Knowledge-Based Systems 146 (2018) 104–117

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Subtype dependent biomarker identification and tumor classification from gene expression profiles

Aiguo Wang[a], Ning An[a,*], Guilin Chen[b], Li Liu[c], Gil Alterovitz[d,e]

[a] School of Computer and Information, Hefei University of Technology, Hefei, China
[b] School of Computer and Information Engineering, Chuzhou University, Chuzhou, China
[c] School of Software Engineering, Chongqing University, Chongqing, China
[d] Center for Biomedical Informatics, Harvard Medical School, Boston, USA
[e] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA

ARTICLE INFO

Article history:
Received 2 October 2017
Revised 19 January 2018
Accepted 26 January 2018
Available online 31 January 2018

Keywords:
Biomarker identification
Tumor subtype
Gene selection
Microarray data
Subtype dependent

ABSTRACT

Gene expression profiles are being used to categorize disease specific genes and classify different tumor subtypes at the molecular level. Due to the inherent nature of these data having high dimensionality and small sample sizes, current conventional machine learning and statistical techniques have drawbacks in achieving satisfactory predictive classification performance in clinical samples. The typical approach to handling this situation is to eliminate noisy and redundant genes from the original gene space. There are currently multiple gene selection methods available, but most of them seek to find a common subset of genes for all tumor subtypes and fail to reflect the unique characteristics of each subtype. Consequently, in this study, we propose a general framework that aims to identify subset of genes for each tumor subtype, and also give another gene selection framework that combines the obtained subtype specific gene subsets into a single gene subset. We then present a corresponding classification model for distinguishing different tumor subtypes, and implement three specific gene selection algorithms within the two frameworks. Finally, extensive experimental results on the six benchmark microarray data validate the proposed tumor subtype dependent selection process to predict and rank specific molecular biomarkers to define tumor subtypes. This new process contributes significantly to the enhancement of tumor-predictive classification performance.

1. Introduction

Tumor metastasis and subsequent mortality place a heavy social and fiscal burden on our society. Early diagnosis of tumor is more cost effective and plays a significant role in better management, treatment, and outcomes [1]. Traditional diagnostic methods include cell based observational and biochemical examination in an organ based context, both of which rely on vast and varied domain knowledge of pathological research. Guidelines and standards of care have progressed yet maintain inherent disadvantages of bias, time, and limited accuracy. Gene mutation and subsequent loss of function or alteration in molecular pathways is a defining occurrence for most metastatic events, and measuring the differential gene expression patterns in tumor cells compared with those of a normal population is increasingly accepted to diagnose cancer, define treatments, and predict outcomes in personalized cancer care plans [2].

The rapid development and wide use of microarray technology enables simultaneous measures of expression perturbations of thousands of genes under multiple experimental conditions. These early multivariate analyses have increased our capacity to identify disease genes, drug targets, and tumor subtypes [2–5]. Accordingly, various methods of analysis, including machine learning algorithms, have been created to compare gene expression profiles. The intrinsic nature of these microarray data collections is usually characterized by high dimensionality (with thousands of gene observations over time and context) and often using a small sample size of specimens or patients to limit the statistical power for clinical use [6]. This multiplicity of classifiers and data dimensionality often causes pattern profiles to be overfit and thus predictions will suffer from poor generalization capacity [7]. There are studies suggesting that there are a few important genes that are associated with a specific classification of cancer subtypes and may be (ideally) submitted for Food and Drug Administration (FDA) validation and used for diagnosis [8]. Also, the affected gene space often

* Corresponding author.
E-mail addresses: wangaiguo2546@163.com (A. Wang), ning.g.an@acm.org (N. An), glchen@chzu.edu.cn (G. Chen), dcsliuli@cqu.edu.cn (L. Liu), gil_alterovitz@hms.harvard.edu (G. Alterovitz).

https://doi.org/10.1016/j.knosys.2018.01.025
0950-7051/

consists of a large number of noisy and redundant genes, which can diminish the performance of a classifier [9–11]. For example, *k*-nearest-neighbor algorithm is sensitive to irrelevant features in classification [12]. One feasible way to mitigate this problem is to select a subset of discriminant genes from the original gene space by filtering noisy and redundant genes using effective feature selection methods [13,14].

Gene selection, also known as feature selection and variable selection, is defined as a process of selecting a small subset of genes that contains the most discriminant information with well-defined evaluation metrics [6]. In addition to reducing the dimensionality of original gene space, effective gene selection methods bring us significant enhancements of quality measures for defining gene sets that validate the drug targets in biological and medical research. These enhancements include better generalization capacity of the constructed classifier, reducing the classifier training time, and improving the interpretability of obtained biomarkers [15].

According to whether using a classifier to evaluate the quality of a candidate feature in the feature selection process, existing feature selection methods can be broadly divided into four categories: (1) filter methods, (2) wrapper methods, (3) embedded methods, and (4) hybrid methods [16,17]. Filter methods are independent of a classification model and measure the quality of a feature, or a subset, using only the intrinsic nature of training samples. Filter methods are flexible in combination with various classifiers and have lower computational complexity. They also have better generalization ability [16]. Furthermore, commonly used metrics in filter methods mainly include distance, consistency, dependency, and information theory-based metrics [18]. Distance-based methods define separability as the metric and try to find those features that can best discriminate the target class. One such method is the ReliefF algorithm [19]. Consistency-based methods use the inconsistency rate as the criterion and seek to select a subset of features with better consistency, such as Focus and LVF algorithms [20]. Dependency-based methods evaluate the importance of candidate features with statistical theory, and there are a variety of methods available such as Pearson correlation coefficient, partial least squares, and Fisher score [21,22]. Information theory based feature selectors have efficiency and effectiveness because of their capacity in capturing higher order statistics of data and reflecting the non-linear relationships between variables [23]. Consequently, researchers have proposed and developed a number of feature selectors from the view of mutual information, including information gain, minimum redundancy maximum relevance (mRMR) [24], and fast correlation based filter (FCBF) [25]. In contrast, wrapper-based methods are specific to a given learning algorithm to extending non-filter features of selection to evaluate the quality of a selected candidate. These methods often use the classification error rate or classification accuracy as an evaluation criterion [26–28]. Due to the specific interaction between the obtained features and the learning algorithm, wrapper methods tend to obtain better classification results but at the cost of high time complexity [27]. Embedded methods are essentially a special case of wrapper methods and more tightly coupled with a specified learning algorithm. Feature subsets are generated during training the classifier, which makes embedded methods usually more tractable and time efficient than wrapper methods. Decision tree and Lasso algorithms are two typical embedded cases [29,30]. Besides, a hybrid scheme has been proposed to take advantage of both filter and wrapper methods, and researchers have proposed to combine filter and wrapper methods [31,32]. Essentially, a filter is initially used to remove a large number of noisy and redundant features from the original feature space, and then a wrapper method is used to find a discriminant feature subset from the reduced subset [33].

According to the final output style, we can group existing feature selection methods into feature ranking and feature subset selection categories. Feature ranking methods return a ranked list of the original features in descending order according to the predictive power of each feature [34]. We are required to specify the number of how many features are to be selected after ranking. Alternatively, we can determine the optimal size of a feature subset with the help of a learning algorithm. Feature ranking methods include single feature ranking and multiple feature ranking methods. The former evaluates the quality of each candidate feature individually, and does not consider the redundancy and interaction between features [19]. These feature ranking methods often fail to obtain a feature subset of high quality. Multiple feature ranking methods take the relationship between the candidate feature and previously selected features into account in the process of feature selection [25]. Ranking methods belonging to this category have a sequential forward or backward selection scheme to rank original features [6]. Unlike feature ranking methods, feature subset selection methods explicitly or implicitly consider the relevance and redundancy between features, and finally return a feature subset without involving a further step to determine the optimal size [25].

Currently, there are a wealth of feature selection methods available [35–38], but most of them seek to find a common subset of genes for subtypes within a defined pathology, and fail to reflect the unique characteristics of each subtype based on molecular differences. In fact, a unique subset of genes is likely to exist within each tumor subtype. Identifying these molecularly based tumor subtypes will increase the clinical efficacy of treatments with such predictive biomarkers [2,39]. Obtaining molecular subtype dependent biomarkers helps design a personalized treatment plan. These plans have been shown to often reduce the toxicity and side effects in treatment, concurrent with significant slowing of tumor progression. These biomarkers also accelerate structural and cell-based refinement in drug development research on these molecular subtypes, reducing the time and cost of bring drugs to clinic.

There are studies from related fields that propose to select a possible different feature subset for each class. For example, de Lannoy et al. propose a method to perform class-specific feature selection in multiclass support vector machines and experimentally validate its performance [40]. Zhou and Wang use class separability measure to select different feature subsets for different classes and compare their method with class independent feature selection method by applying the method on several biomedical data with support vector machine [41]. A major limitation of these methods is that they are related to the use of a particular classifier, which limits its applicability. To alleviate this problem, Pineda-Bautista et al. propose a class specific feature method that can be used with any classifiers and they use classifier ensemble to classify an unseen sample. Their experimental results on low dimensional datasets show the effectiveness of the proposed method [42]. However, classifying new test samples under an ensemble framework without utilizing the confidence of each sub-classifier may makes poor decisions when we face the problem of voting conflict. Besides, the aim of these studies is to return multiple feature subsets for feature analysis and classification model construction, and few studies, to the best of our knowledge, explore the fusion of multiple class-specific feature subsets and further evaluate the effectiveness of these combined features in classification. Furthermore, they conduct experiments on low dimensional datasets without considering a more difficult case that is characterized by high dimensionality and small sample sizes. Accordingly, in this study, we propose to select gene profiles that are associated tumor subtypes, enabling us to define unique genes for a tumor subtype as well as common genes for all tumor subtypes. We will enhance the performance in classifying different tumor subtypes and further reduce the chance of partially overfitting in future algorithms. The main contributions of this study are as follows:

1) We propose a general framework for subtype dependent biomarker identification that returns a filtered profile of genes for each tumor subtype. Subsequently, we provide another gene selection framework, called fusion based gene selection that merges the obtained subtype dependent gene profiles and finally returns a single defining gene profile. We then present corresponding classification (training and testing) model, associated with subtype dependent method, for distinguishing different tumor subtypes.
2) Under each of the frameworks, we implemented three specific gene selection algorithms with Fisher score, mRMR and FCBF as the building blocks, respectively. We have detailed how to obtain the optimal feature subset for feature ranking based as well as feature subset based feature selection methods in this study.
3) We integrate three classification models with different metrics, including support vector machine, Naïve bayes, and *k*-nearest-neighbor, into the framework to construct classifiers, and detail how to estimate the confidence that a sample belongs to a specific class to solve the problem of voting conflict.
4) We tested the proposed methods on six benchmarked microarray datasets that contain multiple tumor subtypes, and compared the performance of support vector machine, Naïve bayes, and *k*-nearest-neighbor. Extensive experiments demonstrate the superiority of subtype-dependent feature selection methods over subtype-independent feature selection methods and the superiority of support vector machine over Naïve bayes and *k*-nearest-neighbor in obtaining a feature subset of high quality.

The paper is structured in the following way. Section II details the proposed subtype dependent biomarker identification framework and its fusion version, and present corresponding subtype-dependent classification model. Section III illustrates the experimental data, three baseline feature selectors and support vector machine classifier, as well as the evaluation metrics. Section IV presents the experimental results. The last section concludes this study.

2. Subtype dependent gene selection and tumor classification framework

In this section, we first present the two proposed gene selection frameworks for biomarker identification: subtype dependent framework and its fusion version. We then show the classifier training and testing model under the subtype dependent framework.

2.1. Notations

In the analysis of gene expression profiles, we denote the microarray data as a matrix. Generally, in this study, we use $\mathbf{X} \in \mathbb{R}^{m\times n}$ to represent the data matrix, where m is the number of samples and n is the number of genes. Specifically, we use g_1, g_2, ..., g_n to represent the n genes and use $\mathbf{g}_1$, $\mathbf{g}_2$, ..., $\mathbf{g}_n$ ($\mathbf{g}_i \in \mathbb{R}^m, 1 \le i \le n$) to represent its vector forms. We also usex_1, x_2, ..., x_m to denote the m instances, and $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_m$ ($\mathbf{x}_i \in \mathbb{R}^n, 1 \le i \le m$) are corresponding vectors. We then have $\mathbf{X} = (\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_n) = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m)^{\mathrm{T}}$. Suppose that $L = \{l_1, l_2, ..., l_C\}$ denotes the label set with C different classes. We use $\mathbf{y} \in \mathbb{R}^{m\times 1}$ to represent corresponding label vector associated with the data matrix, and use y_1, y_2, ..., y_m ($y_i \in L$, $1 \le i \le m$) to denote the target values of the m instances. Then, we can denote the training set with m samples as $\mathbf{D} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbf{X}, y_i \in L, 1 \le i \le m\}$.

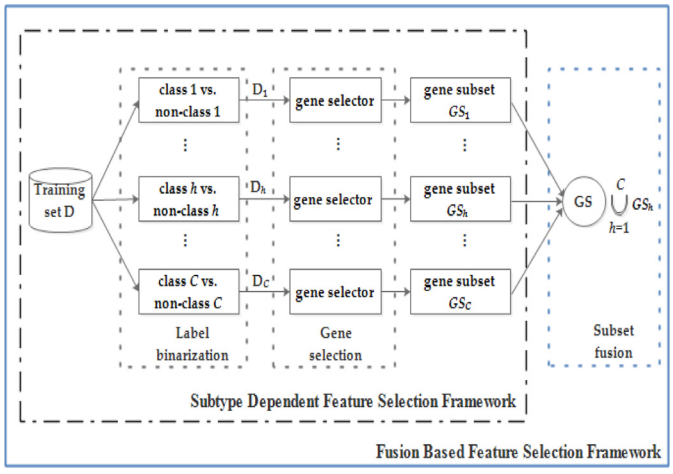

Fig. 1. An illustration to the two proposed feature selection frameworks. (a) *subtype dependent* feature selection framework; (b) *fusion* based feature selection framework. Subtype dependent framework is a building block of fusion based framework with an additional operation of subset fusion.

2.2. Subtype dependent biomarker identification

Most of existing feature selection methods select a gene subset for all tumor subtypes. For the convenience of illustration, we name this kind of method as *subtype independent* feature selectors. Alternatively, *subtype dependent* methods aim to find a subset of genes for each tumor subtype, primarily consisting of two steps as shown in Fig. 1. In the first step, we are required to convert the C-class classification problem into C two-class classification problems. Specifically, for problem h ($1 \le h \le C$), we obtain corresponding training set $\mathbf{D}_h$ by coding the class label of instance x_i ($x_i \in \mathbb{R}^n, 1 \le i \le m$) with the following strategy as represented in formula (1): label it with $+1$, if x_i belonging to class h originally; otherwise, label it with -1, if x_i belonging to other classes. We call this step the label binarization. Then, we can obtain C training sets {$\mathbf{D}_1$, ..., $\mathbf{D}_h$..., $\mathbf{D}_C$} associated with the C two-class classification problems.

$$coding_h(x) = \begin{cases} +1, y_i = h \\ -1, y_i \neq h \end{cases} \tag{1}$$

After obtaining the training set $\mathbf{D}_h$, in the second step, we conduct feature selection for the binary classification problem h to obtain its optimal gene subset GS_h. In this stage, various subtype independent feature selectors, also called traditional feature selectors, can be used on $\mathbf{D}_h$ to search for the best gene subset GS_h by working on a binary class dataset rather than on a multi-class dataset. Specifically, for feature subset based feature selection methods (e.g., fast correlation based filter), they generally return a subset of features as the finally obtained feature subset. For feature ranking based methods (e.g., fisher score), they return a ranking of the original features according to the importance of each feature. To obtain the finally obtained feature subset, we pre-select p top-ranked features, sequentially evaluate p feature subsets, and select the one with best classification performance as the finally obtained feature subset. We can then obtain C gene subsets {GS_1, ..., GS_h..., GS_C}, each of which is related to a specific tumor subtype.

To a further step, we can merge the C gene subsets using formula (2). This is the union of the C gene subsets, which covers the discriminant genes of each tumor subtype and should provide quality classification of tumors. We note this scheme as *fusion* based gene selection framework, and Fig. 1 presents the diagram.

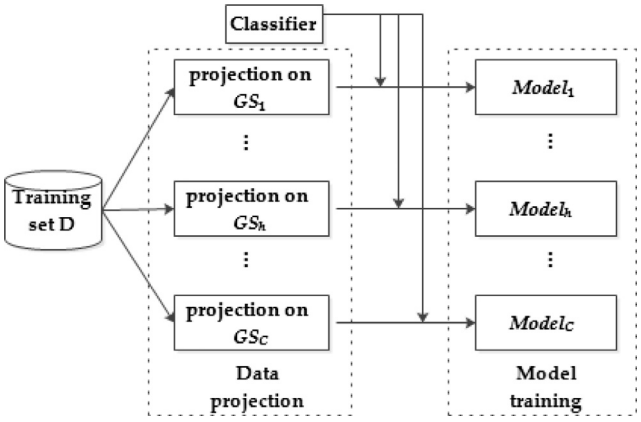

Fig. 2. Subtype dependent classifier training.

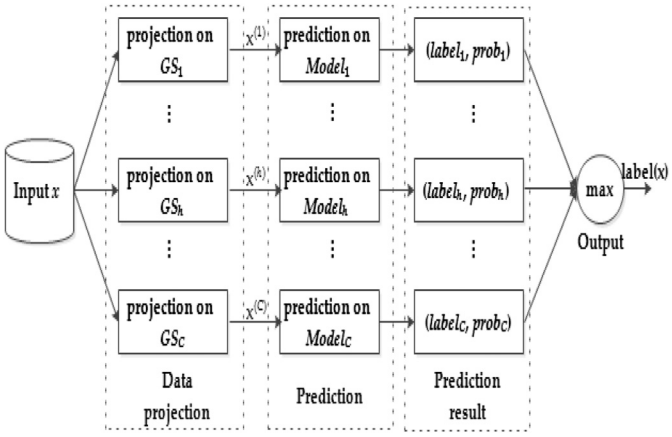

Fig. 3. Classifying test samples using the trained subtype-dependent classifier.

$$GS = \bigcup_{h=1}^{C} GS_h \tag{2}$$

2.3. Subtype dependent classification model

One of the main purposes of feature selection is to build a robust classifier for classifying tumor subtypes. Also, the performance of the classifier built on the obtained gene subsets can be used for effectiveness evaluation of the proposed method. In this subsection, we detail the process of how to construct a corresponding classifier and how to apply it for tumor classification.

2.3.1. Subtype dependent classifier training

For the classification problem h, after subtype dependent feature selection, we can obtain a corresponding training set $\mathbf{D}_h$ and the optimal gene subset GS_h. Then, our task is to train a binary classifier on $\mathbf{D}_h$. Specifically, the classifier is trained on $\mathbf{D}_h$ projected over the selected gene subset GS_h, and we get a model $Model_h$ that can be used to predict whether a test sample belongs to class h. In this step, we can use various learning algorithms, e.g., Naïve bayes, k nearest neighbor, and support vector machine. Repeating the above process, we obtain multiple C classifiers. Finally, we get the classification model by combining these classifiers together as shown in Fig. 2.

2.3.2. Classification with subtype dependent classifier

To conduct predictions, we take the test sample x as an input to the classification model. Then, each classifier $Model_h$ predicts whether x belonging to class h. After all the C classifiers have made their own predictions, we can determine to assign which label to x. However, there may exists the situation where x is assigned multiple classes (e.g. $Model_i$ labels x as class i, while $Model_j$ claims that the label of x is class j $(i \neq j)$), also called voting conflict, which makes it difficult to determine a single label. It is also possible that all the C classifiers reject x. To mitigate the above problems, we parse the classifiers with a probability output that can be used to estimate the confidence that x belongs to a certain class. Specifically, for classifier $Model_h$, it first projects x over GS_h and gets $x^{(h)}$. Then $Model_h$ works on $x^{(h)}$ and outputs the predicted label $label_h$ and probability estimation $prob_h$. Also, $label_h$ takes the value of 1 if $Model_h$ predicts that x belongs to class h; otherwise, $label_h$ equals 0. For C classifiers, we have C groups of outputs (shown in Fig. 3). With this, the predicted label of x is determined by the maximum probability rule (3).

$$label(x) = \max_i \{label_i * prob_i, 1 \leq i \leq C\} \tag{3}$$

Table 1
Experimental dataset descriptions, including the sample size, the gene size, and the number of classes.

ID	Dataset	#Samples	#Genes	#Classes	#SFR
1	Leukemia2	72	11225	3	0.006
2	Brain2	50	10367	4	0.005
3	Brain1	90	5920	5	0.015
4	9_Tumor	60	5726	9	0.010
5	11_Tumor	174	12533	11	0.014
6	14_Tumor	308	15009	26	0.021

Particularly, if the predicted labels of the C classifiers are all zeros, we determine the label of x with formula (4), which is associated with the classifier having the minimal confidence in making a prediction of zero.

$$label(x) = \min_i \{prob_i, 1 \leq i \leq C\} \tag{4}$$

3. Experimental setup

In this section, we first describe the microarray data used in our experiments, and then introduce three gene selection methods that are building blocks of the two proposed frameworks. Finally, we present the widely used support vector machine classifier, and give the evaluation metrics to measure the performance of the proposed methods.

3.1. Microarray data

In our experiments, six publicly available benchmark gene expression profiles are used in this study [43], and they are all related to multi-category classification problems. A brief summary to the six datasets is presented in Table 1. The last column SFR denotes the ratio between the number of samples and the number of genes. From SFR, we can see that there exists a great difference between the number of samples and the number of genes in each microarray data. Constructing a classifier on such a dataset easily leads to overfitting. We can also see that the dimensionality of classes in these datasets ranges from 3 to 26, which has previously posed a great challenge in computer capacity and predictive quality for parsing tumor subtypes for biomarker testing. Particularly, for 14_Tumor microarray data, it consists of 26 subtypes, and has a SFR ratio of 0.021. All the microarray datasets used in this study can be downloaded from http://www.gems-system.org/.

3.2. Baseline feature selection method

As we discussed, there are a variety of gene selection methods that are suitable for the proposed subtype dependent gene selection framework. In this study, we explore to use three proven effective feature selectors with different metrics, covering both feature ranking and feature subset selection methods, to evaluate the quality of candidate features. There are Fisher score, mRMR, and FCBF.

3.2.1. Fisher score

Statistical metrics to rank features are accomplished with the Fisher score, a dependency-based filter method. Fisher score assigns a higher weight to a feature with similar values to instances from the same class and different values to instances from different classes [21], and separates data points into different classes and clusters the data points within the same class. The evaluation metric that Fisher score uses to measure the importance of a candidate feature **g** is given by the following formula:

$$J(\mathbf{g}) = \frac{\sum_{j=1}^{C} n_j(\mu_j - \mu)^2}{\sum_{j=1}^{C} n_j \sigma_j^2} \tag{5}$$

where μ is the mean of the gene **g**, n_j is the number of samples in class j,μ_j and σ_j^2 are the mean and the variance of the samples in class j, respectively.

Therefore, Fisher score provides a basis to rank and cluster the genes. The greater Fisher score, the more important it is in contributing to the classification performance. Fisher score ranks genes in descending order.

3.2.2. Minimal redundancy maximal relevance

The minimal Redundancy Maximal Relevance (mRMR) method is built on information theory, so it has the capacity to capture higher order statistics of data and reflect non-linear relationships between variables [24]. In information theory, mutual information measures the amount of information shared by two discrete random variables **x** and **y**. This can be used to measure the relevance between two variables. Mutual information is defined as:

$$I(\mathbf{x};\mathbf{y}) = \sum_{x\in\mathbf{x}}\sum_{y\in\mathbf{y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{6}$$

where $p(x)$ is the probability distribution of variable x in **x** and $p(x, y)$ is the joint probability distribution of variable x and y in **x** and **y**, respectively.

To select a subset of genes that are relevant to the disease with minimal redundancy, mRMR uses the following cost function to evaluate the quality of a candidate gene **g** given the selected subset **S**,

$$J(\mathbf{g}) = I(\mathbf{g};\mathbf{y}) - \frac{1}{|\mathbf{S}|}\sum_{\mathbf{s}\in\mathbf{S}} I(\mathbf{g};\mathbf{s}) \tag{7}$$

in which |**S**| represents the number of selected genes, **y** is the target class, and in this study, **y** represents the tumor subtypes, I(**g**; **y**) denotes the relevance between gene **g** and class **y**, and I(**g**; **s**) measures the redundancy between genes **s** and **g**.Criteria of the mRMR select the gene that is maximally relevant to the disease and lest redundant to those already selected. Starting from an empty set and working sequentially forward, mRMR initially selects the gene that is most relevant to the disease and parsed into the selected subset, subsequently removing it from the candidate set; then (7) is used to choose the next gene that maximizes J(**g**) into the selected subset. The process is continued until all the genes are ranked or a pre-defined number of genes are obtained.

The mutual information difference form of mRMR, used by Peng et al. [24], also provides a quotient form (8) as well, which ranks features in descending order. Both of products of the mRMR are feature ranking methods. In this study, we use its difference form as a representative.

$$J(\mathbf{g}) = I(\mathbf{g};\mathbf{y}) / \left(\frac{1}{|\mathbf{S}|}\sum_{\mathbf{s}\in\mathbf{S}} I(\mathbf{g};\mathbf{s})\right) \tag{8}$$

3.2.3. Fast correlation based filter

Fast correlation based filter (FCBF) is a feature subset selection method, and selects features by identifying high relevant features and simultaneously removing redundant features [25]. FCBF defines the relevance and redundancy between two features and the relevance of a feature to the target class on the basis of the symmetric uncertainty (9). Symmetric uncertainty (SU) is a normalized mutual information and can be used to measure the relevance between two variables **g** and **y** with (9). H(X) measures the entropy of a variable X.

$$SU(\mathbf{g},\mathbf{y}) = \frac{2 * I(\mathbf{g};\mathbf{y})}{H(\mathbf{g}) + H(\mathbf{y})} \tag{9}$$

Definition 1 (*C-Relevance*)**.** Given a predictive feature **g** and the target variable **y**, the relevance between them is referred to as *C-Relevance*, denoted by SU(**g, y**).

A feature with a larger *C-Relevance* value contains more information about the class than a feature with a smaller *C-Relevance* value.

Definition 2 (*F-Relevance*)**.** The correlation between two predictive features **g** and **h** is referred to as *F-Relevance*, and noted as SU(**g, h**).

Definition 3 (*Approximate Markov blanket*)**.** Given two predictive features **g** and **h**, and the target variable **y**, if both (10) and (11) are satisfied, then **h** is redundant to **g**. **g** is called an approximate Markov blanket of **h**.

$$SU(\mathbf{g},\mathbf{y}) \geq SU(\mathbf{h},\mathbf{y}) \tag{10}$$

$$SU(\mathbf{g},\mathbf{h}) \geq SU(\mathbf{h},\mathbf{y}) \tag{11}$$

According to Markov blanket technique [25], we know that, in feature selection, if **g** is selected, then it is not necessary to choose **h** as it can not provide extra information beyond the information from **g**.For FCBF, selection of final feature subset requires two steps. In the first step, FCBF filters out those features whose relevance with the target class is less than a predefined threshold.

$$SU(\mathbf{g},\mathbf{y}) < \gamma \tag{12}$$

FCBF then eliminates redundant features using approximate Markov blanket technique in the second step. FCBF can then obtain a subset of features that are highly relevant to the target class and less redundant to each other. In this study, we set $\gamma = 0$, indicating that only features that are independent from the target class are removed in the first stage.

3.3. Support vector machine

They are many learning algorithms with probability estimation that may be integrated into the subtype dependent classification model as we have defined. In this study, we selected the proven support vector machine (SVM). SVM is a state-of-the-art classification and regression tool, and has been successfully applied in many fields, such as computer vision, text classification and bioinformatics [44]. SVM was originally designed to parse binary classification. The traditional binary classification problem was solved in SVM by finding an optimal separating hyperplane that can separate two classes with a maximal margin between the two classes. Specifically, assume that we have a training set with m labeled samples

Table 2
Confusion matrix for three-class classification problem.

		True label			
		class1	class2	class3	sum
Inferred Label	class1	TP_{11}	FP_{12}	FP_{13}	NI_1
	class2	FP_{21}	TP_{22}	FP_{23}	NI_2
	class3	FP_{31}	FP_{32}	TP_{33}	NI_3
	sum	NT_1	NT_2	NT_3	Total

$\{(x_i, y_i)\}_{i=1}^{m}$, in which $x_i \in \mathbb{R}^n$ is an instance, and $y_i \in \{-1, 1\}$ is its class label. To separate the two classes, SVM aims to solve the following optimization problem (13).

$$\begin{cases} \min \frac{1}{2}||w||^2 + \lambda \sum_{i=1}^{m} \xi_i \\ s.t. \\ y_i[(w^T x_i) + b] \geq 1 - \xi_i, i = 1, 2, ..., m \\ \xi_i \geq 0, i = 1, 2, ..., m \end{cases} \tag{13}$$

In (13), $\lambda > 0$ is the penalty parameter to make a tradeoff between the training error and the margin.

Particularly, for the non-linear case, SVM can map the data points to a higher dimensional space, and find the optimal separating hyperplane in that space. It is often difficult and even impossible to explicitly define an appropriate mapping function, but we can solve this problem with kernel trick [44]. There are a variety of kernel functions available such as polynomial function, radial basis function and sigmoid function, and users can define various kernel functions as well.

For probability estimation, SVM transforms the decision values into probability values using (14),

$$prob(x) = \frac{1}{\exp(A\hat{y} + B)} \tag{14}$$

where $\hat{y}$ is the decision value of x, A and B are estimated by minimizing the negative log likelihood of training set. Thus, SVM can make a probability estimate that the test instance derived from the predicted label.

3.4. Evaluation measures

To show the superiority of the proposed methods in selecting biomarkers and their performance in classifying different tumor subtypes, we evaluate the proposed frameworks from two different aspects: the selected biomarkers and corresponding classification performance.

In analyzing the selected biomarkers, we compare the number of selected genes for subtype independent, subtype dependent, and fusion based methods. We also compare time costs in obtaining the optimal gene subsets. We evaluate gene relationships and classification to show whether the proposed methods enable us to obtain unique genes for a tumor subtype with a faster, better predictive model.

3.4.1. Classification performance measures

To evaluate the quality of a gene selector, classification performance is a direct and effective criterion and is much more important. A feature selector obtaining poor classification results is of little use in tumor subtype identification. In the evaluation, a confusion matrix that contains the actual labels and predicted labels is applicable to measure the classification performance [45]. Table 2 presents a confusion matrix for tumor subtype classification in the case of three classes. Accordingly, we use accuracy, precision, recall, and F1 to show the classification performance, and the higher their values, the better the constructed classifier with the selected features.

Accuracy is the probability of correctly classifying each sample, it equals the number of samples that are correctly grouped and can be obtained with (15).

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} TP_{ii}}{total} = \frac{\sum_{i=1}^{C} TP_{ii}}{\sum_{i=1}^{C} NI_i} = \frac{\sum_{i=1}^{C} TP_{ii}}{\sum_{i=1}^{C} NT_i} \tag{15}$$

Precision represents the weighted average of the fraction of the inferred labels that are correctly predicted for each tumor subtype. For a classification problem with C classes, *Precision* can be calculated with (16),

$$\text{Precision} = \frac{1}{C} \sum_{i=1}^{C} \frac{TP_{ii}}{NI_i} \tag{16}$$

where TP_{ii} is the number of test samples that are correctly classified for the inferred label i; NI_i shows the total number of test samples that are classified as label i, and equals the sum of the numbers in corresponding row.

$$NI_i = TP_{ii} + \sum_{j=1, j\neq i}^{C} FP_{ij} \tag{17}$$

Recall refers to the weighted average of the fraction of the true labels that are correctly classified for each tumor subtype. For a classification problem with C classes, we can measure *Recall* using (18).

$$\text{Recall} = \frac{1}{C} \sum_{i=1}^{C} \frac{TP_{ii}}{NT_i} \tag{18}$$

in which NT_i indicates the number of test samples with true label i, and can be obtained by totaling the numbers of corresponding column.

$$NT_i = TP_{ii} + \sum_{j=1, j\neq i}^{C} FP_{ji} \tag{19}$$

F1, calculated using formula (20), provides a way to combine precision and recall into a single metric, and is often used when making evaluations on imbalanced datasets. F1 takes a real number between 0 and 1, and 1 indicates that the classifier can correctly classify all test samples.

$$\text{F1} = \frac{2 * precision * recall}{precision + recall} \tag{20}$$

3.5. Experimental setup

In this study, we use Fisher score, mRMR, and FCBF, respectively, as the foundation of the proposed subtype dependent and independent gene selection frameworks. Each feature selector is required to obtain an optimal gene subset. Since FCBF is a feature subset selection method, we can directly use that output for selected genes. For feature ranking methods, such as Fisher score and mRMR, to find the best feature subset, we investigate the p top-ranked features and further search the best ranked feature subsets. Specifically, we first generate feature subsets by picking the top p features sequentially. Then, we obtain p feature subsets and construct classifiers on training set projected over each of the feature subsets. The classifier that achieves the best classification performance corresponds to the best feature subset [34]. If two or more feature subsets produce equal classification accuracy, the one with smallest number of features is selected. As there are studies suggesting that only a few important genes are associated with a certain cancer and therefore sufficient for the diagnosis [8]. In this study, we assess the proposed method with a few candidate genes, and consider $p = 50$. Also, we use the average classification accuracy, which is obtained by the leave-one-out cross validation coupled with SVM, to measure the quality of a feature subset [6].

Table 3
Number of selected genes for subtype-dependent and independent methods on Leukemia2.

Methods	classes	Fisher	mRMR	FCBF
Subtype Independent	all	46	6	94
Subtype	class1	18	2	97
Dependent	class2	13	5	65
	class3	43	35	48
	AVE	24	14	70
Fusion	all	74	42	207
	Common	28	4	70
	Distinct	46	38	137

Table 4
Number of selected genes for subtype-dependent and independent methods on Brain2.

Methods	classes	Fisher	mRMR	FCBF
Subtype Independent	all	40	5	113
Subtype	class1	18	10	41
Dependent	class2	7	10	44
	class3	5	26	43
	class4	4	18	40
	AVE	8	16	42
Fusion	all	33	64	167
	Common	10	2	33
	Distinct	23	62	134

To evaluate the quality of the terminally selected feature subset, we use the stratified k-fold cross validation to evaluate the subtype independent method, subtype dependent method, and fusion based method. In the k-fold cross validation scheme, the data is divided with equal sizes, and one of the k folds is used as the test set and the remaining (k-1) folds are used as the training set for the classifier construction. The final classification accuracy is the average of the k results [6]. For each selected subset, a SVM classifier with linear kernel and default C parameter in LIBSVM [44] is trained on the training set, and tested on corresponding test set. Due to small number of instances of some tumor subtypes, we use three-fold cross validation. Also, each feature selector is evaluated under the same training and test sets. Additionally, to handle the multi-class classification problem with SVM, "one-against-one" strategy is used to label a test instance, that is, when classifying an instance, it constructs all possible two-class classifiers and determines its label with a simple majority rule.

4. Experimental results and analysis

4.1. Selection of discriminant genes

In this section, we report the selected genes of each feature selector from two aspects: the number of selected genes, and the relations between genes selected with subtype dependent and independent methods.

4.1.1. Number of selected genes

For each of the baseline feature selectors, including Fisher score, mRMR and FCBF, we record the number of selected features for subtype independent method, subtype dependent method, and fusion based method. Tables 3–8 present experimental results of the six microarray datasets, respectively. The last three columns represent the three used baseline feature selectors, in which the column "Fisher" is Fisher score for short. Of note, subtype independent method selects a subset of genes for all tumor subtypes, thus it returns a single gene subset. In contrast, subtype dependent method finds a unique gene subset for each tumor subtype, so the number of selected gene subsets equals the number of tumor subtypes. Fusion based method returns a gene subset, which is a union of the subsets obtained using subtype dependent method, thus it returns a gene subset for all tumor subtypes. Additionally, in each table, the row "AVE" presents the average number of genes selected by subtype dependent method over all tumor subtypes.

Table 5
Number of selected genes for subtype-dependent and independent methods on Brain1.

Methods	classes	Fisher	mRMR	FCBF
Subtype Independent	all	31	18	244
Subtype	class1	16	40	50
Dependent	class2	17	8	48
	class3	2	12	51
	class4	1	6	40
	class5	8	4	28
	AVE	8	14	43
Fusion	all	43	69	209
	Common	9	7	82
	Distinct	34	62	127

Table 6
Number of selected genes for subtype-dependent and independent methods on 9_Tumor.

Methods	classes	Fisher	mRMR	FCBF
Subtype Independent	all	13	36	502
Subtype	class1	4	36	49
Dependent	class2	3	9	40
	class3	8	32	49
	class4	4	9	40
	class5	16	5	55
	class6	7	6	44
	class7	4	12	47
	class8	2	15	22
	class9	5	6	38
	AVE	5	14	42
Fusion	all	53	129	368
	Common	10	11	135
	Distinct	43	118	233

Table 7
Number of selected genes for subtype-dependent and independent methods on 11_Tumor.

Methods	classes	Fisher	mRMR	FCBF
Subtype Independent	all	47	37	2008
Subtype	class1	27	2	112
Dependent	class2	6	30	69
	class3	14	13	120
	class4	41	26	146
	class5	4	5	87
	class6	8	8	94
	class7	1	2	74
	class8	1	1	92
	class9	1	1	56
	class10	48	6	64
	class11	22	6	86
	AVE	15	9	90
Fusion	all	173	100	965
	Common	9	3	459
	Distinct	164	97	506

From Tables 1 and 3–8, we can see that all the feature selection methods can significantly reduce the original feature dimensionality, indicating that there exist a substantial number of irrelevant and redundant features in the gene expression profiles. Then, we observe that for the three types of feature selection methods, using a different baseline feature selector leads to a different gene subset. For example, on Leukemia2, when using Fisher score, subtype independent method selects 46 genes in comparison to 6 genes and 96 genes obtained using mRMR and FCBF, respectively. Also, subtype dependent method with FCBF selects 97, 65, and 48 genes for the three tumor subtypes, respectively, while

Table 8
Number of selected genes for subtype-dependent and independent methods on 14_Tumor.

Methods	classes	Fisher	mRMR	FCBF
Subtype Independent	all	50	49	1109
Subtype	class1	30	13	114
Dependent	class2	3	8	19
	class3	1	34	18
	class4	23	15	15
	class5	50	28	38
	class6	45	17	20
	class7	13	7	18
	class8	25	13	18
	class9	10	42	20
	class10	40	15	18
	class11	8	32	18
	class12	23	49	20
	class13	4	20	27
	class14	40	35	64
	class15	4	1	18
	class16	26	32	20
	class17	1	1	16
	class18	11	34	23
	class19	2	31	21
	class20	5	31	23
	class21	1	32	15
	class22	2	5	18
	class23	7	4	24
	class24	21	49	21
	class25	1	8	10
	class26	3	14	31
	AVE	15	21	25
Fusion	all	386	540	642
	Common	9	10	201
	Distinct	377	530	441

it selects 18, 13, and 43 genes using Fisher score. Accordingly, the number of selected genes for fusion based method also varies with the used baseline feature selector. This is reasonable, since Fisher score, mRMR, and FCBF use different metrics to evaluate the goodness of a candidate feature.

We also observe that subtype dependent method selects gene subsets of different sizes for different subtypes. For instance, on Leukemia2 dataset, when using Fisher score, it selects 18 genes for class 1, 13 genes for class 2, and 43 genes for class 3; in the case of mRMR, it obtains 2, 5, and 35 genes for three tumor subtypes, respectively; and when using FCBF, the number of selected genes are 97, 65, and 48 in turn. Similar results can be observed on other datasets. This indicates that there probably exist different gene subsets for different tumor subtypes, which is the main reason that motivates us to conduct this study.

In addition, we can see that the average number of selected genes using subtype dependent method is generally smaller than that of subtype independent method, and that fusion based method always obtains a gene subset with a larger size. For example, on Leukemia2 dataset, using Fisher score feature selector, subtype independent method selects 46 genes, subtype dependent method selects 24 genes for each tumor subtype on average, and fusion based method leads to a subset with 74 genes. On the 14_Tumor dataset with 26 tumor subtypes, using FCBF, subtype independent method selects 1109 genes, fusion based method returns 642 genes, while subtype dependent method selects 25 genes on average. From the results of column “AVE”, we see that using Fisher score or mRMR as the baseline feature selector shows advantages over FCBF in terms of the average number of genes selected. It is worth noting that selecting a smaller subset of genes helps biologists test the underlying biological mechanisms and pathways in cell based assays.

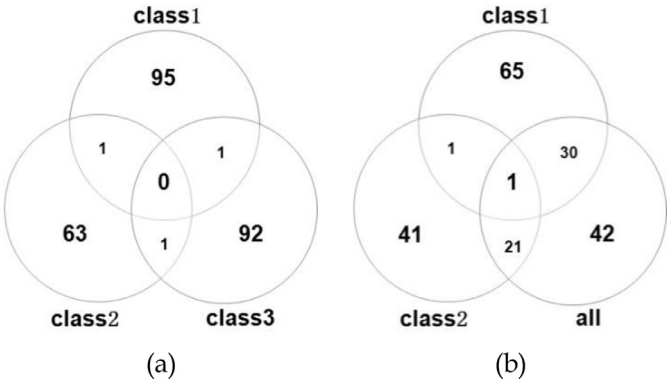

Fig. 4. Relations between genes on Leukemia2 dataset. Class1, class2 and class3 represent genes selected by subtype dependent method for each tumor subtype, and all represents genes selected by subtype independent method for all subtypes.

4.1.2. Relations between subtype-dependent and Subtype-independent biomarkers

We initially investigated the difference between subtype dependent and subtype independent method in the terminally selected genes. As shown in Tables 3–8, the row “Distinct” shows the number of genes that are selected by subtype dependent method but not by subtype independent method, and the row “Common” indicates the number of common genes selected by the two methods. From Tables 3–8, we can observe that there does exist a difference between the genes selected by the two methods. The subtype independent method tends to select a subset of those present in dependent method.

To investigate whether subtype dependent method can obtain a gene subset that is specific to a tumor subtype and contains common genes for all tumor subtypes, we look into the selected genes and use Venn diagram to represent their relationship. Due to the graphical representation power of Venn diagram, we show these kinds of relationships with an example of three tumor subtypes and choose Leukemia2 and 14_Tumor for illustration. Analysis on other datasets can be conducted in a similar way. Accordingly, Fig. 4 presents the relations between the genes related to the Leukemia2 tumor subtypes. From the left one, we can see that subtype dependent method obtains a unique gene subset for each tumor subtype, since the intersection between two of them is very small. From the right one, we can observe that about 33% (31/94) genes selected by subtype independent method intersect with the genes for class1 obtained by subtype dependent method, and that 23% (22/94) of the genes intersect with the genes for class2. This indicates that subtype independent method also has the capacity to select a subset of genes that are related to each tumor subtype. However, subtype independent method fails to locate the tumor subtype specific genes. Fig. 5 shows a similar diagram from 14_Tumor. Due to the large number of tumor subtypes, we randomly choose class24, class25 and class26 as an illustration. Specifically, we can see that subtype dependent method can obtain a unique gene subset for each tumor subtype, because the intersection between two of them is empty. We can also observe that ten genes selected by subtype independent method intersect with the genes for class24 and that three genes intersect with the genes for class25.

4.2. Performance of tumor subtype classification

Tables 9–14 show the experimental results regarding classification performance for each of the experimental datasets, respectively. A higher value indicates better quality of the selected genes in tumor subtype classification in each of the four metrics used

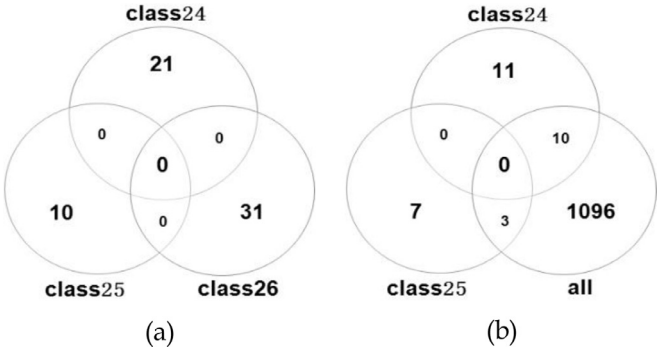

Fig. 5. Relations between genes on 14_Tumor dataset. Class24, class25 and class26 represent genes selected by subtype dependent method for each tumor subtype, and all represents genes selected by subtype independent method for all subtypes.

Table 9
Performance of tumor subtype classification using different gene selection methods on Leukemia2.

Methods		Accu	Prec	Rec	F1
No feature selection		0.8889	0.8979	0.8901	0.8940
Fisher	Independent	0.9722	0.9718	0.9694	0.9706
	Dependent	**0.9861**	0.9885	0.9833	**0.9859**
	Fusion	**0.9861**	0.9885	0.9833	**0.9859**
mRMR	Independent	0.9583	0.9558	0.9528	0.9543
	Dependent	**1.0000**	1.0000	1.0000	**1.0000**
	Fusion	**1.0000**	1.0000	1.0000	**1.0000**
FCBF	Independent	**1.0000**	1.0000	1.0000	**1.0000**
	Dependent	**1.0000**	1.0000	1.0000	**1.0000**
	Fusion	0.9861	0.9885	0.9833	0.9859

Table 10
Performance of tumor subtype classification using different gene selection methods on Brain2.

Methods		Accu	Prec	Rec	F1
No feature selection		0.7576	0.7562	0.7536	0.7549
Fisher	Independent	0.7586	0.7825	0.7941	0.7882
	Dependent	**0.8799**	0.8964	0.8964	**0.8964**
	Fusion	0.8603	0.8798	0.8786	0.8792
mRMR	Independent	0.8578	0.8600	0.8405	0.8501
	Dependent	**0.9596**	0.9677	0.9464	**0.9569**
	Fusion	**0.9596**	0.9677	0.9464	**0.9569**
FCBF	Independent	0.9203	0.9335	0.9107	0.9220
	Dependent	**0.9804**	0.9844	0.9821	**0.9833**
	Fusion	0.8995	0.9196	0.8929	0.9061

in this study. The best values achieved by the three types of feature selection methods in terms of accuracy and F1 score are highlighted in bold. For comparison, the second row "No feature selection" presents the classification results without using feature selection.

We observe in Tables 9–14, that either mRMR or FCBF as the baseline feature selector tends to obtain better classification performance versus that baseline selector from the Fisher score. For example, on Brain2 dataset, subtype dependent method with Fisher score obtains an accuracy of 0.8799, which is less than 0.9596 of mRMR and 0.9804 of FCBF. This is largely because that in contrast to mRMR and FCBF, the Fisher score is a single feature ranking method that evaluates the quality of candidate features independently and does not consider the interaction between features. This indicates that it is preferred to integrate an effective feature selector into the two frameworks to get high-quality gene subsets. Unexpectedly, subtype independent method fails to improve the classification performance especially when working on a dataset with a large number of classes such as 11_Tumor and

Table 11
Performance of tumor subtype classification using different gene selection methods on Brain1.

Methods		Accu	Prec	Rec	F1
No feature selection		0.8877	0.7565	0.6900	0.7217
Fisher	Independent	0.8551	0.6941	0.7133	0.7036
	Dependent	**0.8995**	0.7255	0.7100	0.7176
	Fusion	0.8888	0.7928	0.7533	**0.7725**
mRMR	Independent	0.8988	0.8639	0.800	0.8307
	Dependent	**0.9552**	0.9875	0.8500	**0.9136**
	Fusion	0.9322	0.9021	0.8267	0.8628
FCBF	Independent	0.9437	0.9846	0.8167	**0.8928**
	Dependent	**0.9444**	0.9693	0.8167	0.8865
	Fusion	0.9437	0.9846	0.8167	**0.8928**

Table 12
Performance of tumor subtype classification using different gene selection methods on 9_Tumor.

Methods		Accu	Prec	Rec	F1
No feature selection		0.5851	0.5369	0.5104	0.5233
Fisher	Independent	0.6310	0.5616	0.5840	0.5726
	Dependent	**0.9165**	0.9388	0.8838	**0.9105**
	Fusion	0.8121	0.8516	0.7908	0.8201
mRMR	Independent	0.8163	0.7490	0.7496	0.7493
	Dependent	**0.9675**	0.9778	0.9286	**0.9525**
	Fusion	0.7838	0.8402	0.7542	0.7949
FCBF	Independent	0.8005	0.7213	0.7264	0.7238
	Dependent	**0.9666**	0.9730	0.9722	**0.9726**
	Fusion	0.7687	0.7476	0.6894	0.7173

Table 13
Performance of tumor subtype classification using different gene selection methods on 11_Tumor.

Methods		Accu	Prec	Rec	F1
No feature selection		0.8969	0.9091	0.8390	0.8727
Fisher	Independent	0.7884	0.8036	0.7574	0.7799
	Dependent	**0.9484**	0.9502	0.9172	**0.9334**
	Fusion	0.9382	0.9422	0.9045	0.9230
mRMR	Independent	0.9317	0.9386	0.8918	0.9146
	Dependent	**0.9721**	0.9697	0.9551	**0.9623**
	Fusion	0.9498	0.9543	0.9156	0.9345
FCBF	Independent	0.9542	0.9469	0.9313	0.9390
	Dependent	**0.9944**	0.9899	0.9924	**0.9912**
	Fusion	0.9714	0.9674	0.9533	0.9603

Table 14
Performance of tumor subtype classification using different gene selection methods on 14_Tumor.

Methods		Accu	Prec	Rec	F1
No feature selection		0.6006	0.6353	0.5424	0.5852
Fisher	Independent	0.4381	0.4159	0.3618	0.3870
	Dependent	**0.6755**	0.7664	0.6448	**0.7003**
	Fusion	0.6726	0.6805	0.6400	0.6596
mRMR	Independent	0.5483	0.5301	0.5183	0.5241
	Dependent	**0.7666**	0.8079	0.7342	**0.7693**
	Fusion	0.7178	0.7926	0.6768	0.7301
FCBF	Independent	0.6465	0.7028	0.5947	0.6442
	Dependent	0.7011	0.7501	0.6560	0.6999
	Fusion	**0.7015**	0.7749	0.6643	**0.7153**

14_Tumor, whereas our proposed process has been shown to have higher performance. For example, on 11_Tumor, subtype independent method with Fisher score only obtains an accuracy of 0.7884, which is less than 0.8969 obtained without using feature selection. For 14_Tumor, subtype independent method with mRMR obtains an accuracy of 0.5483, in comparison with 0.6006 obtained using all the features.

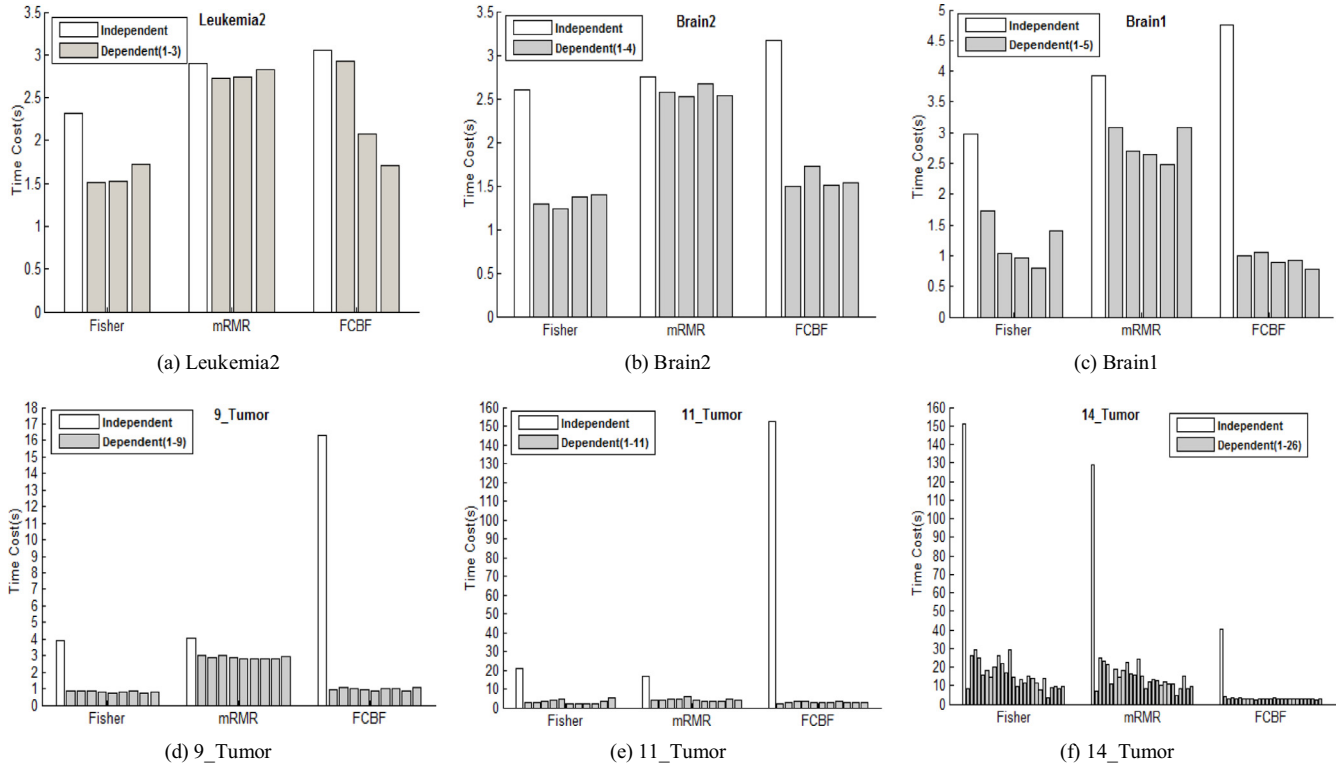

Fig. 6. Time cost comparisons between subtype-independent and subtype-dependent methods in obtaining the optimal gene subset.

We also observe that whichever baseline feature selector is used, subtype dependent method consistently obtains better classification performance than that of the subtype independent method, but this is not the case for the fusion based method. For example, when using mRMR, fusion based method outperforms subtype independent method on Brain1 dataset in terms of accuracy (0.9322 vs. 0.8988) and F1 (0.8628 vs. 0.8307), but on 9_Tumor dataset, fusion based method obtains an accuracy of 0.7838, which is lower than 0.8163 of subtype independent method.

Furthermore, we can see that subtype dependent method outperforms fusion based method in the majority of cases except that the one using FCBF on 14_Tumor dataset. But, in this case, the difference between them is quite small as subtype dependent method has an accuracy of 0.7011, which is only 0.0004 smaller than that of fusion based method. This demonstrates the overall superiority of the proposed subtype dependent feature selection method in achieving better classification performance.

4.3. Running time cost comparison

In the previous section, we presented data that the proposed subtype dependent method has a better classification performance. In this section, we investigate the three types of gene selection methods in terms of the time cost to obtain the optimal gene subset and time cost in classification. All experiments are conducted on a desktop with a Quad-core Intel Core i5 CPU (3.2 GHz processor and 4G RAM).

4.3.1. Time cost in obtaining the optimal gene subset

Due to the independence of the training set in selecting genes that are specific to tumor subtypes, subtype dependent method can work in a parallel way. Thus, we present the time costs for each tumor subtype rather than show them as a whole to compare the time costs. For a C-class gene selection problem, we obtain C time costs, each of which is associated with a tumor subtype. Since the genes selected by fusion based method is the union of the gene subsets that are obtained with subtype dependent method, the time cost is then the maximum of C time costs, and we do not take it into account. Fig. 6 presents the time cost comparison between subtype independent method and subtype dependent method on the six microarray datasets. In each subfigure, "Independent" indicates subtype independent method, and "Dependent(1-N)" indicates subtype dependent method. Specifically, the first bar in this category represents the time cost associated with class1, and the last bar is for class$N$. N is the number of classes of a dataset.

From Fig. 6, we can observe that no matter which baseline feature selector is adopted, subtype independent method is consistently much more time consuming than subtype dependent method, which demonstrates the superiority of subtype dependent method in reducing time costs. Particularly, the difference between them in time cost increases with the number of classes, and the time cost in selecting the subtype specific genes is very small. For example, on Leukemia2, it takes subtype independent method with Fisher score 2.31 seconds, and correspondingly it takes subtype dependent method 1.51, 1.53 and 1.72 seconds for each of the tumor subtypes, respectively. On 14_Tumor, however, with the same parameter setting as that of Leukemia2 dataset, it takes subtype independent method 151.33 seconds, but the maximum execution time of subtype dependent method is 29.28 seconds. The main reason to the above observations is that subtype independent method finds the optimal gene subset on a multi-class dataset, it involves more computations in evaluating the quality of candidate features. Also, for subtype dependent method, the training set size is the same for all the tumor subtypes, so the time costs do not vary much within dependent methods regardless of N.

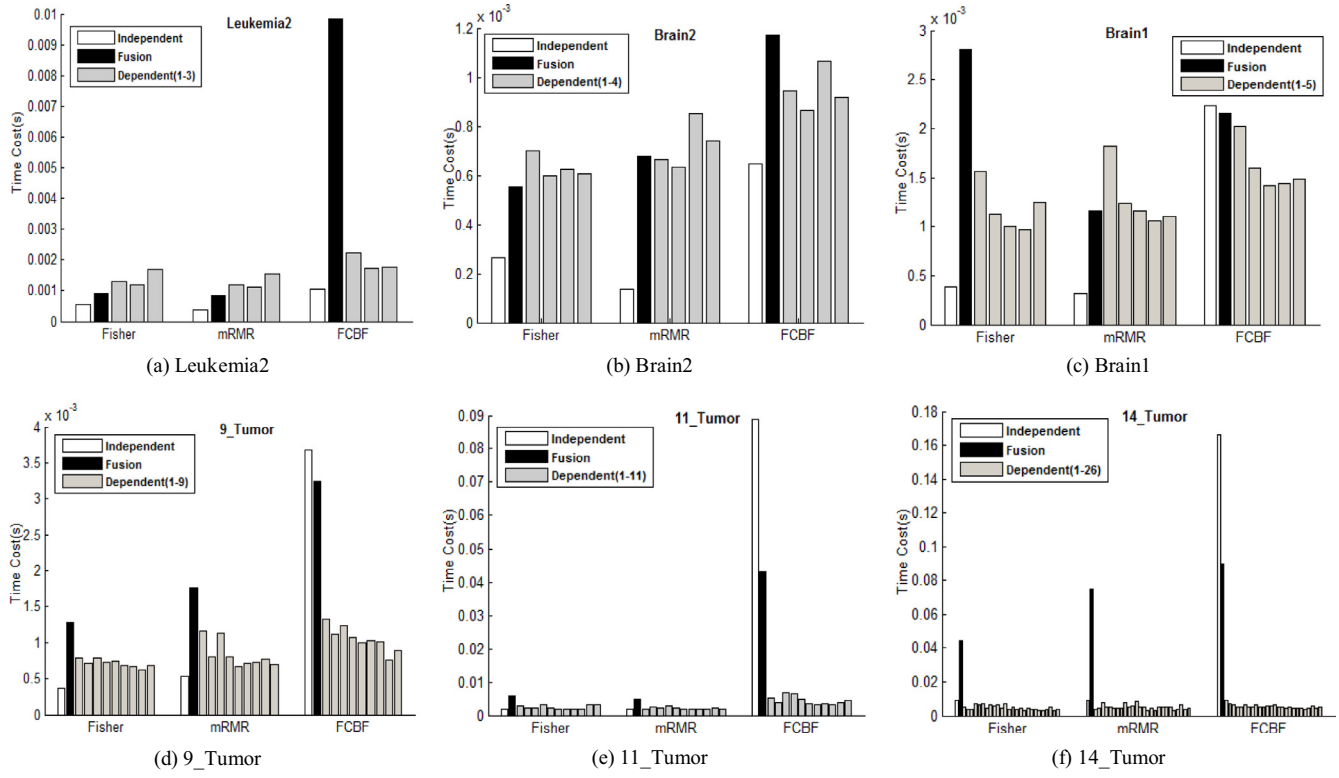

Fig. 7. Time cost comparisons of subtype-independent, subtype-dependent, and fusion based gene selection methods in performing tumor subtype classification.

4.3.2. Time cost in tumor subtype classification

In this subsection, we present the time costs of the three types of methods in tumor subtype classification. Because subtype dependent method can work in parallel, we present the time costs associated with each tumor subtype. For a *C*-class classification problem, we obtain *C* time costs, each of which is related to a two-class classification problem. Besides, we present the time costs of fusion based method. Fig. 7 presents the time cost comparison on the experimental datasets. In each subfigure, "Independent" indicates subtype independent method, "Fusion" means fusion based method, and "Dependent(1-*N*)" indicates subtype dependent method. The first bar in this category represents the time cost associated with class1, and the last bar represents the time cost for class*N*.

From Fig. 7, we can see that the time costs of three types of methods vary with the baseline feature selectors. When using Fisher score or mRMR, the time costs taken by subtype independent method are comparable to that of subtype dependent method. But if using FCBF, in dealing with datasets with a large number of classes, such as 9_Tumor, 11_Tumor, and 14_Tumor, subtype dependent method greatly benefits from the proposed scheme and is much more efficient than subtype independent method. In addition, we can observe that whichever baseline feature selector is used, the fusion based method generally has a higher time cost than subtype dependent method. Also, the difference in time costs between subtype dependent method and fusion based method increases with the number of classes. These results and analysis demonstrate the efficiency of the subtype dependent method.

4.4. Exploration of other classifiers

As we discussed in previous sections, the two proposed feature selection frameworks, including subtype dependent feature selection and fusion based feature selection, are not designed for a specific classifier and any classifier can be integrated into them. Furthermore, in addition to SVM, we preliminarily investigate two other commonly used classification models, Naïve bayes (NB) and *k*-nearest-neighbor (KNN), and compare their performance to that of SVM.

Both NB and KNN basically take a similar procedure as SVM in selecting subtype-dependent features and in training a classification model, while they take a slightly different way to deal with voting conflict in classifying test samples. In the case of SVM, we use formula (14) to estimate the confidence that a test sample belongs to a certain class.

For NB, it outputs the posteriori probability for determining the label l of an unseen sample x using formula (21). Accordingly, we can directly use the posteriori probability to estimate the confidence for solving voting conflict.

$$p(l|x) = \frac{p(l)p(x|l)}{\sum_i p(l_i)p(x|l_i)} \tag{21}$$

For KNN, it calculates the distance between a test sample and every training sample to assign test samples labels. To measure the assignment confidence of a test sample x, we adopt the following steps to get a posteriori probability. For x and a training set Tr, let k be the number of nearest neighbors used in KNN, $nbhd(Tr, x)$ be the k nearest neighbors to x in Tr, $Y(nbhd)$ be the labels of the points in $nbhd(Tr, x)$, $prior$ be the priors of the classes in Tr, $W(nbhd)$ be the weights of the points in $nbhd(Tr, x)$ and being normalized to sum to the priors, we calculate the posteriori probability $p(l|x)$ that x is assigned to class l using formula (22).

$$p(l|x) = \frac{\sum_{i \in nbhd(x)} W(i)\mathrm{I}\{Y(i) = l\}}{\sum_{i \in nbhd(x)} W(i)} \tag{22}$$

in which $\mathrm{I}\{a = b\}$ is an indicator function and equals 1 if a equals b. Afterwards, we use the posteriori probability to estimate the

Table 15
Performance of tumor subtype classification using different classifiers and gene selection methods on Leukemia2.

Metrics		Accu			Prec			Rec			F1		
Methods		SVM	NB	KNN	SVM	NB	KNN	SVM	NB	KNN	SVM	NB	KNN
No feature selection		0.9728	0.9728	0.8627	0.9714	0.9055	0.8627	0.9714	0.9714	0.8615	0.9714	0.9714	0.8636
Fisher	Independent	0.9722	0.7347	0.7230	0.9718	0.9697	0.7178	0.9694	0.7087	0.7139	0.9706	0.7138	0.7158
	Dependent	**1.0000**	0.9021	0.7648	1.0000	0.8595	0.8363	1.0000	0.8937	0.7306	**1.0000**	0.8995	0.7799
	Fusion	1.0000	0.8610	0.8604	1.0000	0.7189	0.8555	1.0000	0.8559	0.8552	1.0000	0.8577	0.8554
mRMR	Independent	0.9461	0.8220	0.7103	0.9394	0.9841	0.6927	0.9417	0.8123	0.6905	0.9405	0.8196	0.6916
	Dependent	**1.0000**	0.9733	0.8488	1.0000	0.9444	0.8667	1.0000	0.9762	0.8290	**1.0000**	0.9729	0.8474
	Fusion	1.0000	0.9450	0.8615	1.0000	0.8270	0.8671	1.0000	0.9524	0.8587	1.0000	0.9484	0.8629
FCBF	Independent	1.0000	0.9728	0.9867	1.0000	0.9055	0.9841	1.0000	0.9714	0.9861	1.0000	0.9714	0.9851
	Dependent	**1.0000**	0.9867	0.9855	1.0000	0.9841	0.9885	1.0000	0.9881	0.9833	**1.0000**	0.9861	0.9859
	Fusion	1.0000	0.9867	1.0000	1.0000	0.9714	1.0000	1.0000	0.9881	1.0000	1.0000	0.9861	1.0000

Table 16
Performance of tumor subtype classification using different classifiers and gene selection methods on Brain2.

Metrics		Accu			Prec			Rec			F1		
Methods		SVM	NB	KNN	SVM	NB	KNN	SVM	NB	KNN	SVM	NB	KNN
No feature selection		0.8186	0.7194	0.6177	0.8559	0.8067	0.6677	0.7869	0.6464	0.5940	0.8199	0.7177	0.6287
Fisher	Independent	0.7390	0.6422	0.5772	0.7579	0.6292	0.5613	0.7357	0.6298	0.5619	0.7466	0.6295	0.5616
	Dependent	0.9387	0.8003	0.6985	0.9488	0.8039	0.7373	0.9488	0.8095	0.6714	0.9488	0.8067	0.7028
	Fusion	**0.9583**	0.8003	0.8603	0.9706	0.7942	0.8983	0.9643	0.7905	0.8631	**0.9674**	0.7923	0.8804
mRMR	Independent	0.8419	0.6422	0.5233	0.8837	0.5994	0.4873	0.7691	0.5952	0.4917	0.8224	0.5973	0.4895
	Dependent	**1.0000**	0.8787	0.5625	1.0000	0.8803	0.6250	1.0000	0.8964	0.5107	**1.0000**	0.8883	0.5621
	Fusion	0.9608	0.8787	0.8186	0.9677	0.8774	0.8452	0.9464	0.8774	0.8262	0.9569	0.8774	0.8356
FCBF	Independent	0.9596	0.8591	0.8799	0.9665	0.8841	0.9041	0.9643	0.8417	0.8774	0.9654	0.8624	0.8905
	Dependent	**0.9596**	0.8419	0.8407	0.9655	0.8625	0.8938	0.9655	0.8048	0.8286	**0.9655**	0.8326	0.8599
	Fusion	0.9400	0.8419	0.6765	0.9514	0.8502	0.7903	0.9464	0.8048	0.6476	0.9489	0.8269	0.7119

Table 17
Performance of tumor subtype classification using different classifiers and gene selection methods on 11_Tumor.

Metrics		Accu			Prec			Rec			F1		
Methods		SVM	NB	KNN	SVM	NB	KNN	SVM	NB	KNN	SVM	NB	KNN
No feature selection		0.8736	0.8674	0.7417	0.8917	0.8751	0.7349	0.8127	0.8168	0.6630	0.8504	0.8449	0.6971
Fisher	Independent	0.7994	0.3107	0.2242	0.7941	0.1449	0.1494	0.7679	0.1911	0.1502	0.7808	0.1648	0.1498
	Dependent	0.9309	0.9025	0.6839	0.9510	0.8872	0.7995	0.8910	0.8965	0.6439	0.9200	0.8918	0.7133
	Fusion	**0.9428**	0.8506	0.8677	0.9583	0.8320	0.8517	0.9186	0.8272	0.8483	**0.9380**	0.8296	0.8500
mRMR	Independent	0.8968	0.3327	0.2248	0.8880	0.1833	0.1757	0.8547	0.2291	0.1763	0.8710	0.2036	0.1760
	Dependent	0.9598	0.9081	0.6835	0.9654	0.9079	0.8025	0.9270	0.8935	0.6242	0.9458	0.9007	0.7022
	Fusion	**0.9600**	0.8735	0.8564	0.9660	0.8620	0.8462	0.9452	0.8542	0.8281	**0.9555**	0.8581	0.8370
FCBF	Independent	0.9544	0.9023	0.8909	0.9463	0.9247	0.9061	0.9272	0.8195	0.8738	0.9367	0.8689	0.8896
	Dependent	**0.9942**	0.9540	0.8905	0.9899	0.9709	0.9497	0.9924	0.9275	0.8690	**0.9912**	0.9487	0.9076
	Fusion	0.9715	0.9252	0.9199	0.9628	0.9402	0.9043	0.9582	0.8983	0.9021	0.9605	0.9188	0.9032

confidence that x belongs to a certain class and use one-nearest-neighbor (1NN) in this study.

To evaluate the performance of NB and 1NN, we conducted experiments on three representative datasets with the same experimental setup as we adopted above. Tables 15–17 show the experimental results regarding classification performance for Leukemia2, Brain2, and 11_Tumor, respectively. A higher value indicates better quality of the selected genes in tumor subtype classification in terms of each of the four metrics used in this study. The best values achieved by the three types of feature selection methods and three classification models in terms of accuracy and F1 are highlighted in bold. For comparison, the third row "No feature selection" presents the classification results without using feature selection. We can observe the following results. (1) In comparison to subtype independent methods, subtype dependent biomarker identification helps obtain higher classification performance whatever SVM, NB, or KNN is used, which demonstrates the effectiveness of the two proposed frameworks. For example, for 11_Tumor dataset in the case of using fisher score, NB only achieves 0.1648 F1 in subtype independent case, while subtype dependent method increases it to 0.8918 and fusion based method an F1 of 0.8296. For KNN, subtype dependent method and fusion based method increase the 0.1498 F1 of subtype independent method to 0.7133 and 0.8500, respectively. (2) In comparison with NB and KNN, integrating SVM into the framework is a better choice towards better tumor classification performance. For example, for 11_Tumor dataset in the case of fisher score with the framework of subtype dependent method, SVM obtains an F1 of 0.9380, which outperforms both NB with an F1 of 0.8296 and KNN with an F1 of 0.8500; with the framework of fusion based method, SVM obtains an F1 of 0.8710, while NB has an F1 of 0.2036 and KNN has an F1 of 0.1760. This indicates the superiority of SVM over NB and KNN in classifying gene expression profiles of high-dimensionality and small sample sizes.

5. Conclusion

Tumor progression is a social and economic problem that affects the life quality of a large number of individuals, thus accurately distinguishing tumor subtypes contributes to the bet-

ter management, treatment, and outcomes. Microarray technology provides us a way to identify disease genes and classify tumor subtypes, but the intrinsic nature of microarray data characterized by high dimensionality and small sample sizes limits their capacity. Correspondingly, researchers have put forward a wealth of feature selection methods. However, most of them seek to find a common subset of genes for all tumor subtypes within a pathological derived context that may be unable to reflect the unique gene profiles of each molecular subtype specific for personalized targeting. Therefore, in this study, we first propose a framework, called subtype dependent method that selects gene subsets for each tumor subtype, combined with another framework, named fusion based method that merges the outputs of subtype dependent method into a single gene subset. In addition, we give a corresponding classification model, including classifier training and testing schemes. We then detailed how to obtain the optimal feature subset for ranking based as well as subset based feature selection methods in this study, and detailed how to estimate the confidence that a sample belongs to a specific class to solve the problem of voting conflict.

To evaluate the performance of the two proposed methods in gene selection and tumor subtype classification, under each of the two proposed frameworks, we implement three specific gene selection algorithms with Fisher score, mRMR, and FCBF as the building blocks, respectively, and use three different classification models with different metrics (support vector machine, Naïve bayes, and *k*-nearest-neighbor) as the learning algorithm in feature selection and classifier construction. Finally, we conducted extensive experiments on six publicly available microarray datasets in terms of the number of selected genes, classification performance, and the time costs in gene selection and tumor subtype classification. Experimental results show that in comparison with subtype independent method, our subtype dependent method selects a subset of genes with a smaller size that is specific for each tumor subtype. Besides, compared with NB and KNN, integrating SVM into the framework is a better choice towards better tumor classification performance. This helps design personalized treatment plans, also accelerates drug discovery and aids drug design. Furthermore, the subtype dependent method outperforms the subtype independent and fusion based methods in terms of accuracy, precision, recall, and F1 score, particularly for the microarray dataset with a large number of categories, such as 14_Tumor with 26 classes. Also, the experimental results show that a powerful baseline feature selector is preferred to be used in the proposed frameworks. Additionally, time cost comparisons demonstrate the efficiency of subtype dependent method in gene selection as well as subtype identification.

In the future, we plan to follow up with three lines of questions. First, although we tested the effectiveness of the proposed method in classifying tumor subtypes, it is actually a general framework that can be applied to other situations such as protein structure and function prediction. Second, we will explore other feature selectors and classification models as the building blocks of the proposed frameworks and analyze corresponding results. Third, in recent years, deep learning models have gained great popularity due to their powerful feature representation ability and they have been successfully applied in a variety of fields such as drug-drug interaction identification [46], protein-protein interaction detection [47], and protein structure prediction [48]. Though working well, methods built on deep learning models generally suffer from the difficulty of interpretation of features. Consequently, it is often difficult for researchers to understand the obtained features, which greatly limits their further use in identifying drug targets and locating disease genes. In contrast, the two proposed methods provide a way of alleviating this problem, and exploring their power remains another research topic.

Acknowledgment

This work was partially supported by the China Postdoctoral Science Foundation (No. 2016M592046), the National Natural Science Foundation of China (No. 71661167004), the Fundamental Research Funds for the Central Universities(No. JZ2016HGBH1053), the “111 Project” of the Ministry of Education and State Administration of Foreign Experts Affairs(Grant No. B14025), and the Science and Technology Innovation Project of Foshan City, China(Grant No. 2015IT100095).

References

[1] T.R. Golub, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (5439) (1999) 531–537.
[2] J. Welsh, et al., Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer, Cancer Res. 61 (16) (2001) 5974–5978.
[3] A. Wang, N. An, J. Yang, G. Chen, L. Li, G. Alterovitz, Wrapper-based gene selection with Markov blanket, Comput. Biol. Med. 81 (2017) 11–23.
[4] D. Singh, et al., Gene expression correlates of clinical prostate cancer behavior, Cancer Cell 1 (2) (2002) 203–209.
[5] M. Wu, D. Dai, Y. Shi, H. Yan, X. Zhang, Biomarker identification and cancer classification based on microarray data using Laplace Naive Bayes model with mean shrinkage, IEEE/ACM Trans. Comput. Biol. Bioinform. 9 (6) (Nov./Dec. 2012) 1649–1662.
[6] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (1-3) (2002) 389–422.
[7] L. Ein-Dor, O. Zuk, E. Domany, Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer, *Proc. Natl. Acad. Sci.* 103 (15) (2006) 5923–5928.
[8] W. Li, Y. Yang, How many genes are needed for a discriminant microarray data analysis?, in: S.M Lin, K.F. Johnson (Eds.) Methods of Microarray Data Analysis, Kluwer Academic, 2002, pp. 137–150.
[9] G. Piatetsky-Shapiro, P. Tamayo, Microarray data mining: facing the challenges, Sigkdd Explorations 5 (2003) 1–5.
[10] S. Bandyopadhyay, S. Mallik, A. Mukhopadhyay, A survey and comparative study of statistical tests for identifying differential expressionfrom microarray data, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 11 (1) (Jan./Feb. 2014) 95–115.
[11] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, Y. Saeys, Robust biomarker identification for cancer diagnosis with ensemble feature selection methods, Bioinformatics 26 (3) (2010) 392–398.
[12] A. Wang, N. An, G. Chen, L. Li, G. Alterovitz, Accelerating wrapper-based feature selection with K-nearest-neighbor, Knowl.-Based Syst. 83 (2015) 81–91.
[13] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[14] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (4) (2005) 491–502.
[15] S. Rathore, M. Hussain, A. Khan, GECC: gene expression based ensemble classification of colon samples, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 11 (6) (Jan./Feb. 2014) 1131–1145.
[16] I. Inza, P. Larranaga, R. Blanco, A. Cerrolaza, Filter versus wrapper gene selection approaches in DNA microarray domains, Artif. Intell. Med. 31 (2) (2004) 91–103.
[17] R. Ruiz, J.C. Riquelme, J.S. Aguilar-Ruiz, Incremental wrapper-based gene selection from microarray data for cancer classification, *Pattern Recognit.* 39 (12) (2006) 2383–2392.
[18] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, A. Nowe, A survey on filter techniques for feature selection in gene expression microarray analysis, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 9 (4) (July/Aug. 2012) 1106–1119.
[19] M. Robnik-Sikonja, I. Kononenko, Theoretical and empirical analysis of reliefF and rrelieff, Mach. Learn. 53 (1–2) (2003) 23–69.
[20] M. Dash, H. Liu, Consistency-based search in feature selection, Artif. Intell. 151 (1) (2003) 155–176.
[21] Z. Zhao, L. Wang, H. Liu, J. Ye, On similarity preserving feature selection, IEEE Trans. Knowl. Data Eng. 25 (3) (2013) 619–632.
[22] A. Boulesteix, K. Strimmer, Partial least squares: a versatile tool for the analysis of high-dimensional genomic data, *Brief. Bioinf.* 8 (1) (2007) 32–44.
[23] G. Brown, A. Pocock, M. Zhao, M. Lujan, Conditional likelihood maximisation: a unifying framework for information theoretic feature selection, J. Mach. Learn. Res. 13 (2012) 27–66.
[24] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[25] L. Yu, H. Liu, Feature selection for high-dimensional data: a fast correlation-based filter solution, in: Proc. 20th Int'l Conf. Machine Learning, 2003, pp. 856–863.
[26] R. Kohavi, G. John, Wrappers for feature subset selection, Artif. Intell. 97 (0) (1997) 273–324.
[27] S. Maldonado, R. Weber, A wrapper method for feature selection using support vector machines, Inf. Sci. 179 (13) (2009) 2208–2217.

[28] L. Li, C.R. Weinberg, T.A. Darden, L.G. Pedersen, Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method, Bioinformatics 17 (12) (2001) 1131–1142.
[29] J. Huang, H. Fang, X. Fan, Decision forest for classification of gene expression data, Comput. Biol. Med. 40 (8) (2010) 698–704.
[30] F. Nie, H. Huang, X. Cai, C. Ding, Efficient and robust feature selection via joint 2, 1-norms minimization, in: Proc. Adv. Neural Inf. Process. Syst., 2010, pp. 1813–1821.
[31] E. Huerta, A. Hernández-Montiel, R. Morales-Caporal, M. Arjona López, Hybrid framework using multiple-filters and an embedded approach for an efficient selection and classification of microarray data, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 13 (1) (Jan./Feb. 2016) 12–26.
[32] S. Shreem, S. Abdullah, M. Nazri, Hybridising harmony search with a Markov blanket for gene selection problems, Inf. Sci. 258 (2014) 108–121.
[33] A. El Akadi, A. Amine, A. El Ouardighi, D. Aboutajdine, A two-stage gene selection scheme utilizing MRMR filter and GA wrapper, Knowl. Inf. Syst. 26 (3) (2011) 487–500.
[34] A. Sharma, S. Imoto, S. Miyano, A top-r feature selection algorithm for microarray gene expression data, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 9 (3) (2012) 754–764.
[35] F. Yang, K.Z. Mao, Robust feature selection for microarray data based on multicriterion fusion, *IEEE/ACM Trans. Comput. Biol. Bioinf.* 8 (4) (July/Aug. 2011) 1080–1092.
[36] X. Zhou, D.P. Tuck, MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data, Bioinformatics 23 (9) (2007) 1106–1114.
[37] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[38] A. Wang, N. An, G. Chen, L. Li, G. Alterovitz, Improving PLS–RFE based gene selection for microarray data classification, Comput. Biol. Med. 62 (2015) 14–24.
[39] L. Eindor, I. Kela, G. Getz, D. Givol, E. Domany, Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 21 (2) (2005) 171–178.
[40] G. de Lannoy, Gaël, D. François, M. Verleysen, Class-specific feature selection for one-against-all multiclass SVMs, in: Proceedings of European Symposium on Artificial Neural Networks, 2011, pp. 263–268.
[41] N. Zhou, L. Wang, Processing bio-medical data with class-dependent feature selection, in: Proceedings of Advances in Neural Networks, 2016, pp. 303–310.
[42] B. Pineda-Bautista, J. Ariel Carrasco-Ochoa, J. Martínez-Trinidad, General framework for class-specific feature selection, Expert Syst. Appl. 38 (8) (2011) 10018–10024.
[43] A. Statnikov, C. Aliferis, I. Tsamardinos, D. Hardin, S. Levy, A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis, Bioinformatics 21 (5) (2005) 631–643.
[44] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27.
[45] A. Wang, N. An, G. Chen, L. Li, G. Alterovitz, Predicting hypertension without measurement: a non-invasive, questionnaire-based approach, Expert Syst. Appl. 21 (42) (2015) 7601–7609.
[46] Wang Y, S. Liu, M. Rastegar-Mojarad, L. Wang, F. Shen, F. Liu, H. Liu, Dependency and AMR embeddings for drug-drug interaction extraction from biomedical literature, in: Proc. 8th ACM Int'l Conf. Bioinformat. Comput. Biol. Health Inform., 2017, pp. 36–43.
[47] H. Zhang, M. Yang, X. Feng, W. Yang, W. Tong, R. Guan, Protein-protein interaction extraction using attention-based convolution neural networks, in: Proc. 8th ACM Int'l Conf. Bioinformat. Comput. Biol. Health Inform., 2017, pp. 770–771.
[48] R. Heffernan, Y. Yang, K. Paliwal, Y. Zhou, Capturing non-local interactions by long short term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers, and solvent accessibility, Bioinformatics 33 (2017) 2842–2849.