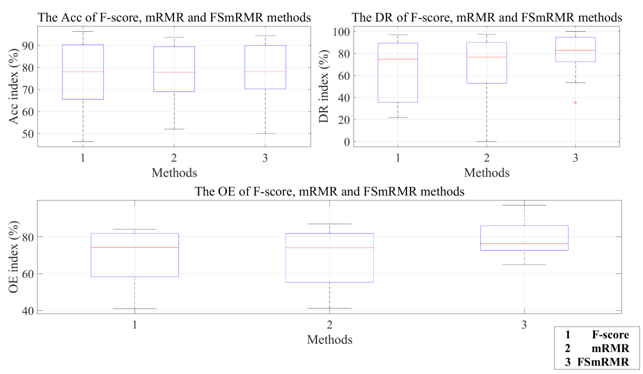

FIGURE 8. Comparison of F-score and mRMR with FSmRMR method.

found that MF reaches 26.67-97.33%, while the mRMR is 0.00-97.33% and F-score is 21.67-96.99%. Therefore, in OE index, MF is better than mRMR and F-score, and the maximum value of OE differs from the minimum value by about 35%, while the mRMR and F-score is about 45%.

The experimental results illustrate that the combination of mRMR and F-score can improve the DR index on the premise of ensuring the Acc, as well as OE of feature selection method. In sum, the combination of mRMR and F-score has a preferable feature selection performance.

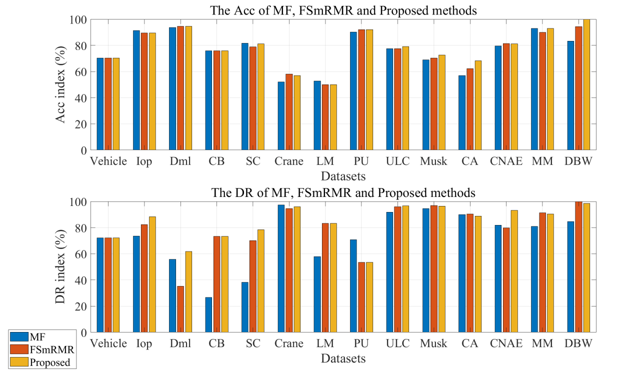

FIGURE 9. Compare the Acc and DR indexes of MF and FSmRMR with the proposed method.

D. EXPERIMENT II: COMBINATION OF FA WITH MRMR (FSMRMR) AND F-SCORE (FSF-SCORE)

Considering the limitations of traditional feature selection methods for aiming to the single feature, we explore combining FA with mRMR and F-score to form two new methods FSmRMR and FSF-score. The FSmRMR selects feature subsets without sorting feature subsets after removing redundant features of each subset, while the FSF-score selects feature subsets by sorting feature subsets without removing redundancy. Finally, the selected feature subsets are obtained through FS_SFS.

In Experiment II, FA is combined with mRMR and F-score respectively based on the idea of mRMR remove redundancy and F-score evaluate feature subset. The results of Experiment II is shown in Figure 7.

Compared with F-score, FSF-score shows the feature selection performance with great span and instability, and only has better performance in CA, CNAE, MM and DBW datasets, with its dimensions reach 279-4702.

However, the FSmRMR method gives completely different results. Compared with mRMR method, the Acc and DR of FSmRMR reach 50.00-94.49% and 35.29-99.85% respectively, while the mRMR reach 52.00-93.58% and 0.00-97.33%. In terms of OE, the mean of FSmRMR is 78.71% and the variance is 84.72, while the mRMR is 69.73% and 260.40, while the mean and variance of OE of F-score is 69.63% and 200.22. Therefore, the FSmRMR method has better and more stable feature selection performance than mRMR and F-score.

Overall, compared with mRMR and F-score, FSmRMR method can improve the classification accuracy and dimension reduction rate at the same time, and FSmRMR presented obvious advantages in feature selection performance. Generally speaking, it is effective to use FA to generate feature subsets and achieve feature selection. In addition, through the different performance of FSmRMR and FSF-score, it can be found that it is necessary to remove the redundancy of each feature subset in feature selection while using FA to generate feature subsets.

E. EXPERIMENT III: COMPARISON OF FAFS_HFS WITH MF AND FSMRMR

In Experiment III, we combined FA with FSmRMR, FSF-score methods together as the proposed method FAFS_HFS, and endowed with the functions of "subsets generation", "redundancy removal" and "subsets evaluation", respectively.

Experiment III compares the proposed FAFS_HFS with MF and FSmRMR methods. The number of features extracted and the feature subsets generated are shown in Table 2. And the experimental results are listed in Table 3.

Compared Acc and DR indexes of F-score and FSmRMR with MF methods in Figure 9, it can be explored that the proposed FAFS_HFS method can effectively improve the classification accuracy from 0.92% to 16.67%. However, it can be noted that the degree of accuracy improvement is related to the dimension of the dataset. The improvement performance is not obvious or increased in low-dimensional datasets, but it will be significantly improved when the dimension of datasets reaches 147-4702. Therefore, FAFS_HFS method has an obvious effect on improving classification accuracy in middle-dimensional and high-dimensional datasets.

Moreover, FAFS_HFS maintains a relatively good dimension reduction rate. Compared with FSmRMR method, FAFS_HFS illustrates obvious effect on dimension reduction in Iop, Dml, SC, Crane, ULC and CNAE datasets from 0.68% to 26.47%. And compared with MF method, better dimension reduction rate can be obtained in Iop, Dml, CB, SC, LM, ULC, Musk, CNAE, MM, DBW datasets from 1.81% to 46.67%.

TABLE 2. The number of features selected and subsets generated.

Dataset	Raw features	Feature subsets	MF # features	FSmRMR # features	FAFS_HFS # features
Vehicle	18	1	5	5	5
Iop	34	4	9	6	4
Dml	34	14	15	22	13
CB	60	11	44	16	16
SC	60	4	37	18	13
Crane	75	5	2	4	3
LM	90	3	38	15	15
PU	103	1	30	48	48
ULC	147	3	12	6	5
Musk	166	3	9	5	6
CA	279	28	28	27	31
CNAE	856	70	155	173	59
MM	1300	77	247	112	125
DBW	4702	22	719	7	71

Seen from Figure 10, 13 out of 14 datasets, FAFS_HFS method has higher or equal OE index against MF and FSmRMR methods. In Iop, Dml, SC, Crane, ULC, Musk, CA, CNAE, MM and DBW datasets, the improvement of OE index of FAFS_HFS method reaches 0.17-6.38%, especially in high-dimensional datasets as CNAE, MM and DBW, whose dimension reaches 856 - 4702.

In conclusion, FA is an effective method to generate feature subsets for feature selection. Combining FA with other traditional feature selection methods will obtain optimal feature subset and outperform each single method.

TABLE 3. Comparison of MF and FSmRMR with FAFS_HFS method.

Dataset	Index	Raw	MF	FSmRMR	FAFS_HFS
Vehicle	ACC (%)	67.924	70.283	70.283	70.283
	DR (%)	0.000	72.222	72.222	72.222
	OE (%)	33.622	71.253	71.253	**71.253***
Iop	ACC (%)	89.523	91.429	89.524	89.524
	DR (%)	0.000	73.529	82.353	88.235
	OE (%)	44.762	82.479	85.938	**88.879***
Dml	Acc (%)	92.661	93.578	94.495	94.495
	DR (%)	0.000	55.882	35.294	61.765
	OE (%)	46.330	74.730	64.895	**78.130***
CB	Acc (%)	75.806	75.806	75.806	75.806
	DR (%)	0.000	26.667	73.333	73.333
	OE (%)	37.903	51.237	74.570	**74.570***
SC	Acc (%)	78.333	81.667	78.889	81.111
	DR (%)	0.000	38.333	70.000	78.333
	OE (%)	39.166	60.000	74.444	**79.722***
Crane	Acc (%)	49.000	52.000	58.000	57.000
	DR (%)	0.000	97.333	94.667	96.000
	OE (%)	24.500	74.667	76.333	**76.500***
LM	Acc (%)	44.444	52.778	50.000	50.000
	DR (%)	0.000	57.778	83.333	83.333
	OE (%)	22.222	55.278	66.667	**66.667***
PU	Acc (%)	89.000	90.222	92.111	92.111
	DR (%)	0.000	70.874	53.398	53.398
	OE (%)	44.500	**80.548***	72.755	72.755
ULC	Acc (%)	64.881	77.381	77.381	79.167
	DR (%)	0.000	91.837	95.918	96.599
	OE (%)	32.440	84.609	86.650	**87.883***
Musk	Acc (%)	63.380	69.014	70.423	72.535
	DR (%)	0.000	94.578	96.988	96.386
	OE (%)	31.690	81.796	83.705	**84.460***
CA	Acc (%)	26.667	57.037	62.222	68.148
	DR (%)	0.000	89.964	90.323	88.889
	OE (%)	13.333	73.501	76.272	**78.519***
CNAE	Acc (%)	79.012	79.629	81.482	81.173
	DR (%)	0.000	81.893	79.790	93.108
	OE (%)	39.506	80.761	80.636	**87.140***
MM	Acc (%)	90.000	92.941	90.000	92.941
	DR (%)	0.000	81.000	91.385	90.385
	OE (%)	45.000	86.971	90.692	**91.663***
DBW	Acc (%)	72.222	83.333	94.444	100.000
	DR (%)	0.000	84.709	99.851	98.490
	OE (%)	36.111	84.021	97.148	**99.245***

V. DISCUSSION

A. COMPARISON OF FAFS_HFS WITH OTHER FEATURE SELECTION METHODS

In the previous experiments, we made vertical comparison about the feature selection performance of FAFS_HFS. In this section, we will evaluate the classification ability of FAFS_HFS from a horizontal perspective. We compared FAFS_HFS with seven widely used feature selection methods: information gain (IG), symmetrical uncertainty (SU), gain ratio (GR), reliefF (RlF), correlation-based feature selection (CBF), fast correlation-based filter (FCBF) and the recent state of the art works in feature selection method by

TABLE 4. Comparison of proposed method with other feature selection methods.

Dataset	Index	IG	SU	GR	RlF	CBF	FCBF	MIM	FAFS_HFS
Vehicle	Acc (%)	67.924	68.868	68.396	67.924	67.924	52.830	71.698	70.283
	DR (%)	0.000	38.889	44.444	33.333	0.000	89.474	38.889	72.222
	OE (%)	33.962	53.878	56.420	50.629	33.962	71.152	55.293	**71.253***
Iop	Acc (%)	90.476	91.429	91.429	89.524	89.524	84.762	89.524	89.524
	DR (%)	76.471	23.529	70.588	5.882	58.823	91.429	64.706	88.235
	OE (%)	83.473	57.479	81.008	47.703	74.174	88.095	77.115	**88.879***
Dml	Acc (%)	95.413	95.413	95.413	95.413	93.578	93.578	95.413	94.495
	DR (%)	58.823	55.882	38.235	61.765	44.118	54.286	44.118	61.765
	OE (%)	77.118	75.648	66.824	**78.589***	68.848	73.932	69.765	78.130
CB	Acc (%)	75.806	75.806	75.806	75.806	75.806	53.226	75.806	75.806
	DR (%)	8.333	6.667	6.667	1.667	3.333	96.721	25.000	73.333
	OE (%)	42.070	41.237	41.237	38.737	39.570	**74.973***	50.403	74.570
SC	Acc (%)	78.333	81.667	81.111	79.444	80.000	57.222	79.444	81.111
	DR (%)	0.000	33.333	26.667	33.333	20.000	96.721	45.000	78.333
	OE (%)	39.167	57.500	53.889	56.389	50.000	76.972	62.222	**79.722***
Crane	Acc (%)	60.000	56.000	51.000	50.000	52.000	47.000	53.000	57.000
	DR (%)	92.000	89.333	88.000	93.333	90.667	97.368	92.000	96.000
	OE (%)	76.000	72.667	69.500	71.667	71.333	72.184	72.500	**76.500***
LM	Acc (%)	48.148	48.148	47.222	46.296	46.296	14.815	48.148	50.000
	DR (%)	32.222	57.778	30.000	45.556	45.555	97.802	55.556	83.333
	OE (%)	40.185	52.963	38.611	45.926	45.926	56.308	51.852	**66.667***
PU	Acc (%)	89.667	89.333	90.333	89.444	89.333	60.111	89.222	92.111
	DR (%)	62.136	69.903	78.641	35.922	85.437	98.077	20.388	53.398
	OE (%)	75.901	79.618	84.487	62.683	**87.385***	79.094	54.805	72.755
ULC	Acc (%)	67.262	69.643	66.667	73.809	68.452	27.381	73.214	79.167
	DR (%)	74.150	68.707	27.891	68.027	70.748	98.649	95.918	96.599
	OE (%)	70.706	69.175	47.280	70.918	69.600	63.015	84.566	**87.883***
Musk	Acc (%)	66.197	64.084	69.718	66.197	69.718	54.930	65.493	72.535
	DR (%)	94.578	99.398	98.193	93.976	96.988	98.802	99.398	96.386
	OE (%)	80.388	81.741	83.955	80.086	83.353	76.866	82.445	**84.460***
CA	Acc (%)	57.778	58.518	58.518	62.222	54.815	53.333	54.074	68.148
	DR (%)	89.606	89.964	95.699	68.459	90.681	94.643	99.642	88.889
	OE (%)	73.692	74.241	77.109	65.340	72.748	73.988	76.858	**78.519***
CNAE	Acc (%)	79.630	79.630	79.630	79.938	81.481	73.765	80.864	81.173
	DR (%)	34.112	33.995	33.995	5.958	77.804	94.049	80.724	93.108
	OE (%)	56.871	56.812	56.812	42.948	79.643	83.907	80.794	**87.140***
MM	Acc (%)	90.000	90.000	90.588	91.765	92.353	61.176	90.000	92.941
	DR (%)	0.000	0.000	1.846	39.308	75.000	99.154	73.231	90.385
	OE (%)	45.000	45.000	46.217	65.536	83.676	80.165	81.615	**91.663***
DBW	Acc (%)	88.889	88.889	88.889	100.000	83.333	83.333	83.333	100.000
	DR (%)	96.746	97.448	96.044	97.214	84.560	99.490	84.687	98.490
	OE (%)	92.817	93.168	92.466	98.607	83.946	91.411	84.010	**99.245***

using mutual information: mutual information maximization (MIM) [9]. The experiment results are shown in Table 4.

The results illustrate that the Acc of FAFS_HFS is superior to IG, SU, GR, RlF, CBF, FCBF and MIM methods in high-dimensional datasets. When dimension reaches 90-4702, the improvement of Acc of FAFS_HFS is from 0.58% to 5.92%. In the terms of dimension reduction rate DR, Table 4 shows that due to the strong dimension reduction capability of FCBF method, FAFS_HFS has no advantage in DR. Compared with the other six methods from DR index, FAFS_HFS has advantages in both low-dimensional and high-dimensional datasets, with a maximum increase of 48.33%. However, because of the lack of prediction accuracy of FCBF method, in OE index, FAFS_HFS has still maintained a higher index value than FCBF. And in all 14 datasets, the OE index of FAFS_HFS is better than IG, SU, GR, RlF, CBF, FCBF and MIM methods in 11 datasets, especially in high-dimensional datasets, and the improvement reaches from 0.10% to 10.36%.

To further verify the performance of the proposed FAFS_HFS method, Friedman test is conducted on the OE index of IG, SU, GR, RlF, CBF, FCBF, MIM and the proposed FAFS_HFS feature selection methods.

Let r_i^j denote the ranking of the jth $(1 \leq j \leq k)$ algorithm on the ith $(1 \leq i \leq N)$ dataset. The Friedman test compares the average ranking of algorithms, and the equation is shown in (11).

$$R_j = \frac{1}{N} \sum\nolimits_i r_i^j \tag{11}$$

TABLE 5. The results of sensitive analysis of proposed method under different α_w.

Dataset	α_w value range with strong sensitivity
Vehicle	weak sensitivity
Iop	weak sensitivity
Dml	0.2-0.3
CB	0.2-0.3
SC	0.2-0.3
Crane	0.6-0.7
LM	0.2-0.3
PU	weak sensitivity
ULC	weak sensitivity
Musk	weak sensitivity
CA	0.4-0.5
CNAE	0.1-0.2
MM	0.3-0.4
DBW	0.2-0.3

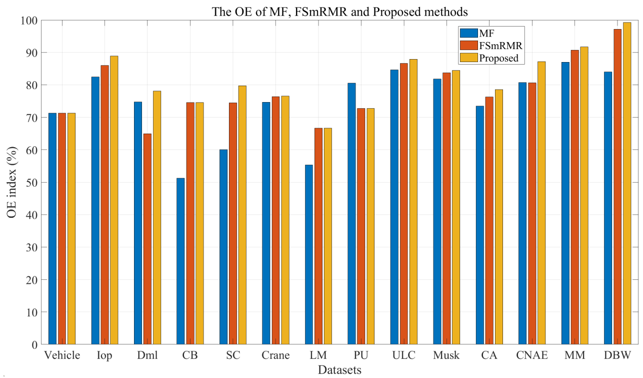

FIGURE 10. Compare the OE index of MF and FSmRMR with FAFS_HFS method.

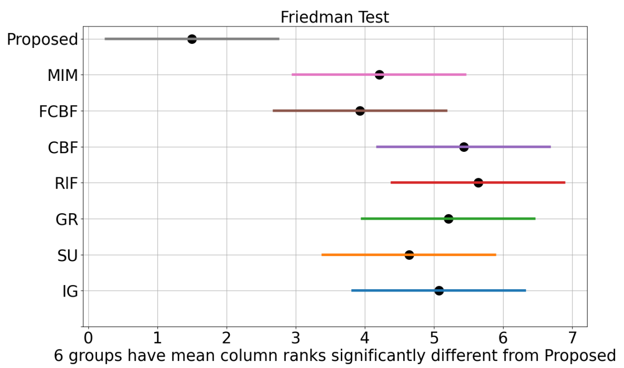

FIGURE 11. The Friedman test of IG, SU, GR, RIF, CBF, FCBF, MIM and the proposed feature selection method.

Under the null hypothesis that all algorithms are equivalent and therefore their ranking R_j should be equal, equation (12) shows the distribution of the Friedman statistic.

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{12}$$

where χ_F^2 represents the chi-square value. Iman and Davenport show that Friedman's χ_F^2 is too conservative and come up with a better statistical formula, which is shown in equation (13).

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \tag{13}$$

where k and N represents the number of algorithms and datasets respectively. This formula follows an F-distribution with degrees of freedom of $(k-1)$ and $(k-1)(N-1)$. In this paper, $k = 8$, $N = 14$, and the confidence interval $\alpha = 0.05$. After calculation, $F_F = 3.28$ and $F(3, 91) = 2.43$ can be obtained. It is shown that $F(3, 91) < F_F$. This indicates that there are significant differences between the algorithms, and subsequent tests can be carried out. In the study, Bonferroni-Dunn follow-up test is used, and FAFS_HFS method is set as the control algorithm. In general, critical difference (CD) controls the family-wise error rate, and the equation is shown in (14).

$$CD = q_\alpha \times \sqrt{\frac{k(k+1)}{6N}} \tag{14}$$

where, q_α is the critical value. By querying the critical value table, $q_{0.05} = 2.69$ can be obtained, and the result of equation (14) can be calculated as shown in (15).

$$CD = 2.69 \times \sqrt{\frac{8 \times (8+1)}{6 \times 14}} = 2.49 \tag{15}$$

Figure 11 shows the results of the Friedman test.

It is obvious that the proposed FAFS_HFS has significant differences with IG, SU, GR, RIF, CBF and MIM methods, indicating that FAFS_HFS has better performance and generalization ability, especially in high-dimensional datasets.

B. SENSITIVE ANALYSIS

In the previous chapter 4, OE index was taken as the main evaluation index when determining evaluation indexes. And the Acc and DR indexes are given the same weight $\alpha_w = 0.5$ in equation (10). However, different weights may bring different classification results. For example, for low-dimensional datasets, DR will fluctuate dramatically with the change of the number of selected features, while the high-dimensional datasets focus more on classification accuracy. Therefore, in this part, we will discuss the influence of different weights α_w on our results.

We took α_w from 0.1 to 0.9 with the step size of 0.1, so as to compare the changes of Acc and DR indexes in various datasets under different α_w through FAFS_HFS method.

The results in Table 5 show that when $\alpha_w \geq 0.3$, Acc and DR indexes will basically keep stable, only CA, Crane and MM datasets when α_w defines as 0.5, 0.7 and 0.4, respectively, Acc and DR indexes start to maintain stability. In general, when $\alpha_w \geq 0.3$ the Acc and DR indexes of the most datasets will be stable. The optimal scenario is to define $\alpha_w = 0.7$ at which point Acc and DR indexes of all datasets remain stable.

VI. CONCLUSION

This paper proposed a hybrid feature selection method FAFS_HFS based on feature subset and FA. It uses FA to generate feature subsets with strong internal correlation and weak external correlation, then uses FSmRMR based on feature subsets to carry out redundancy removal for each feature subset, and finally evaluates and selects subsets by FSF-score method. In this study, we have compared the classification performance of the proposed FAFS_HFS method with the traditional mRMR, F-score and the proposed MF and FSmRMR methods on 14 datasets. The experimental results show that the proposed FAFS_HFS method has the advantage of high OE index. Compared with IG, SU, GR, RlF, CBF, FCBF and MIM methods, FAFS_HFS also has advantages in classification performance and generalization ability, especially in high-dimensional datasets.

In the future, we will try to use different methods of redundancy removal and subset evaluation to explore the effectiveness of FAFS_HFS feature selection method, and other hybrid feature methods such as clustering and various optimization algorithms will be deeply studied and explored. Moreover, extend the proposed method to higher dimensional datasets, and further optimized the running time to achieve real-time function.

ACKNOWLEDGMENT

(Lizeng Gong and Shanshan Xie contributed equally to this work.)

REFERENCES

[1] T. Dokeroglu, A. Deniz, and H. E. Kiziloz, "A comprehensive survey on recent metaheuristics for feature selection," *Neurocomputing*, vol. 494, pp. 269–296, Jul. 2022, doi: 10.1016/j.neucom.2022.04.083.

[2] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Comput. Statist. Data Anal.*, vol. 143, Mar. 2020, Art. no. 106839, doi: 10.1016/j.csda.2019.106839.

[3] A. Kaur, K. Guleria, and N. K. Trivedi, "Feature selection in machine learning: Methods and comparison," in *Proc. Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, Mar. 2021, pp. 789–795.

[4] H. A. L. Thi, V. V. Nguyen, and S. Ouchani, "Gene selection for cancer classification using DCA," in *Proc. Int. Conf. Adv. Data Mining Appl.*, 2008, pp. 62–72.

[5] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997, doi: 10.1016/S0004-3702(97)00043-X.

[6] U. Kaya and M. Fidan, "Parametric and nonparametric correlation ranking based supervised feature selection methods for skin segmentation," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 2, pp. 821–833, Feb. 2022, doi: 10.1007/s12652-021-02936-0.

[7] M. Eftekhari, A. Mehrpooya, F. Saberi-Movahed and V. Torra "A hybrid filter-based feature selection method via hesitant fuzzy and rough sets concepts," M.S. thesis, Dept. Comp. Engn., Shahid Bahonar Univ. Kerman, Kerman, Iran, 2022, pp. 147–156, doi: 10.1007/978-3-030-94066-9_10.

[8] B. Khozaei and M. Eftekhari, "Unsupervised feature selection based on spectral clustering with maximum relevancy and minimum redundancy approach," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 35, no. 11, Sep. 2021, Art. no. 2150031, doi: 10.1142/S0218001421500312.

[9] J. Gonzalez-Lopez, S. Ventura, and A. Cano, "Distributed selection of continuous features in multilabel classification using mutual information," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2280–2293, Oct. 2019.

[10] J. Gonzalez-Lopez, S. Ventura, and A. Cano, "Distributed multi-label feature selection using individual mutual information measures," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 105052, doi: 10.1016/j.knosys.2019.105052.

[11] S. Xie, Y. Zhang, D. Lv, X. Chen, J. Lu, and J. Liu, "A new improved maximal relevance and minimal redundancy method based on feature subset," *J. Supercomput.*, pp. 1–24, Aug. 2022, doi: 10.1007/s11227-022-04763-2.

[12] D. Jain and V. Singh, "A two-phase hybrid approach using feature selection and adaptive SVM for chronic disease classification," *Int. J. Comput. Appl.*, vol. 43, no. 6, pp. 524–536, Jul. 2021, doi: 10.1080/1206212X.2019.1577534.

[13] S. Abasabadi, H. Nematzadeh, H. Motameni, and E. Akbari, "Hybrid feature selection based on SLI and genetic algorithm for microarray datasets," *J. Supercomput.*, vol. 78, no. 18, pp. 19725–19753, Dec. 2022, doi: 10.1007/s11227-022-04650-w.

[14] M. S. Uzer, O. Inan, and N. Yılmaz, "A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA," *Neural Comput. Appl.*, vol. 23, nos. 3–4, pp. 719–728, Sep. 2013, doi: 10.1007/s00521-012-0982-6.

[15] M. Alzaqebah, S. Jawarneh, R. M. A. Mohammad, M. K. Alsmadi, I. Al-marashdeh, E. A. E. Ahmed, N. Alrefai, and F. A. Alghamdi, "Hybrid feature selection method based on particle swarm optimization and adaptive local search method," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, pp. 2414–2422, 2021, doi: 10.11591/ijece.v11i3.pp2414-2422.

[16] R. M. Aziz, "Nature-inspired metaheuristics model for gene selection and classification of biomedical microarray data," *Med. Biol. Eng. Comput.*, vol. 60, no. 6, pp. 1627–1646, Jun. 2022, doi: 10.1007/s11517-022-02555-7.

[17] R. M. Aziz, "Application of nature inspired soft computing techniques for gene selection: A novel frame work for classification of cancer," *Soft Comput.*, vol. 26, no. 22, pp. 12179–12196, Nov. 2022, doi: 10.1007/s00500-022-07032-9.

[18] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021.

[19] A. Tiwari and A. Chaturvedi, "A hybrid feature selection approach based on information theory and dynamic butterfly optimization algorithm for data classification," *Expert Syst. Appl.*, vol. 196, Jun. 2022, Art. no. 116621, doi: 10.1016/j.eswa.2022.116621.

[20] J. Xie and W.-X. Xie, "Several feature selection algorithms based on the discernibility of a feature subset and support vector machines," *Chin. J. Comput.*, vol. 37, pp. 1704–1718, Aug. 2014.

[21] M. Li, H. Wang, L. Yang, Y. Liang, Z. Shang, and H. Wan, "Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113277, doi: 10.1016/j.eswa.2020.113277.

[22] Q. Song, J. Ni, and G. Wang, "A fast clustering-based feature subset selection algorithm for high-dimensional data," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 1–14, Jan. 2013.

[23] Z. Dehghan and E. G. Mansoori, "A new feature subset selection using bottom-up clustering," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 57–66, Feb. 2018, doi: 10.1007/s10044-016-0565-8.

[24] M. García-Torres, F. Gómez-Vela, F. Divina, D. P. Pinto-Roa, J. L. V. Noguera, and J. C. M. Román, "Scatter search for high-dimensional feature selection using feature grouping," in *Proc. Genetic Evol. Comput. Conf. Companion*, Jul. 2021, pp. 149–150, doi: 10.1145/3449726.3459481.

[25] M. R. Mahmoudi, D. Baleanu, S. S. Band, and A. Mosavi, "Factor analysis approach to classify COVID-19 datasets in several regions," *Results Phys.*, vol. 25, Jun. 2021, Art. no. 104071, doi: 10.1016/j.rinp.2021.104071.

[26] M. S. Bartlett, "Multivariate analysis," *Suppl. J. Roy. Stat. Soc.*, vol. 9, no. 2, pp. 176–197, 1947, doi: 10.2307/2984113.

[27] J. Decoster. (1998). *Overview of Factor Analysis*. Accessed: May 31, 2022. [Online]. Available: http://www.stat-help.com/notes.html

[28] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[29] M. Billah and S. Waheed, "Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection," *Multimedia Tools Appl.*, vol. 79, nos. 33–34, pp. 23633–23643, Sep. 2020, doi: 10.1007/s11042-020-09151-7.

[30] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," 2012, *arXiv:1202.3725*.

[31] Z. Q. Li, J. Q. Du, B. Nie, W. P. Xiong, C. Y. Huang, and H. Li, "Summary of feature selection methods," *Comput. Eng. Appl.*, vol. 55, no. 24, pp. 10–19, 2019, doi: 10.3778/j.issn.1002-8331.1909-0066.

[32] N. Sevani, I. Hermawan, and W. Jatmiko, "Feature selection based on F-score for enhancing CTG data classification," in *Proc. IEEE Int. Conf. Cybern. Comput. Intell. (CyberneticsCom)*, Aug. 2019, pp. 18–22.

[33] M. Sasikala and N. Kumaravel, "Comparison of feature selection techniques for detection of malignant tumor in brain images," in *Proc. Annu. IEEE India Conf. (Indicon)*, Jun. 2005, pp. 212–215.

[34] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994, doi: 10.1016/0167-8655(94)90127-9.

[35] *UC Irvine Machine Learning Repository*. Accessed: Mar. 20, 2022. [Online]. Available: https://archive-beta.ics.uci.edu/ml/datasets

[36] *Grupo De Inteligencia Computacional*. Accessed: Mar. 21, 2022. [Online]. Available: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

LIZENG GONG received the degree in communication engineering from Southwest Forestry University, Kunming, China, in 2019. His current research interests include machine learning and intelligent information processing.

SHANSHAN XIE received the B.E. degree from Southwest Forestry University, Kunming, China, in 2020, where she is currently pursuing the M.A. degree in systems science. She has published five academic papers. Her current research interests include machine learning and intelligent information processing.

YAN ZHANG received the Ph.D. degree from Beijing Forestry University, Beijing, China. She is currently a Professor with the College of Mathematics and Physics, Southwest Forestry University. She is the author of five books and more than 50 articles. Her research interests include machine learning, speech recognition, and intelligent information systems.

MENGYAO WANG received the M.A. degree from Yunnan University, Kunming, China, in 2016. She is currently working at the College of Mathematics and Physics, Southwest Forestry University. She has more than five patents. Her current research interests include artificial intelligence and blockchain.

XIAOYAN WANG received the Graduate degree from the Taiyuan University of Technology, Taiyuan, China. She is currently working at the College of Mathematics and Physics, Southwest Forestry University. She has published more than 20 academic papers. Her current research interests include mobile communications and machine learning.