

Received August 31, 2018, accepted September 27, 2018, date of publication October 5, 2018, date of current version November 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2873634

A New Filter Feature Selection Based on Criteria Fusion for Gene Microarray Data

WENJUN KE¹, CHUNXUE WU¹, YAN WU², AND NEAL N. XIONG¹

¹School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

²School of Public and Environmental Affairs, Indiana University Bloomington, Bloomington, IN 47405, USA

³Department of Mathematics and Computer Science, Northeastern State University Tahlequah, Tahlequah, OK 74464, USA

Corresponding author: Neal N. Xiong (xiong31@nsuok.edu)

This work was supported in part by the Shanghai Science and Technology Innovation Action Plan Project under Grant 16111107502 and Grant 17511107203, in part by the National Natural Science Foundation of China under Grant 61502220, and in part by Henan Science and Technology through the Program for Tackling Key Problems under Grant 172102310636.

ABSTRACT In machine learning and data mining, feature selection aims to seek a compact and discriminant feature subset from the original feature space. It is usually used as a preprocessing step to improve the prediction performance, understandability, scalability, and generalization capability of classifiers. A typical gene microarray data set has the characteristics of high dimensionality, limited samples, and most irrelevant features, and these characteristics make it difficult to discover a compact set of features that really contribute to the response of the model. In this paper, a score-based criteria fusion feature selection method (SCF) is proposed for cancer prediction, and this method aims at improving the prediction performance of the classification model. The SCF method is evaluated on five open gene microarray data sets and three low-dimensional data sets, and it shows superior performance over many well-known feature selection methods when employing two classifiers SVM and KNN to measure the quality of selected features. Experiments verify that SCF is able to find more discriminative features than the competing methods and can be used as a preprocessing algorithm to combine with other methods effectively.

INDEX TERMS Dimension reduction, feature selection, high-dimensional data, criteria fusion, cancer prediction.

I. INTRODUCTION

In the information area, enormous amounts of high-dimensional data have been continuously produced in assorted areas, e.g. social medias, genomics, biomedical engineering, e-commerce data and surveillance video. The Big data analysis can discover meaningful information from these massive data and promise new levels of scientific discovery and economic value [1]. It has two main goals [2]: constructing robust methods that can accurately predict query observations, and understanding the relationship between the features and response for scientific purpose. In general, the big data analysis aims to understand heterogeneity and commonality across different subpopulations. In recent years, healthcare has received increasing attention, and has benefited from big data analysis in diagnosing disease, planning treatment, and determining the effectiveness of a treatment based on information about tumors.

Cancer classification is key part of healthcare. Conventional diagnostic method is conducted based on morphological and clinical results, and it is limited by the professional

competence of the doctor. Benefiting from DNA microarray technology, researchers can measure expression levels of thousands genes through the gene expression microarray (GEM) experiment. It has been reported that gene expression represents the process of using genetic information to produce functional gene products, such as proteins and functional RNA. Tumors are primarily caused by the changes in the genetic information, and the gene expression levels of tumor's tissue differ from those of the normal tissues. Therefore, gene expression levels can be utilized to determine whether a person has cancer or not and to categorize different tumor types. Compared to conventional cancer diagnostics techniques, some classical classification methods (such as SVM, KNN, Random Forest and so on) from data mining and machine learning can attain impressive results on GEM data.

However, the tough challenges come from the characteristics of GEM data (i.e., high dimensionality, severely limited samples and containing large portion of irrelevant genes), resulting in difficulties of discovering key genes for cancer prediction [3]. An important research objective in GEM data

analysis is to discover genes relevant for particular target from different populations, and these genes are called informative genes or biomarkers [4]. In addition, the data characteristics of high dimensionality and limited samples could render classification methods prone to be overfitting [5]. Given that high-dimensional data often reside into subspaces of lower dimension [6], [7], their crucial information can be presented in several subspaces. Hence, for GEM data, a high-dimensional data analysis (such as dimension reduction) can be used to dig out informative genes. The dimension reduction aims to represent high-dimensional data with fewer features while preserving most of the key information within the original data [8], [9]. In order to alleviate the overfitting of classifiers for GEM data, feature selection mechanism is employed as a preprocessing algorithm for classification methods. It is one of the main dimension techniques and can reduce the dimensionality via selection of a small subset of the original feature set, which helps alleviate the dimension disaster [10], reduces the time complexity of classification algorithm and improves the prediction accuracy.

Gene selection is the application of feature selection technology in gene microarray data [11]. Among the original genes, the informative genes associated with cancer are rather few [12]. Through gene selection informative genes could be efficiently identified, which improves the prediction accuracy of classification models and removes the noise, irrelevant and redundant genes interfering with the performance of the model [13]. Gene selection has two main purposes [14], [15]: (1) to identify informative genes associated with a cancer. (2) to spot a compact gene set that has discriminative power to construct a more robust pattern classifier for generalization. However, due to the relatively higher dimensionality and limited sample size of GEM data [11], the feature selection could be not effectively applied in discovering informative genes.

Despite recent literature of feature selection [5], [9], [16]–[19] is abundant, our purpose is to propose a new filter feature selection for cancer classification based on GEM data. The remainder of this paper is as follows. In section II, different types of feature selection methods and some related studies of filter methods are introduced. In addition, a summary of our contributions is made. Then, in section III, two feature selection criteria are described in detail, and a new feature selection method is proposed. Next, experiments reveal the efficiency of the proposed method and the experimental results are demonstrated in section IV. Section V summarizes this work and discusses the future works.

II. RELATED WORK

In the feature selection, according to the relevance and redundancy of features, feature subsets are mainly divided into three categories: noisy and irrelevant features, relevant features, and relevant and non-redundant features. To achieve a satisfactory model prediction, relevant features should be selected, while irrelevant, redundant, or noisy features should

be eliminated due to their worthless to improve prediction accuracy. Specifically, irrelevant and noisy features contain valueless information for prediction, and redundant features have significantly statistical relations with other features [20].

Generally speaking, feature selection is the process of discovering little informative features associated with a specific task and simultaneously discarding noise, irrelevant and redundant features [8]. It involves a multitude of advances, including reducing the high dimensionality of data and time cost in training classifier, improving the prediction accuracy of classifier, and enhancing the understandability and generalization capability of the classifier.

In terms of the principle of the methods, feature selection methods can be divided into: filter [4], [21], wrapper [22], [23], embedded [24], hybrid [18], [25] and ensemble [26]. A filter method does not involve a classification model and only employs the intrinsic properties of data (such as dependency, mutual information, distance, and consistency) to measure the quality of a feature subset [4], [22], which make filter methods have lower time complexity. It can be flexibly used in other types of feature selection. A hybrid method is a combination of two different methods such as filter and wrapper, where filter feature selection is used as a preprocess algorithm to remove the noise genes and reduce the data dimensionality [25], [27], [28]. A ensemble method utilizes filters as criteria to produce a set of feature subsets by using various strategies, and then generates an aggregated result from these subsets [26]. Compared with filter methods, wrapper methods involve a specific classifier used to evaluate the quality of the feature subset [20]. They achieve better classification performance since they search for a subset of features that are best suited to the classifier [22]. An Embedded method is similar to wrapper, and also involves a specific classifier. However, its selection of feature subsets is embedded in the classifier [24] and its feature selection performs together with training of classifier. The complexity of classification computation is vital for the feature selection [5], but either wrapper or embedded method has a high time complexity, which makes them difficult to be applied to other kinds of data, e.g., low dimensional data. Compared with other types of feature selection methods, filter method has no advances in prediction accuracy, but it is fundamental considering that it is frequently included in hybrid and ensemble methods and compatible with wrapper or embedded methods. Filter method has advances of flexibility and low time complexity, making them able to be effectively combined with other algorithms. Moreover, filter method has been broadly used. They could be applied to low dimensional data, rather than just to high dimensional data. Therefore, the filter feature selection is investigated herein.

As the relation between variables can be explicitly expressed using mutual information according to information theory, many filter methods [29]–[31] based on mutual information have been proposed. According to the selected features, there are mainly two categories of feature selection methods [14], [15]: ranking-based feature selection and

set-based feature selection. Ranking-based feature selection ranks features according to their individual importance, but it does not take into account the dependencies between features [15]. Set-based feature selection evaluates the importance of features and at the meantime considers the dependencies between features [11]. To date, a large number of literatures on ranking-based feature selection methods have been proposed, such as gini index [32], information gain [33], Chi-square test [34], Fisher score [35], Laplacian score [36] and ReliefF [37], [38]. The main advantages of these methods is the low time complexity, which exhibits a linear relationship with the data dimensionality. Therefore, these methods are usually used as a pre-selection algorithm to remove irrelevant and noisy features in the data.

Recent work has shown that set-based methods significantly outperform ranking-based methods in feature selection [39], [40]. When selecting a feature subset, a set-based method typically optimizes a cost function to ensure maximum relevance between features and class labels and to avoid redundancy of the feature subset. There are some famous set-based methods based on mutual information, such as MIFS [41], MIFS-U [42], mRMR [40], DISR [16], JMIM [43], NMIFS [44] and FBCF [13]. Battiti *et al.* [41] established a MIFS method, which considers the relevance between features and classes as well as the redundancy between features in the cost function. However, the coefficient of the redundancy term is difficult to determine. Kwak and Choi [42] proposed MIFS-U based on MIFS, estimating the relevance by using normalized mutual information instead of mutual information. Normalized mutual information corrects the biased of mutual information towards features with more values, so it makes a better estimation for the relevance than MIFS. Peng *et al.* [40] proposed a variant of MIFS, MRMR, where the reciprocal of the number of the selected feature is defined as the coefficient of the redundant term. NMIFS suggested by Estévez *et al.*, aims to obtain a better estimation for redundancy between features by using normalized mutual information [44]. Yu and Liu [13], [39] built a fast correlation-based filter (FCBF) method that circularly removes redundant features and retains predominant features through Markov blanket. However, the number of genes selected by the FCBF is uncontrollable due to fcbf removes features through a pre-set threshold. Even if the threshold is low, the number of genes it chooses is sometimes very small. Clara *et al.* proposed a feature selection method called Double Input Symmetrical Relevance (DISR). This method is the normalized second-order approximation of the criterion MASSIVE that maximizes the lower bound of the mutual information of a subset [16]. Additionally, Bennasar *et al.* [43] propose a feature selection method called Joint Mutual Information Maximisation (JMIM), which is the goal function based on joint mutual information and the ‘maximum of the minimum’ approach.

In this paper, our purpose is to construct a more versatile filter feature selection framework for cancer prediction

through fusing various feature selection methods. The proposed method developed in this study fuses two completely different feature selection methods to estimate the relevance between features and the class label, and it evaluates the redundancy in the selected features. Briefly, the key idea of the proposed method is to ensure the features that are highly related to the class and independent of each other. Compared with other filter methods, the proposed method can spot a subset of genes reflecting more cancer information from the dataset.

In summary, our contributions are as following:

- A novel filter feature selection method, which fuses SU and ReliefF to assess the importance of the genes for classification and adopt a redundant measurement to confirm the dependence of the genes, is proposed. In addition, its variants of the proposed method are presented.
- The proposed method could improve the performance of the classification model, and we validate its superiority with five GEM datasets and three low dimensional UCI datasets.
- In addition, experimental results verify that the proposed method can be effectively combined with a wrapper method.

III. FEATURE SELECTION BASED ON FUSION OF CRITERIA

A. MOTIVATION

In machine learning, several weak classifiers can be integrated into a complex composite classifier that could produce better results than a strong classifier [45]. Inspired by this idea, a new score-based criteria fusion feature selection method (SCF) is proposed. This method combines two feature selection methods, i.e., SU and ReliefF. The idea of criteria fusion is useful because [11]: Firstly, feature subsets produced by different criteria are diverse. However, no consensus has been reached on which criterion is best and no guideline is available for determining which method is most suitable for practical applications, while the fusion of multiple criteria is of highly practical applicability. Secondly, the model based on multiple weak assumptions is usually more effective than one based on a strict assumption. Each criterion is based on an assumption about data distribution, but the actual data distribution is usually complex and unknown. The fusion of multiple criteria could at least partially weaken the precondition and capture results that better reflect the actual data distribution. Thirdly, the search ability of each feature criterion is limited, which may overlook some informative features. The fusion of multiple criteria is able to produce a complementary result, conducive to the selection of informative features.

Consequently, a new filter method based on criterion fusion is proposed for cancer classification. In contrast with the existed filter methods, the proposed method has three improvements and superiority. Firstly, some filter methods do not consider the redundancy between the selected genes, while our method adopts the redundant measurement to

minimize the redundancy, through which only the most informative features are selected. Secondly, many filter methods are based on mutual information, but mutual information bears a bias toward genes that have large values. SCF employs the variant of the mutual information to avoid the bias, rendering the evaluation of relevance more convincing. Thirdly, the estimation of relevance by SCF relies on two criteria rather than a single one, which ensures that its outcome is more robust and more applicable across different datasets.

B. BASIC CRITERIA

There have been many literature on feature ranking methods [4], [20], and feature selection methods chosen for fusion are called basis criteria. To select appropriate basis criteria for fusion, a appropriate benchmark is needed to be followed. Benchmark in [11]: basis criteria should present diversity, is adopted here. Hence, feature subsets generated by basis criteria should be completely dissimilar, and the fusion of different feature selection criteria can assure a complementary result. Provided that two feature selection criteria produce similar subsets, their fusion is useless.

In this paper, feature ranking methods SU and ReliefF are chosen as basis criteria for relevance estimation because: 1) the diversity should be followed. SU and ReliefF are completely different relevance evaluation criteria. SU estimates the relevance based on mutual information, while ReliefF is related with distance estimation and it estimates the relevance according to the ability of distinguishing near-distance samples; 2) many literatures have verified the effectiveness of mutual information, but mutual information has a bias towards variable with large values, and thus SU that normalizes mutual information is adopted; 3) ReliefF is a simple and efficient algorithm to evaluate the ability of attributes, and it outperforms Maxrel, Sum Minority and partially wins information gain [38]; 4) both SU and ReliefF are time efficient, which keeps our method's time complexity from reaching too high. The details of the SU and ReliefF are described below.

1) SYMMETRICAL UNCERTAINTY(SU)

In information theory, mutual information refers to the amount of information about another random variable contained in one random variable, which is intuitively understood as the amount of information shared by two variables [24]. It is usually often used to measure the relevance between variables. The definition of mutual information is as:

$$I(X, Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (1)$$

where $p(x, y)$ is the joint probability distribution function for X and Y , $p(x)$ and $p(y)$ are the edge probability distribution function of X and Y respectively. The mutual information can also be following formula::

$$\begin{aligned} I(X, Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y), \end{aligned} \quad (2)$$

where $H(X)$ and $H(Y)$ denote edge entropy, $H(X|Y)$ is conditional entropy, and $H(X, Y)$ is joint entropy. From equation (2), the value range of mutual information is as follows:

$$0 \leq I(f_i, f_s) \leq \min\{H(f_i), H(f_s)\}. \quad (3)$$

However, there is a big problem with mutual information, which has a bias towards features with more values [13]. To avoid this problem, SU is used here to compensate for the bias of mutual information. The definition of SU is as follows:

$$SU(X, Y) = \frac{2I(X, Y)}{H(X) + H(Y)}, \quad (4)$$

where the value of SU belongs to the interval $[0, 1]$ and it is symmetric. When the value of SU is 1, it means that the knowledge of one random variable X can completely predict the knowledge of another variable Y , and its value of 0 means that X and Y are independent.

2) RELIEFF

ReliefF is an attribute estimation algorithm, which is an extended version of Relief. Original relief is limited to only two-class problems, while ReliefF is extended to multiple-class problems [37]. The main idea of ReliefF is based on the feature's ability of distinguishing near-distance instances to estimate the weight of features, it evaluates the weight of features through multiple iterations. At each iteration, an instance R is randomly selected from a dataset, and ReliefF searches k nearest neighbors (called nearest hit, H) from the class of instance and searches k nearest neighbors (called nearest miss, M) from each of the other class [37]. For a feature, if the distance between instance R and the nearest hit is less than that between sample R and nearest miss M , this feature is useful for distinguishing instance from different classes and its weight should be elevated; otherwise, it is useless and its weight should be reduced. For a dataset with n features, the above update process needs to be repeated n times to get the weight estimation W of all final features. The formula for updating W is as follows:

$$\begin{aligned} W[i] &= W[i] - \sum_{j=1}^k \frac{\text{diff}(i, R, H_j)}{m \cdot k} \\ &\quad + \sum_{C \neq \text{class}(R)} \frac{P(C)}{1 - P(\text{class}(R))} \cdot \frac{\sum_{j=1}^k \text{diff}(i, R, M_j(C))}{mk}, \end{aligned} \quad (5)$$

$$\begin{aligned} \text{diff}(i, R, H) &= \frac{|value(i, I_1) - value(i, I_2)|}{\max(i) - \min(i)}, \end{aligned} \quad (6)$$

where function $\text{diff}(i, R, H_j)$ represents the difference of feature i between sample R and H_j . $P(C)$ is the prior probability of the class C and $1 - P(\text{class}(R))$ is the sum probability for the misses' class $\text{class}(R)$.

C. SCORE-BASED CRITERIA FUSION(SCF)

Being mindful of the aforementioned challenges, we propose a new feature selection method, score-based criteria fusion (SCF), which maximizes the relevance between features and class and simultaneously emphasizes dependence between genes. SCF is consisted of two estimation measures. One is a relevance estimate for selecting features that are highly associated with class labels, and the other is a redundancy estimate used to select features that are non-redundant to other selected features.

To select informative genes for cancer classification, the first step is to locate genes that are highly associated with the category, and then remove the redundant genes from these candidate genes. Therefore, good relevance estimation measure is essential for the selection of informative genes. In SCF, the relevance estimation measure consists of two parts, i.e., SU and ReliefF, combined by using a specific fusion method. According to the combined objects, the fusion methods are divided into two categories: score-based multicriterion fusion and ranking-based multicriterion fusion [11]. In this study, score-based multicriterion is employed to fuse two basis criteria. Via score-based methods, each basis criteria generates a score vector containing scores of all features, and multiple score vectors are aggregated into one vector by a combination algorithm [15]. Finally features are sorted according to their values in the final score vector. Concretely, combination algorithm fuses two score vectors from two basis criteria by multiplying the weight parameter. The detail of combination is as follows:

$$R_{i,c} = \mu S_{U,i,c} + (1 - \mu) W_i, \quad (7)$$

where the weighted parameter μ makes a trade-off between SU and ReliefF, and its value determines the contributions of two basis criteria to the estimated relevance. The value of μ belongs to $[0, 1]$.

It is worth noting that score produced by two basis criteria should be comparable, suggesting that the range of values for both criteria should be the same. In this paper, score generated by SU belongs to the interval $[0, 1]$, while the interval of scores generated by ReliefF is unsure. Hence, score normalization is performed for ReliefF before the score combination. The score normalization is as follows:

$$u'_i = \frac{u_i - u_{min}}{u_{max} - u_{min}}. \quad (8)$$

A set of good features should be highly relevant to classes and non-redundant with each other. In this work, the estimation measure of redundancy mentioned in [44] is adopted to evaluate the redundancy in the selected feature subset. This measure normalizes mutual information to eliminate the bias of mutual information towards features with large values. The detail of this measure is as follows:

$$NI(f_i, f_s) = \frac{I(f_i, f_s)}{\min\{H(f_i), H(f_s)\}}. \quad (9)$$

$$D_i = \frac{1}{|S|} \sum_{f_i, f_s \in S} NI(f_i, f_s). \quad (10)$$

Algorithm 1 SCF

```

input: train data  $D(f_1, f_2, \dots, f_N)$ , predefined threshold  $\delta$ ,
        the number of selected  $K$ ;
output: selected optimal subset  $S_{best}$ ;
1: begin
2:   for  $i = 1 \rightarrow n$  do
3:     calculate  $R_i$  for  $F_i$  by (6);
4:     if  $R_i \geq \delta$  then
5:       Add  $R_i$  to  $S_{list}$ ;
6:     end if
7:   end for
8:   Order  $S_{list}$  in descending the value of  $R$ ;
9:    $F_p = getFirstElement(S_{list})$ ;
10:  Add  $F_p$  to  $S_{best}$ ;
11:  Remove  $F_p$  from  $S_{list}$ ;
12:  while  $|S_{best}| \leq K$  do
13:     $F_{curlast} = getLastElement(S_{best})$ ;
14:    for  $i = 1 \rightarrow |S_{list}|$  do
15:      calculate  $D_i(F_i, F_{curlast})$  by (9);
16:    end for
17:     $F = \max_{F \in S_{list}}(G)$ ,  $G$  is calculated by (10);
18:    Add  $F_q$  to  $S_{best}$ ;
19:    Remove  $F_q$  from  $S_{list}$ ;
20:  end while
21:  return  $S_{best}$ 
22: end
```

where $|S|$ is the number of selected features, and its reciprocal is used to balance relevance term with redundant term.

The above estimation measures of relevance and redundancy construct a new set-based algorithm SCF. The complete definition is as follows:

$$G = R_{i,c} - D_i. \quad (11)$$

Finally, a relevant and non-redundant feature subset can be selected by maximizing cost function (11). Note that when $\mu = 0$ and $\mu = 1$, (11) respectively becomes (12) and (13). (12) and (13) is the variant of SCF, and they can be treated as adding redundancy estimates to SU and ReliefF.

$$G_1 = W_i - D_i. \quad (12)$$

$$G_2 = S_{U,i,c} - D_i. \quad (13)$$

In addition, the search strategy of SCF algorithm follows the incremental forward selection. For details of the algorithm, it can be seen algorithm 1. Given a dataset with N features and a class variable C , and S_{best} denote predominant features selected by the algorithm SCF for cancer classification. Line 2-7 represents a calculation for the relevance between features and classes, and S_{list} is the features list after removal noise and irrelevant features by predefined threshold δ . In line 9-20, algorithm calculates the score of each feature by subtracting redundancy from relevance, and then adding the feature with highest score into S_{best} . In line 21, algorithm stops when K features are selected, and return S_{best} .

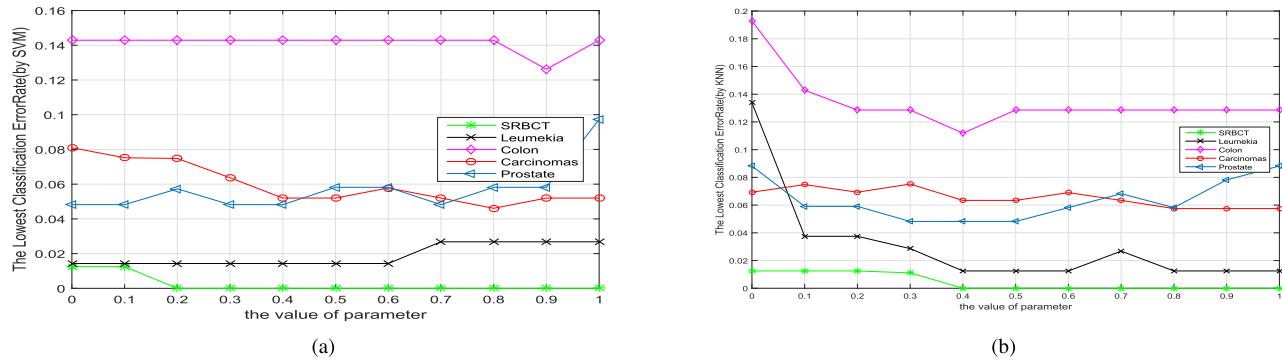


FIGURE 1. Performance comparisons of different value of control parameter. (a) Classification error by SVM. (b) Classification error by KNN.

IV. PERFORMANCE ANALYSIS

A. DATASET

In order to validate the algorithm SCF, experiments are conducted on the following five gene expression microarray datasets:

SRBCT data [46]. There are four types of childhood tumor and 83 samples, including 29 EWS samples, 18 NB samples, 11 BL samples and 25 RMS samples. Each sample contains 2308 genes. This dataset can be downloaded at <http://www.biolab.si/supp/bi-cancer/projections/info/SRBCT.html>.

Leukemia data [47]. This dataset is a collection of expression measurements reported by Golub *et al.* This microarray dataset contains 72 samples, including 25 acute myeloid leukemia (AML) samples and 47 acute lymphoblastic leukemia (ALL) samples. Each sample contains 7129 genes. This dataset can be downloaded at <http://cilab.ujn.edu.cn/datasets.html>.

Colon data [48]. This dataset is a collection of expression measurements of colon biopsy samples reported by Alon *et al.* This microarray dataset contains 62 colonic epithelial cell samples, including 22 normal samples and 40 tumor samples. Each sample contains 2000 genes. This dataset can be downloaded at <http://featureselection.asu.edu/datasets.php>.

Carcinomas data [49]. This dataset is consisting of 174 samples, including 11 categories, 26 prostate, 8 bladder/ureter, 26 breast, 23 colorectal, 12 gastroesophagus, 11 kidney, 7 liver, 27 Ovary, 6 pancreas, 14 lung adenocarcinomas, and 14 lung squamous cell carcinomas. Each sample contains 12533 genes in the original dataset, and 9182 genes per samples in preprocessed dataset. This dataset can be downloaded at <http://featureselection.asu.edu/datasets.php>.

Prostate data [17]. This dataset was first reported by Singth *et al.* It is consisted of 102 samples where 50 samples is prostate tumors and 52 sample is normal. Each samples contains 10509 genes. This dataset can be downloaded at <http://www.gems-system.org/>.

Table 1 briefly summarizes these datasets.

B. EXPERIMENT SETUP

In the experimental study, proposed algorithm, SCF, firstly is used to select a set of genes from each gene dataset, then

TABLE 1. Gene data sets characteristics.

data set	samples	genes	class
SRBCT	83	2308	4
Leukemia	72	7129	2
colon	62	2000	2
Carcinomas	174	9182	11
Prostate	102	10509	2

the selected genes are evaluated by two classifiers, SVM and KNN ($k = 3$). Additionally, the performance of SCF is compared with ReliefF, SU, FCBF, MRMR, DISR, JMIM and NMIFS. The linear SVM is implemented through the LIBSVM toolbox. Due to the relatively small sample size of the datasets, the experimental results herein are all obtained via 10-fold cross validation to ensure the robustness of the classification. The average of 10 performance estimations (i.e., classification errors) is taken as the final result. These methods are implemented with Matlab and experiments are conducted on a machine with Intel Core CPU (2.5 GHz and 4G RAM).

In SCF, the quality of the relevance estimation is determined by the weight parameter μ and basis criteria. μ controls the contributions made by basis criteria, and thus its value decides the quality of the relevance estimation between features and classes, and affects the importance of features. In this work, the weight parameter μ is determined via 10-fold cross validation and experimental results on five GEM data are shown in the Fig. 1. The results show that the value of μ determines the optimal classification rate error of classifiers and the value corresponding to the optimal error varies across the datasets. Note that in general low classification error can be obtained when μ has a value of 0.4. Therefore μ is set to be 0.4 in the following experiments. Another point to note is that 0.4 is not the optimal value for SCF but the suggested value for SCF on five GEM datasets, and its suggested value may be not 0.4 on other GEM datasets because the optimal value is different on different datasets and is hard to determined.

The experiment is expected to demonstrate that the elevated efficiency of SCF over the baselines. Given two features sets S_1 and S_2 with the same number of features, it usually thinks that feature space of S_1 is more characteristic than

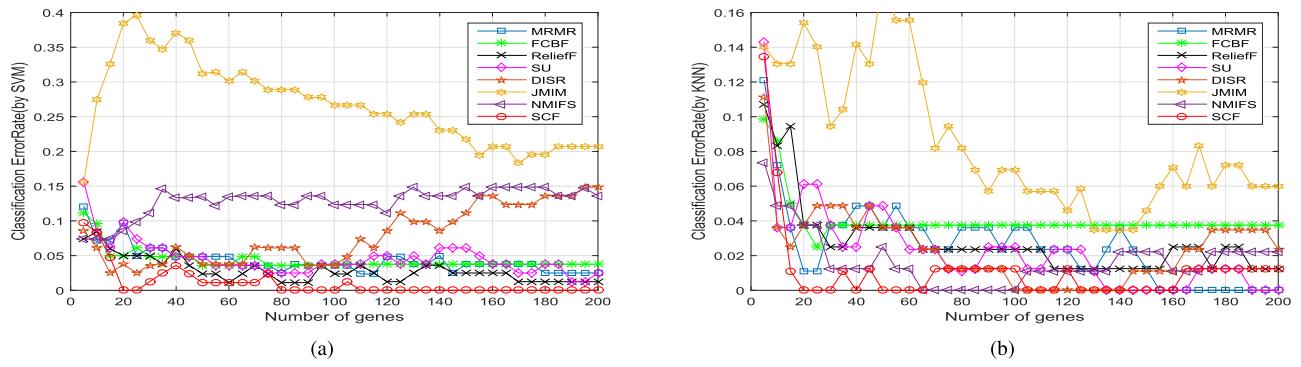


FIGURE 2. Performance comparisons on SRBCT data. (a) Classification error by SVM. (b) Classification error by KNN.

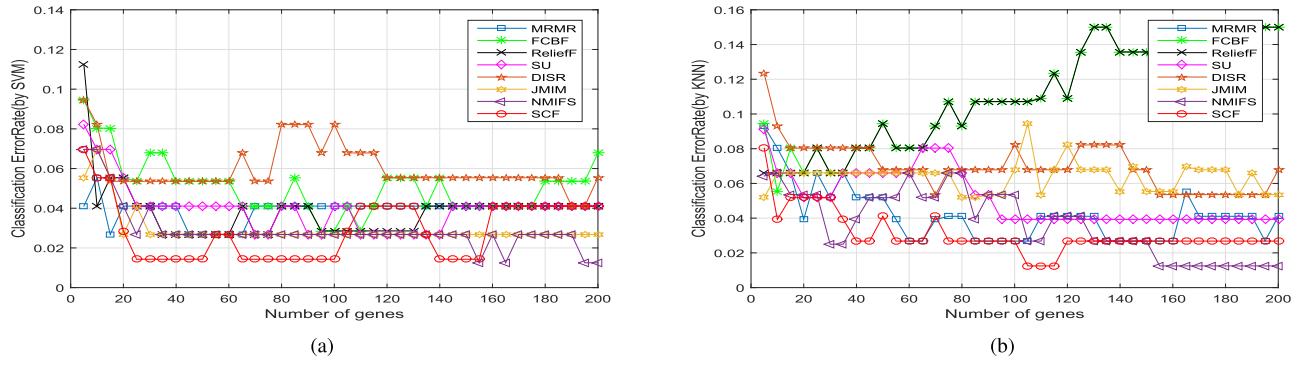


FIGURE 3. Performance comparisons on Leukemia data. (a) Classification error by SVM. (b) Classification error by KNN.

feature space of S_2 if the classification error based on S_1 is lower than that from S_2 . Next, the experiment will prove that SCF can obtain higher prediction accuracy than the baselines when dealing with the same number of sequential genes and nonsequential genes.

C. COMPARATIVE STUDY ON HIGH DIMENSIONAL GENE DATASETS

1) COMPARISON OF CLASSIFICATION ACCURACY

In order to verify that SCF can improve the prediction accuracy, the classification error of the genes selected by SCF is compared with the classification error of the genes selected by seven baselines. In this experiment, the number of genes selected ranges from 1 to 200. As can be seen from Fig. 2,3,4,5 and 6, in terms of classification error rate, SCF is basically superior to the baselines on five GEM datasets.

Fig. 2 shows the results from SVM and KNN for the SRBCT dataset. From the SVM classification error (Fig. 2a), SCF achieves an error rate of 0 with only 20 genes, which outperforms the baselines. DISR and JMIM perform unsatisfactorily, failing to spot the informative genes. Although other methods manage to reduce classification error, they do not capture the key genes from the dataset. However, SCF captures the key genes that reduce the classification errors to zero. According to the classification error from KNN, SCF, SU, MRMRR, NMIFS and DISR all achieves an error of 0. Nevertheless, SCF produces the lowest error within the selected subset of 20 genes.

Fig. 3 show the results for the SVM and KNN classification errors on Leukemia dataset. From the Fig. 3a, SCF and NMIFS both have an error of 1.43 percent, lower than other methods. However, SCF achieves this accuracy by selecting approximately 25 genes while NMIFS requires approximately 155 genes. The results of KNN are alike to those from SVM.

On colon dataset, according to the classification error results of SVM (Fig. 4a), SCF performs well, but inferior to NMIFS that produces an error down to 12.62 percent. A minimum error of 14.29 percent is observed for SCF, SU and ReliefF, while mRMR, DISR and JMIM are able to reach errors of 14.52, 16.19 and 17.86 percent, respectively. In contrast with SU and ReliefF, SCF achieves the lowest error with only 5 genes (Fig. 4a). According to the results from KNN (Fig. 4b), SCF achieves an error of 11.19 percent by screening only 15 genes, exhibiting great superiority over the baselines.

On Carcinomas dataset, SCF apparently outperforms the baselines in terms of the classification error results from SVM and KNN (Fig. 5). From the experimental results of SVM on Prostate dataset (Fig. 6a), SCF shows higher accuracy than MRMRR, FCBF, SU, ReliefF, and NMIFS. It achieves an error of 4.82 percent, slight inferior to DISR and JMIM that respectively produce error rate of 3.82 and 4.73 percent. However, SCF is obviously superior over the other methods according to the results from KNN on Prostate dataset (Fig. 6b).

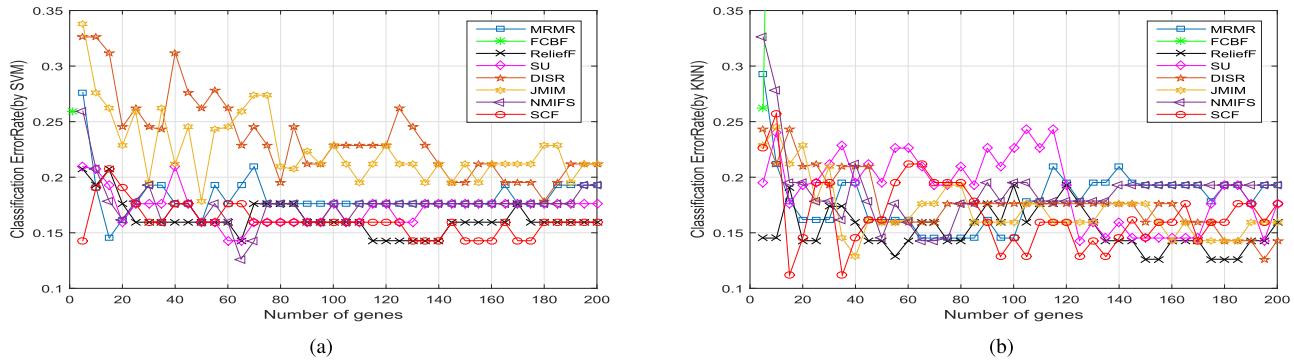


FIGURE 4. Performance comparisons on colon data. (a) Classification error by SVM. (b) Classification error by KNN.

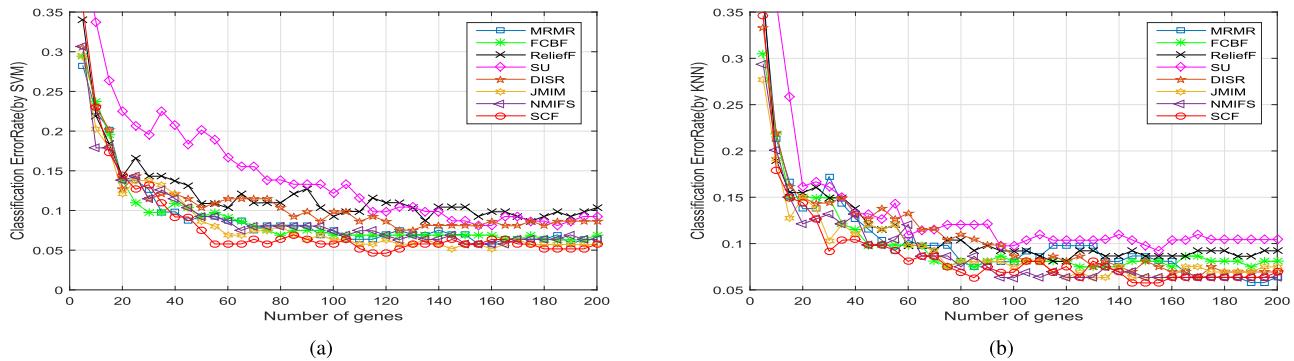


FIGURE 5. Performance comparisons on Carcinomas data. (a) Classification error by SVM. (b) Classification error by KNN.

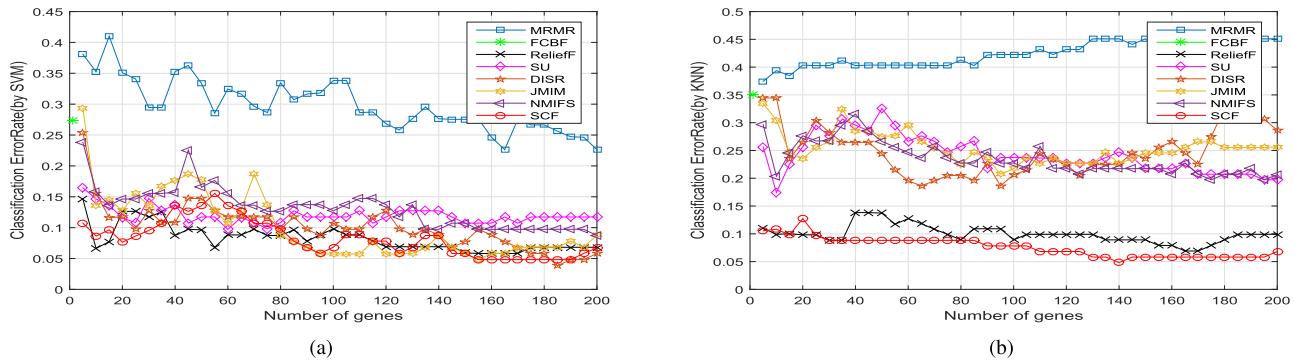


FIGURE 6. Performance comparisons on Prostate data. (a) Classification error by SVM. (b) Classification error by KNN.

In general, SCF exhibits better prediction accuracy than the baselines, and therefore the sequential genes selected by SCF are of higher relevance to the target class, i.e., cancer in this study.

2) COMPARISON OF COMPACT GENE SUBSETS

The experiment in the previous subsection has confirmed that SCF gene space has higher characteristic strength than the baselines when handling the same number of sequential genes. In this subsection, given the same number candidate genes, we will study whether the SCF feature space contains a more characterizing nonsequential feature subspace than the baselines. And the wrapper method

is used for selecting nonsequence genes from candidate genes.

First, 200 sequence genes are selected by SCF and the baseline as candidate gene set, and then use the wrapper based on forward selection strategy is employed to search for an optimal subset from these candidate gene sets. If candidate gene space of SCF is more characteristics than those of competing methods, wrapper could find a subset from SCF genes warranting a better classification.

According to the experimental results from SVM (Fig. 7), only few informative genes, which reduce the classification error, are selected by wrapper from candidate gene set.

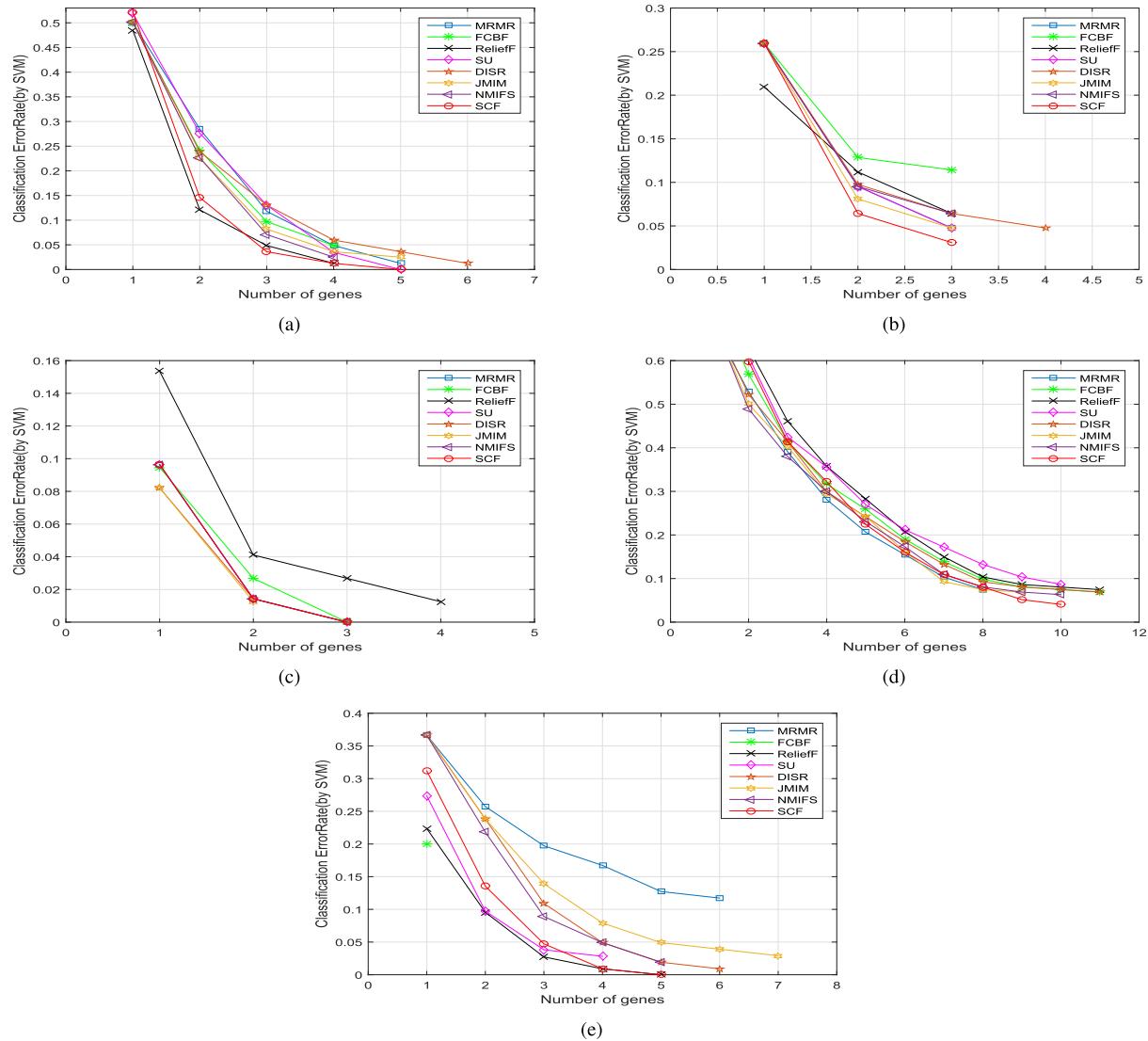


FIGURE 7. The wrapper classification results by SVM. (a) SRBCT. (b) colon. (c) Leukemia. (d) Carcinomas (e) Prostate.

TABLE 2. Low dimensional data sets characteristics.

data set	samples	features	class
Sonar	208	60	2
Breast	569	30	2
Arrhythmia	452	278	16

On SRBCT dataset (Fig. 7a), 5 genes of SCF and SU both produce an error of 0, but the decline rate of SCF curve is steeper than that of SU. The error rates of FCBF, JMIM, NMIFS, MRMRR, ReliefF, and DISR are 4.86, 2.5, 2.5, 1.25, 1.25 and 1.25 percent respectively.

On colon dataset (Fig. 7b), SCF obtains the lowest classification error of 11.43 percent with 3 genes, which greatly outperforms the baselines. On Leukemia dataset (Fig. 7c), the classification error rate of SCF, SU, mRMR and NMIFS are exactly the same and all produces 0 error, which means

candidate gene sets of SCF and these baselines contains identical informative genes. Similarly, FCBF and DISR also generates 0 error rate. In addition, ReliefF and DISR both obtains the worst classification error rate of 1.25 percent, but 4 genes of the former is used and 2 genes of the latter is used. On Carcinomas dataset (Fig. 7d), SCF genes outperform the baselines genes in terms of the error rate. SCF produces an error of 4.05 percent, while DISR and JMIM both produces an error of 6.31 percent that are inferior to SCF, and other methods are more inferior to SCF. On Prostate dataset (Fig. 7e), SCF and ReliefF both produces 0 error that is better than the baselines.

Experimental results indicate that wrapper can design a better combination for SCF candidate genes, which means genes selected by SCF contains informative genes. Hence, SCF candidate gene space contains a more characteristics gene subspace than the baselines.

TABLE 3. The lowest error of SVM and KNN on three low dimensional datasets.

	Sonar		Breast		Arrhythmia	
	SVM	KNN	SVM	KNN	SVM	KNN
FCBF	0.2260	0.1776	0.0456	0.0720	0.4421	0.4114
ReliefF	0.2310	0.1683	0.0281	0.0738	0.3915	0.3738
SU	0.1857	0.1533	0.0246	0.0755	0.4578	0.3783
MRMR	0.2071	0.1681	0.0299	0.0790	0.4533	0.3539
DISR	0.2167	0.1731	0.0263	0.0790	0.4578	0.3828
JMIM	0.1976	0.1774	0.0246	0.0790	0.4578	0.3828
NMIFS	0.1976	0.1774	0.0246	0.0790	0.4578	0.3828
SCF	0.2114	0.1395	0.0246	0.0544	0.4400	0.3582

D. EXTENSIVE COMPARATIVE STUDY ON LOW DIMENSIONAL DATASETS

To further verify the effectiveness of the proposed method, we also conduct the extensive experiment on three low dimensional datasets. Three datasets from the UCI Repository are used here and these datasets have been previously used in the related research. The characteristics of these datasets are summarized in Table 2.

Table 3 shows the lowest classification error of each method on three low dimensional datasets. For example, on Sonar dataset, SCF obtains the lowest error of SVM, 21.14 percent, which is worse than SU, MRMR, JMIM and NMIFS. On Breast dataset, SCF achieves the lowest error, 2.46 percent that is same as SU, JMIM and NMIFS. On Arrhythmia dataset, the results of SCF is not as satisfactory as ReliefF, but is better than those of other methods. Nonetheless, SCF gains the best results from KNN on three datasets. In general, the experimental results verify that SCF is efficient on the low dimensional dataset.

E. TIME COMPLEXITY ANALYSIS

In this subsection, we compare the time complexity of the proposed SCF and the baselines: SU, ReliefF, FCBF, MRMR, DISR, JMIM and NMIFS. Table 4 shows the theoretical time complexity for each method and the average time for 10 runs on the carcinom dataset. The parameters are as follows: $m = 174$ (the number of samples) and $k = 200$ (the number of the selected genes), the former is the number of the instances in dataset and the latter is the number of features that need to be selected. From the table, it is obvious that the time cost of our method is little expensive, which is only less than the time cost of DISR. The time complexity of SCF is $O((k + m\log m)n)$ due to the relevance estimate of SCF combines SU and ReliefF and SCF is based on ‘maximum’ approach. Consequently, SCF is only better than DISR in terms of run time. However, considering the elevated accuracy of SCF, its slightly high time complexity could be tolerated. On the other hand, FCBF gets the best time complexity $O(n)$ when only one feature is selected and get the worst time complexity $O(n^2)$ when all features are selected.

F. DISCUSSION

According to Figs. 2,3,4,5 and 6, it can illustrate that SCF can discover more compact and discriminant features from

TABLE 4. Time complexity analysis of algorithm.

Method	Complexity	Running Time
SU	$O(n)$	0.2496s
FCBF	$O(n)/O(n^2)$	12.1525s
ReliefF	$O(nm\log m)$	13.0573s
MRMR	$O(kn)$	8.8921s
DISR	$O(n^2)$	22.9321s
JMIM	$O(n\log n)$	16.7542s
NMIFS	$O(n\log n)$	15.5281s
SCF	$O((k + m\log m)n)$	20.6857s

the original feature set, and its features are more informative than the baselines.

Next, SRBCR dataset is taken as an example (Fig. 2). In terms of classifier SVM and KNN, with the 20 genes (the top 0.2 percent genes) SCF can obtain a error rate of 0. For the baselines, their minimum error rates and the corresponding number of genes are as following: based on SVM, 190 SU genes obtain an lowest error rate of 1.25 percent, 80 mRMR genes obtain 2.36 percent, 25 FCBF genes produce an error rate of 2.5 percent, genes of DISR, JMIM and NMIFS perform even worse; based on KNN, 135 SU genes obtain 0 error, 155 mRMR genes achieve 0 error, 50 FCBF genes produce an error rate of 2.5 percent, 65 NMIFS genes obtain 0 error, 100 DISR genes achieve 0 error and JMIM genes perform worse. Thus, SCF obtains the lowest classification error and the number of corresponding SCF genes is the smallest, suggesting that SCF can select informative genes efficiently with holding higher the prediction accuracy. Moreover, the results from other gene datasets are consistent with the conclusions made on SRBCR dataset.

Additionally, Fig. 7 illustrates that SCF can be efficiently used as first-stage algorithm combined with wrapper, and this combination can be considered as a new hybrid method. This hybrid method consisting of SCF and wrapper outperforms other combinations of the baselines. Hence, the feature space of SCF is of higher characteristic strength.

Compared with the competing methods, SCF is proved to achieve the goal of improving the classification accuracy on GEM data through the experiments. However, it should be noted that SCF has one limitation: the determination of the optimal value of the weight parameter μ in (7) is cumbersome since the optimal value of parameter differs on different gene dataset.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a new score-based criteria fusion feature selection method (SCF) is proposed for cancer prediction. SCF fuses two feature ranking methods for the estimation of relevance between features and classes, which can obtain a better estimation than baselines. Meanwhile, SCF adopts the redundancy analysis to avoid redundant features within the feature subset. In general, SCF can select a good feature subset exhibiting high correlation with classes while containing low redundancy. To test the effectiveness of the proposed method, experiments are conducted on five GEM datasets and three low dimensional datasets. Experimental results show that SCF greatly improve the prediction accuracy of the model, indicating that SCF can find a set of informative features that have more characteristics strength than SU, ReliefF, FCBF, mRMR, DISR, JMIM and NMIFS when dealing with high-dimensional GEM data. Besides, SCF, as a first-stage method for preprocessing, can be combined with other methods (such as wrapper) to discover a better combination of features. This combination method is superior to other combination methods involving competing methods.

However, it should be noted that SCF has an obvious limitation (i.e., difficulty in assignment of the weighted parameter). In the next step, we will commit to design a new combination method or a score normalization method to cope with this issue.

ACKNOWLEDGMENT

The authors would like to appreciate all anonymous reviewers for their insightful comments and constructive suggestions to polish this paper.

REFERENCES

- [1] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Nat. Sci. Rev.*, vol. 1, no. 2, pp. 293–314, 2014.
- [2] J. Fan and J. Lv, "Sure independence screening for ultrahigh dimensional feature space," *J. Roy. Statist. Soc. Ser. B*, vol. 70, no. 5, pp. 849–911, 2008.
- [3] Y. Saeyns, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
- [4] C. Lazar et al., "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 4, pp. 1106–1119, Jul./Aug. 2012.
- [5] S. Guo, D. Guo, L. Chen, and Q. Jiang, "A centroid-based gene selection method for microarray data classification," *J. Theor. Biol.*, vol. 400, pp. 32–41, Jul. 2016.
- [6] W. Cheng, M. Zhao, N. Xiong, and K. T. Chui, "Non-convex sparse and low-rank based robust subspace segmentation for data mining," *Sensors*, vol. 17, no. 7, p. 1633, 2017.
- [7] M. Luo, F. Nie, X. Chang, Y. Yang, A. G. Hauptmann, and Q. Zheng, "Adaptive unsupervised feature selection with structure regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 944–956, Apr. 2018.
- [8] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 648–660, Feb. 2018.
- [9] S. Guo, D. Guo, L. Chen, and Q. Jiang, "A L1-regularized feature selection method for local dimension reduction on microarray data," *Comput. Biol. Chem.*, vol. 67, pp. 92–101, Apr. 2017.
- [10] J. He, D. Wu, N. Xiong, and C. Wu, "Orthogonal margin discriminant projection for dimensionality reduction," *J. Supercomput.*, vol. 72, no. 6, pp. 2095–2110, 2016.
- [11] F. Yang and K. Z. Mao, "Robust feature selection for microarray data based on multicriterion fusion," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 8, no. 4, pp. 1080–1092, Jul./Aug. 2011.
- [12] W. Li and Y. Yang, "How many genes are needed for a discriminant microarray data analysis," in *Methods of Microarray Data Analysis*. Boston, MA, USA: Springer, 2002, pp. 137–149.
- [13] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 856–863.
- [14] F. K. Ahmade, N. M. Norwawi, S. Deris, and N. H. Othman, "A review of feature selection techniques via gene expression profiles," in *Proc. Int. Symp. Inf. Technol.*, Aug. 2008, pp. 1–7.
- [15] F. Yang and K. Z. Mao, "Improving robustness of gene ranking by multi-criterion combination with novel gene importance transformation," *Int. J. Data Mining Bioinf.*, vol. 7, no. 1, pp. 22–37, 2013.
- [16] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 261–274, Jun. 2008.
- [17] D. Singh et al., "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.
- [18] A. A. Raweh, M. Nassef, and A. Badr, "A hybridized feature selection and extraction approach for enhancing cancer prediction based on dna methylation," *IEEE Access*, vol. 6, pp. 15212–15223, 2018.
- [19] J. Xu, H. Mu, Y. Wang, and F. Huang, "Feature genes selection using supervised locally linear embedding and correlation coefficient for microarray classification," *Comput. Math. Methods Med.*, vol. 2018, no. 2, 2018, Art. no. 5490513.
- [20] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 5, pp. 971–989, Sep. 2016.
- [21] W. Duch, T. Wieczorek, J. Biesiada, and M. Blachnik, "Comparison of feature ranking methods based on information entropy," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 1415–1419.
- [22] A. Wang, N. An, J. Yang, G. Chen, L. Li, and G. Alterovitz, "Wrapper-based gene selection with Markov blanket," *Comput. Biol. Med.*, vol. 81, pp. 11–23, Feb. 2017.
- [23] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Incremental wrapper-based gene selection from microarray data for cancer classification," *Pattern Recognit.*, vol. 39, no. 12, pp. 2383–2392, 2006.
- [24] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "MIFS-ND: A mutual information-based feature selection method," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6371–6385, 2014.
- [25] H.-H. Hsu, C.-W. Hsieh, and M.-D. Lu, "Hybrid feature selection by combining filters and wrappers," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8144–8150, 2011.
- [26] Q. Shen, R. Diao, and P. Su, "Feature selection ensemble," in *Turing*, 2012, pp. 289–306.
- [27] J. J. Chen, A. Song, and W. Zhang, "A novel hybrid gene selection approach based on ReliefF and FCBF," *Int. J. Digit. Content Technol. Appl.*, vol. 5, no. 10, pp. 404–411, 2011.
- [28] Z. Yi, C. Ding, and L. Tao, "Gene selection algorithm by combining reliefF and mRMR," *BMC Genomics*, vol. 9, p. S27, Sep. 2008.
- [29] J.-T. Kim, H.-Y. Kum, and J.-H. Kim, "A comparative study of filter methods based on information entropy," *J. Korean Soc. Marine Eng.*, vol. 40, no. 5, pp. 437–446, 2016.
- [30] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, Jan. 2014.
- [31] G. Brown, A. Pocock, M.-J. Zhao, and M. Lujan, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *J. Mach. Learn. Res.*, vol. 13, pp. 27–66, Jun. 2012.
- [32] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the gini index and information gain criteria," *Ann. Math. Artif. Intell.*, vol. 41, no. 1, pp. 77–93, 2004.
- [33] M. A. Hall and L. A. Smith, *Practical Feature Subset Selection for Machine Learning*. Springer, 1998.
- [34] J. Yang, Y. Liu, Z. Liu, X. Zhu, and X. Zhang, "A new feature selection algorithm based on binomial hypothesis testing for spam filtering," *Knowl.-Based-Syst.*, vol. 24, no. 1, pp. 904–914, 2011.
- [35] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," in *Proc. UAI*, 2012, pp. 266–273.
- [36] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2005, pp. 507–514.

- [37] I. Kononenko, "Estimating attributes: Analysis and extensions of relief," in *Proc. Eur. Conf. Mach. Learn. (ECML)*, vol. 784, Apr. 1994, pp. 171–182.
- [38] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.
- [39] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, no. 12, pp. 1205–1224, 2004.
- [40] 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.
- [41] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.
- [42] N. Kwak and C.-H. Choi, "Input feature selection for classification problems," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 143–159, Jan. 2002.
- [43] M. Bennasar, Y. Hicks, and R. Setchi, "Feature selection using joint mutual information maximisation," *Expert Syst. Appl.*, vol. 42, pp. 8520–8532, Sep. 2015.
- [44] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.
- [45] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. 1st Int. Workshop Multiple Classifier Syst.*, vol. 1857, 2000, pp. 1–15.
- [46] J. Khan *et al.*, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Med.*, vol. 7, no. 6, pp. 673–679, 2001.
- [47] T. R. Golub *et al.*, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, 1999.
- [48] U. Alon *et al.*, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Natl. Acad. Sci. USA*, vol. 96, no. 12, pp. 6745–6750, 1999.
- [49] K. Yang, J. Li, Z. Cai, and G. Lin, "A model-free and stable gene selection in microarray data analysis," in *Proc. 5th IEEE Symp. Bioinf. Bioeng. (BIBE)*, Oct. 2006, pp. 3–10.



YAN WU received the Ph.D. degree with Southern Illinois University Carbondale, with a focus on environmental chemistry and ecotoxicology. He is currently a Post-Doctoral Associate with the School of Public and Environmental Affairs, Indiana University Bloomington. His research interests involve elucidations of environmental fate of contaminants using chemical and computational techniques, predictions of their associated effects on wildlife and public health, and data processing and analysis in environmental-related fields.



NEAL N. XIONG received the Ph.D. degree in sensor system engineering from Wuhan University and the Ph.D. degree in dependable sensor networks from the Japan Advanced Institute of Science and Technology. Before he joined Northeastern State University Tahlequah, Tahlequah, OK, USA, he was with Georgia State University, Wentworth Technology Institution, and Colorado Technical University about 10 years. He is currently an Associate Professor with the Department of Mathematics and Computer Science, Northeastern State University Tahlequah. His research interests include cloud computing, security and dependability, parallel and distributed computing, networks, and optimization theory.

He has published over 280 international journal papers and over 120 international conference papers. Some of his works were published in the IEEE JSAC, the IEEE TRANSACTIONS or *ACM Transactions*, the ACM Sigcomm Workshop, the IEEE INFOCOM, ICDCS, and IPDPS. He has received the Best Paper Award from the 10th IEEE International Conference on High Performance Computing and Communications in 2008 and the Best Student Paper Award from the 28th North American Fuzzy Information Processing Society Annual Conference in 2009. He was the General Chair, the Program Chair, the Publicity Chair, a PC member, and an OC member of over 100 international conferences. He was a reviewer of about 100 international journals, including the IEEE JSAC, the IEEE SMC (Park: A/B/C), the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON MOBILE COMPUTING, and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. He is serving as an Editor-in-Chief, an Associate Editor, or an Editor Member for over 10 international journals, including an Associate Editor for the IEEE TRANSACTION ON SYSTEMS, MAN & CYBERNETICS: SYSTEMS and *Information Sciences*, an Editor-in-Chief of the *Journal of Internet Technology* and the *Journal of Parallel & Cloud Computing*, and a Guest Editor for over 10 international journals, including the *Sensors* Journal, WINET, and MONET.

Dr. Xiong is a Senior Member of the IEEE Computer Society. He is also the Chair of Trusted Cloud Computing Task Force, IEEE Computational Intelligence Society, <http://www.cs.gsu.edu/cscnx/index-TF.html>, and the Industry System Applications Technical Committee, <http://ieeecs.org/technical/isatc/>.



WENJUN KE is currently pursuing the master's degree in computer science and technology with the School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. His research interests include neural networks and machine learning, especially in dimensionality reduction and computer vision.



CHUNXUE WU received the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Beijing, China, in 2006. He is currently a Professor with the Computer Science and Engineering and the Software Engineering Division, School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. His research interests include, wireless sensor networks, distributed and embedded systems, wireless and mobile systems, and networked control systems.