Computers in Biology and Medicine xxx (xxxx) xxx

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

A new feature selection algorithm based on relevance, redundancy and complementarity

Chao Li, Xiao Luo, Yanpeng Qi, Zhenbo Gao, Xiaohui Lin *

School of Computer Science & Technology, Dalian University of Technology, 116024, Dalian, China

ARTICLE INFO

Keywords:
Biological data analysis
Feature selection
Feature relevance
Feature redundancy
Feature complementarity

ABSTRACT

Defining important information from biological data is critical for the study of disease diagnosis, drug efficacy and individualized treatment. Hence, the feature selection technique is widely applied. Many feature selection methods measure features based on relevance, redundancy and complementarity. Feature complementarity means that two features' cooperation can provide more information than the simple summation of their individual information. In this paper, we studied the feature selection technique and proposed a new feature selection algorithm based on relevance, redundancy and complementarity (FS-RRC). On selecting the feature subset, FS-RRC not only evaluates the feature relevance with the class label and the redundancy among the features but also evaluates the feature complementarity. If complementary features exist for a selected relevant feature, FS-RRC retains the feature with the largest complementarity to the selected feature subset. To show the performance of FS-RRC, it was compared with eleven efficient feature selection methods, MIFS, mRMR, CMIM, ReliefF, FCBF, PGVNS, MCRMCR, MCRMICR, RCDFS, SAFE and SVM-RFE on two synthetic datasets and fifteen public biological datasets. The experimental results showed the superiority of FS-RRC in accuracy, sensitivity, specificity, stability and time complexity. Hence, integrating feature individual discriminative ability, redundancy and complementarity can define more powerful feature subset for biological data analysis, and feature complementarity can help to study the biomedical phenomena more accurately.

1. Introduction

With the development of high-throughput technologies, large quantities of biological data have been produced. How to analyze the data and define the important information from the data has been one of the main focuses in biological studies. Feature extraction and feature selection are two efficient techniques for mining meaningful information from high-dimensional data. Feature extraction, such as principal component analysis and partial least squares-discriminant analysis, transforms the original feature space into a new low-dimensional space. Feature selection, such as ReliefF, Fisher Score and Lasso, reduces the original feature space into a low-dimensional feature subspace by directly removing the noise and noninterested input features [1]. Compared to feature extraction, feature selection reduces the dimensionality of the feature space without transformation and is assumed to be superior in terms of readability and biological interpretability. Hence feature selection is widely applied for analyzing the genomics data and metabolomics data to identify biomarkers for disease diagnosis, early warning of malignant tumors and disease prognosis [2,3].

Feature selection can be divided into unsupervised, semi-supervised and supervised. Unsupervised feature selection is used to explore biological data without the class label. It can provide an effective method for discovering the unknown meaningful results. Wolf et al. [4] proposed an unsupervised feature selection method, called Q - α, which defined the feature relevance by the Laplacian spectrum and ranked features based on a least squares optimization process. Semi-supervised feature selection studies both labeled data and unlabeled data and integrates labeled data into unlabeled data as additional information to improve the performance of the feature selection. Benabdeslem and Hindawi [5] proposed a semi-supervised feature selection method based on constraint selection and redundancy elimination. Supervised feature selection uses the labeled data to select the feature subset.

Usually, supervised feature selection algorithms can be classified into three categories, wrapper, embedded, and filter, depending on the relationship with the learning model [6]. Wrapper methods [7,8] use the learning models to evaluate the feature subset by the classification accuracy rates. Support vector machine-recursive feature elimination (SVM-RFE) [9] is one of the classical and efficient wrapper methods.

* Corresponding author.
E-mail address: datas@dlut.edu.cn (X. Lin).

https://doi.org/10.1016/j.compbiomed.2020.103667
Received 26 July 2019; Received in revised form 16 February 2020; Accepted 16 February 2020
Available online 19 February 2020

Embedded methods use learning models to guide feature selection. Regularization is an embedding technique that can perform continuous shrinkage and automatically select a feature subset [10,11]. Filter methods separate the learning model and feature selection and weigh features based on their characteristics [12]. They are fast and easily analyze the high-dimensional data. Some filters measure the redundancy among features in addition to feature importance, such as mutual information feature selection (MIFS) [13], minimal-redundancy -maximal-relevance (mRMR) [14], conditional mutual information maximization (CMIM) [15], minimum conditional relevance-minimum conditional redundancy (MCRMCR) and minimum conditional relevance-minimum intra-class redundancy (MCRMICR) [16], which aim to selecting the feature subset with the maximum relevance to the class label and minimum redundancy. Fast correlation-based filter (FCBF) [17] and predominant group-based variable neighborhood search (PGVNS) [18] also consider feature redundancy, they use the approximate Markov blanket to identify and remove redundant features. Feature relationships are very complex. There are synergism and complementarity between some features. Hence to define a powerful feature subset, some filters simultaneously take advantage of feature relevance, redundancy and complementarity. Redundancy complementariness dispersion-based feature selection (RCDFS) [19] not only considers the relevance between feature and class label and pairwise inter-correlation of features but also extends traditional redundancy analysis to redundancy-complementariness analysis. Self-adaptive feature evaluation (SAFE) [20] uses feature complementarity in the search process, which penalizes redundancy and rewards complementarity based on an adaptive cost function.

Disease progression is a complex process that is usually not influenced by individual molecules but by complex molecule interactions. Hence, neglecting the complementarity between features may lose some important information for studying the nature of biomedical problems. Defining interactive or complementary features is of great significance to understanding the disease progression and prevention, diagnosis, and treatment. In this paper, we studied the feature selection technique and proposed a new feature selection algorithm based on feature relevance, redundancy and complementarity (FS-RRC). For a selected feature f, FS-RRC computes the complementarity of each feature with f. If there exist features that are complementary with f, FS-RRC selects the one having the largest complementarity to the selected feature subset. By combining feature relevance, redundancy and complementarity, FS-RRC can define more important information from the complex biological data.

The rest of this paper is organized as follows. Section 2 first introduces some evaluation criteria about feature relevance, redundancy and complementarity, and then FS-RRC is proposed. Section 3 gives the experimental settings and descriptions of the datasets. Section 4 shows the experimental results of FS-RRC compared with eleven well-known feature selection techniques. Finally, we discuss and conclude this paper.

2. Methods

The biological system is complex, and molecules cooperate and relate to each other in the process of physiological and pathological changes. Hence, the complementarity between features may also contain some meaningful information, and considering feature complementarity in addition to relevance and redundancy may induce a better biological data analysis result [19].

2.1. Relevance, redundancy and complementarity

Relevance means that a feature contains the information of given features or class labels. Wolf et al. measured the relevance by the Laplacian spectrum [4]. Peluffo et al. used distances to explore relevance [5]. Mutual information (MI) and symmetrical uncertainty (SU) are two common methods that measure the relevance between two random variables based on information theory. Let X and Y be two random variables, then the MI between X and Y is defined as follows [6]:

$$I(X;Y)=\begin{cases} H(X)-H(X|Y) \\ H(Y)-H(Y|X) \\ H(X)+H(Y)-H(X,Y) \end{cases}, \tag{1}$$

where $H(X)$ and $H(Y)$ are the entropy of X and Y, $H(X, Y)$ is the joint entropy of X and Y, and $H(X|Y)$ represents the conditional entropy. $I(X; Y)$ measures the level of correlation between X and Y or how much information X contains about Y.

It can be seen that the range of MI is $[0, +\infty)$. When X and Y are statistically independent, MI is zero. Symmetrical uncertainty is the normalized mutual information, and its value is restricted to the range $[0, 1]$. The SU of variables X and Y is defined as follows [21]:

$$SU(X,Y)=\frac{2I(X;Y)}{H(X)+H(Y)}. \tag{2}$$

There are also some other relevance measurements, such as the Pearson correlation coefficient, the Spearman correlation coefficient and the maximal information coefficient [22,23].

Feature redundancy is regarded as the degree of dependency among features. If two features are redundant, they are nonindependent and contain similar information. Correlation, mutual information and symmetrical uncertainty of features are usually used to measure the redundancy between features. In addition, the approximate Markov blanket (AMB) can also be used to determine whether two features are redundant [17].

Approximate Markov blanket [17] *Let $F = \{f_1, f_2, ..., f_m\}$ be the feature set, and C be the class label set. For two relevant features f_i and $f_j \in F$ $(i \neq j)$, f_j forms an approximate Markov blanket for f_i iff $SU(f_j, C) \geq SU(f_i, C)$ and $SU(f_i, f_j) \geq SU(f_i, C)$.*

FCBF [17] is an efficient filter that focuses on feature relevance and redundancy. First, it selects the relevant features based on SU and sorts them in descending order based on their SUs. Then, it checks and removes the redundant features by AMB.

Complementarity between features is also called information synergy or interaction, which means that two features working together could provide more information than the sum of their individual information [24]. Even if two features are redundant, their combination may be more informative, and their cooperation may provide additional information in some cases. Therefore, neglecting the complementarity between features may lose meaningful information. Let $f_i, f_j \in F$ $(i \neq j)$, then $I(f_i; f_j| C)$ reflects the synergy of f_i and f_j with class C. If $I(f_i; f_j| C) - I(f_i; C) > 0$, then f_i and f_j are complementary, and the cooperation of feature f_i and f_j can provide more information about the class label C than feature f_i alone.

2.2. FS-RRC algorithm

Many feature selection algorithms are dedicated to searching the features having strong relevance with the class label and weak redundancy among them. While, some features may be complementary or interact with each other, their co-working could provide more information. Hence, considering feature complementarity may define a more discriminative feature subset.

FS-RRC combines the feature relevance, redundancy and complementarity to select the feature subset. For feature $f_i, f_j \in F$, the relevance between f_i and class label C is measured by symmetrical uncertainty $SU(f_i; C)$, the redundancy between f_i and f_j is evaluated by an approximate Markov blanket, and the complementarity between feature f_i and f_j is measured by

$$Score(f_i, f_j, C) = I(f_i; C|f_j) - I(f_i; C). \tag{3}$$

FS-RRC first selects the features that are strongly relevant to the class label and ranks all the relevant features based on their SUs in descending

order. Then, FS-RRC examines each feature f sequentially based on the current feature rank: (1) removes the features that are behind f in the current rank list and redundant with f by AMB, (2) computes the complementarity score of each feature with f. If there exist features that are complementary with f, then FS-RRC keeps the feature having the largest complementarity score to the selected feature subset. The detailed information of FS-RRC is shown in Algorithm.1. The flowchart of the FS-RRC algorithm is given in the supporting file (see Fig. S1).

Algorithm.1. FS-RRC algorithm

Input: Dataset D, feature set $F = \{f_1, f_2, \ldots, f_m\}$, class label set C.
Output: The selected feature subset S_{best}.
Initialization: $S_{list} = \phi$; $S_{comp} = \phi$; $S_{best} = \phi$;
For each $f \in F$ do
  Calculate $SU(f; C)$;
  If $(SU(f; C) > 0)$
    $S_{list} = S_{list} \cup \{f\}$;
  End if;
End for;
Sort $f \in S_{list}$ in descending order according to $SU(f; C)$;
$S^*_{list} = S_{list}$;
F = getFirstElement (S_{list});
Do begin
  f^* = getNextElement (S_{list}, f);
  Do begin
    If $(SU(f; f^*) \geq SU(f^*; C))$
      $S_{list} = S_{list} - \{f^*\}$;
    f^* = getNextElement (S_{list}, f^*);
  Until f^* == NULL;
  For (each $f^* \in S^*_{list}$) and ($f^* != f$) do
    Calculate $Score(f, f^*, C) = I(f; C \mid f^*) - I(f; C)$;
  $S_{comp} = S_{comp} \cup \{argmax_{f^*}\{I(f; f^*; C): I(f; f^*; C) > 0, f^* \in S^*_{list}, f^* != f\}\}$;
  f = getNextElement (S_{list}, f);
Until f == NULL;
$S_{best} = S_{list} \cup S_{comp}$;
Return S_{best};

Fig. 1 gives an example to illustrate how FS-RRC works. In Fig. 1, six features f_1, f_2, f_3, f_4, f_5 and f_6 in S_{list} are relevant to the class label, they are sorted in descending order by their relevance with the class label, f_1 has the largest relevance and f_6 has the smallest relevance. Hence, (1) f_1 is first selected. Then, f_2 and f_4 are removed from S_{list} because f_2 and f_4 are redundant with f_1 by AMB. Since f_5 and f_6 are the complementary features with f_1, and $0 < Score(f_1, f_5, C) < Score(f_1, f_6, C)$, f_6 is selected to set S_{comp}. (2) Now there are four features (f_1, f_3, f_5 and f_6) in S_{list}. Then f_3 is selected, and f_6 is removed from S_{list} because f_6 is redundant with f_3. f_3 and f_2 have complementarity ($Score(f_3, f_2, C) > 0$), thus, f_2 is put into the set S_{comp}. (3) Finally, f_5 is selected. According to FS-RRC, the final selected feature subset $S_{best} = S_{list} \cup S_{comp} = \{f_1, f_3, f_5\} \cup \{f_6, f_2\} = \{f_1, f_2, f_3, f_5, f_6\}$.

3. Experiments

In this section, FS-RRC is compared with eleven effective feature selection techniques, MIFS, mRMR, CMIM, ReliefF, FCBF, PGVNS, MCRMCR, MCRMICR, RCDFS, SAFE and SVM-RFE, on synthetic datasets and fifteen public real-world biological datasets to show its performance.

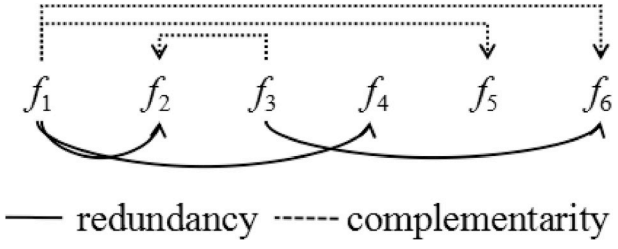

Fig. 1. A simple example to illustrate the workflow of FS-RRC.

3.1. Experimental settings

In the experiments, the feature elimination ratio r of SVM-RFE in each iteration was set as 0.05, parameter β of MIFS was set as 0.5, threshold δ of FCBF and PGVNS was set as 0, and the iteration number It_{max} and the neighborhood structure number k_{max} for PGVNS were set as 10 and 5, respectively. Since all the datasets contain features with continuous values, the minimum descriptive length (MDL) [25] was adopted for data discretization. MIFS, mRMR, CMIM, ReliefF, MCRMCR, MCRMICR ranked the features according to their weights, the top k (k = 1, 2, 3,..., $min\{100, |F|\}$) features from the feature ranking by each algorithm were selected. Then the highest classification accuracy rate as well as the sensitivity and specificity were reported. SVM, NB and kNN were used as the classifiers, respectively. The kernel function of SVM was set to a linear kernel function (SVM-Linear), polynomial kernel function (SVM-Poly) and radial basis function kernel function (SVM-RBF). The penalty factor was set to 1, the degree of SVM-Poly was set to 3, and the parameter k of kNN was set to 3. The implementation of SVM was from the library of support vector machines (LIBSVM) and was available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. FS-RRC, MIFS, mRMR, CMIM, ReliefF, FCBF, PGVNS and SVM-RFE were all implemented in C++. MCRMCR, MCRMICR, RCDFS and SAFE were implemented in MATLAB. Five-fold cross-validation was run fifty times for each method, and the average classification accuracy rate of the fifty runs was calculated. Taking FS-RRC as an example, the experimental scheme is shown in Fig. 2.

3.2. Synthetic datasets

To evaluate the validity of FS-RRC directly, we simulate synthetic datasets based on reference [20]. The first synthetic dataset SD_1 contains fifteen features $f_1, f_2, \ldots, f_{15}$, including ten features $f_1, f_2, \ldots, f_{10} \sim U(0,1)$, $f_{11} = (f_1 - f_2)/2$, $f_{12} = (f_1 + f_2)/2$, $f_{13} = f_3 + 0.1$, $f_{14} = f_4 - 0.2$, $f_{15} = 2 \times f_5$. The multi-class label C, which has eight different class labels, is defined as follows:

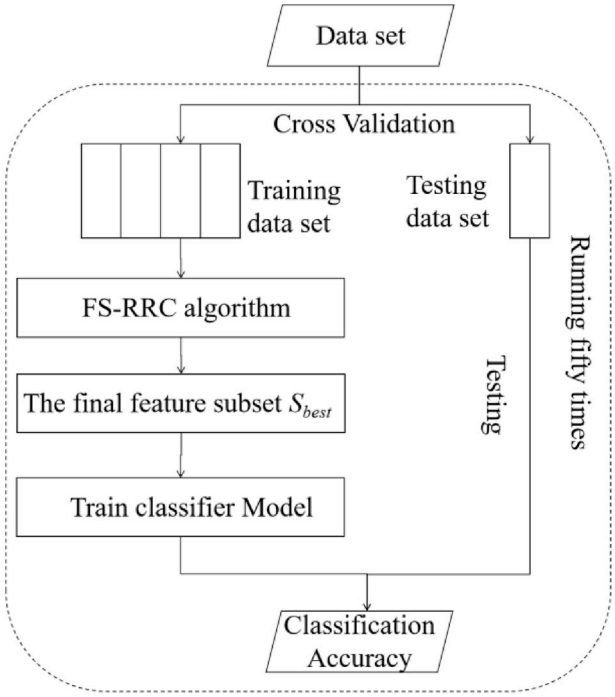

Fig. 2. The experimental scheme.

$$C=\left[C^1C^2C^3C^4\right]_2=\begin{cases}C^1=1 & if\ f_1>f_2\\ C^2=1 & if\ f_4>f_3\\ C^3=1 & if\ C^1+C^2=1\\ C^4=1 & if\ f_5>0.8\\ C^i=0 & otherwise(i=1,2,3,4)\end{cases}, \tag{4}$$

The relevant features are f_1, f_2, f_3, f_4, f_5, f_{11}, f_{13}, f_{14} and f_{15}, but redundancy exists, i.e., f_3 and f_{13} are redundant. The remaining features are irrelevant to the class label.

The second synthetic dataset SD_2 contains fourteen features $f_1, f_2, \ldots, f_{14}$, including ten features $f_1, f_2, \ldots, f_{10} \sim U(0,1)$, $f_{11}=f_1+0.1$, $f_{12}=f_2-0.2$, $f_{13}=2\times f_1$, $f_{14}=2\times f_2$. The multi-class label C is defined as follows:

$$C=\left[C^1C^2C^3C^4\right]_2=\begin{cases}C^1=1 & if\ f_1>0.5\&\&f_2>0.5\\ C^2=1 & if\ f_1<0.5\&\&f_2<0.5\\ C^3=1 & if\ f_1<0.5\&\&f_2>0.5\\ C^4=1 & if\ f_1>0.5\&\&f_2<0.5\\ C^i=0 & otherwise(i=1,2,3,4)\end{cases}, \tag{5}$$

The relevant features are f_1, f_2, f_{11}, f_{12}, f_{13} and f_{14}, but redundancy exists, i.e., f_1 and f_{11} are redundant. The remaining features are irrelevant to the class label. f_1 and f_2 together decide the class label C, f_1 and f_2 are complementary.

The multi-class label C is computed by binary number $[C^1C^2C^3C^4]_2$. For example, $C=[C^1C^2C^3C^4]_2=[1101]_2$ means a class label formed by $C^1=1$, $C^2=1$, $C^3=0$ and $C^4=1$.

3.3. Biological datasets

Fifteen public biological datasets (see Table 1) were also used to validate the performance of our FS-RRC method, which covers both binary-class problems and multi-class problems. The number of features ranges from 30 to 11225, and the number of samples varies from 38 to 569.

The "ratio" in Table 1 is an indicator of the difficulty of the feature selection for each dataset and is defined as follows [26]:

$$Ratio=\frac{N}{M^*n_c}, \tag{6}$$

where N is the sample number, M is the median arity of the features, and n_c is the number of class labels. Hence, a smaller ratio indicates a more challenging feature selection problem.

4. Results

4.1. Comparison in synthetic datasets

In this section, we evaluate the performance of FS-RRC and other competitor algorithms on synthetic datasets. The results of the experiment on SD_1 and SD_2 are given in Table 2. Bold font represents an optimal feature subset without irrelevant and redundant features. "Selected features" shows the features having a selected frequency larger than 75%. "*Sn*" represents the sensitivity of the feature selection method, "*Sp*" represents the specificity of the feature selection method.

All the algorithms can select the relevant features. Three methods, FS-RRC, CMIM and FCBF, can select the optimal feature subset on SD_1 and SD_2. However, CMIM has poor sensitivity and specificity compared with FS-RRC and FCBF. Although the specificity of FS-RRC is lower than that of FCBF on SD_1, it is 98.2% which is also good.

ReliefF selects the maximum number of features including four redundant features. MIFS, mRMR, ReliefF, MCRMCR, MCRMICR, SAFE and SVM-RFE also retain the redundant features on both synthetic datasets.

In SD_1, the sensitivities of FS-RRC, FCBF, MCRMCR and MCRMICR are 100.0%, but the specificities of MCRMCR and MCRMICR are quite low. In SD_2, the sensitivities of FS-RRC, FCBF, PGVNS, MCRMCR, MCRMICR, RCDFS and SAFE are 100.0%, but the specificities of MCRMCR, MCRMICR, RCDFS and SAFE are lower than the specificities of FS-RRC, FCBF and PVGN, which are 100%.

The experiments on the two synthetic datasets show that FS-RRC can select relevant and complementary features and remove irrelevant and redundant features effectively.

4.2. Comparison in biological datasets

In this section, we evaluate the performance of the selected feature by FS-RRC and other algorithms on biological datasets. The average classification accuracy rates of different classifiers for the fifteen datasets are shown in Fig. 3. The classification accuracy rates, sensitivities and specificities of SVM classifiers with linear kernel function are reported in Table 3, Table 4 and Table 5, respectively. The "Avg" in Table 3 (Table 4, Table 5) shows the average performance of each algorithm on the datasets. Bold font represents the highest value among all the methods in a dataset. Moreover, a *t*-test between FS-RRC and each baseline method is performed, "*" identifies that the corresponding baseline method significantly wins or loses FS-RRC at the 0.05 level. "Win/Tie/Loss" means that the number of datasets on which a method has a significantly higher/no significantly different/significantly lower accuracy rate with respect to FS-RRC. Table S1 in the supporting file gives the summary of significant Win/Tie/Loss results of each method compared with FS-RRC.

Table 3 shows that on SVM with linear kernel function, FS-RRC is superior to all the other competitor algorithms in terms of the classification accuracy rate for seven of the fifteen datasets. The standard

Table 1
Fifteen public biological datasets.

No.	Dataset	Features	Samples	Classes	Ratio	Source
1	Breast cancer	162	271	2	136	[27]
2	Breast2	4869	77	2	39	[28]
3	CNS	7129	60	2	30	[29]
4	Colon	2000	60	2	31	[30]
5	DLBCL_GEMS	5469	77	2	39	[31]
6	GSE28700	556	44	2	22	[32]
7	Hepato	7129	60	2	30	[33]
8	Leukemia	3051	38	2	19	[34]
9	Leukemia2_GEMS	11225	72	3	24	[31]
10	Leukemia3	7129	72	2	36	[34]
11	Lymphoma1	4026	62	3	21	[35]
12	Lymphoma2	4026	96	2	48	[36]
13	Prostate.data	6033	102	2	51	[37]
14	SRBCT-GEMS	2308	83	4	21	[38]
15	WDBC	30	569	2	285	[39]

Table 2
Comparison on synthetic datasets.

Dataset	SD_1			SD_2		
	Selected features	*Sn*	*Sp*	Selected features	*Sn*	*Sp*
FS-RRC	$\boldsymbol{f_3, f_4, f_5, f_{11}}$	100.0	98.2	$\boldsymbol{f_1, f_2}$	100.0	100.0
MIFS	$f_3, f_4, f_5, f_{11}, f_{15}$	99.6	84.0	f_1, f_2, f_{11}	99.6	89.5
mRMR	$f_1, f_2, f_3, f_4, f_5, f_{15}$	61.2	81.1	$f_1, f_2, f_{11}, f_{12}, f_{13}$	82.8	84.4
CMIM	$\boldsymbol{f_3, f_4, f_5, f_{11}}$	82.4	95.7	$\boldsymbol{f_1, f_2}$	97.2	98.4
ReliefF	$f_1, f_2, f_3, f_4, f_5, f_{11}, f_{13}, f_{14}, f_{15}$	84.8	59.7	$f_1, f_2, f_{11}, f_{12}, f_{13}, f_{14}$	91.8	71.5
FCBF	$\boldsymbol{f_3, f_4, f_5, f_{11}}$	100.0	100.0	$\boldsymbol{f_1, f_2}$	100.0	100.0
PGVNS	f_5, f_{11}	47.3	99.0	$\boldsymbol{f_1, f_2}$	100.0	100.0
MCRMCR	$f_2, f_3, f_4, f_5, f_{11}, f_{12}$	100.0	71.1	f_1, f_2, f_3	100.0	88.7
MCRMICR	$f_3, f_4, f_5, f_{11}, f_{12}, f_{13}, f_{14}$	100.0	59.6	f_1, f_2, f_3	100.0	88.7
RCDFS	$f_3, f_4, f_5, f_{11}, f_{13}, f_{14}$	97.2	80.0	$\boldsymbol{f_1, f_2}$	100.0	98.5
SAFE	f_5, f_{11}, f_{15}	50.0	90.9	f_1, f_2, f_{11}	100.0	91.7
SVM-RFE	$f_3, f_5, f_{11}, f_{14}, f_{15}$	86.8	72.3	f_{11}, f_{13}, f_{14}	52.6	70.8

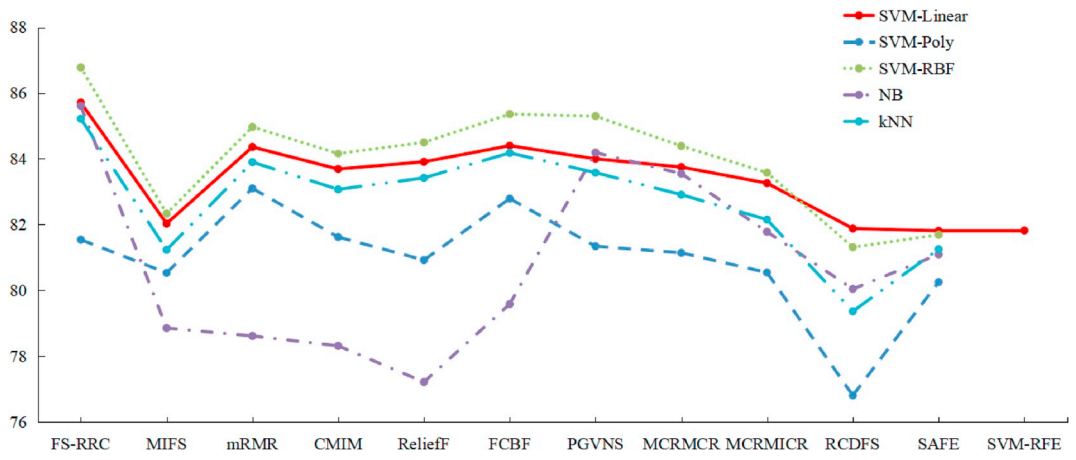

Fig. 3. Comparison among FS-RRC and eleven other feature selection methods in accuracy.

Table 3
Comparison of accuracy (%) on SVM with the linear kernel function.

	FS-RRC	MIFS	mRMR	CMIM	ReliefF	FCBF	PGVNS	MCRMCR	MCRMICR	RCDFS	SAFE	SVM-RFE
1	87.07	86.32*	85.96*	86.10*	86.43*	87.21	86.91	86.52*	86.12*	86.49*	**87.28**	82.14*
2	64.04	61.27*	63.43	**64.37**	60.51*	63.05	63.58	62.16	60.77*	61.95*	61.20*	61.49*
3	61.47	57.37*	60.63	59.93	**61.63**	59.17*	60.00	59.73	59.80	59.03*	57.23*	59.37
4	79.46	80.78	81.17*	80.01	**82.21***	80.97*	78.24	79.47	81.22*	80.02	81.96*	78.18
5	**95.74**	89.92*	92.21*	92.38*	91.27*	92.10*	94.85*	93.42*	89.07*	92.62*	87.13*	87.02*
6	69.61	67.38	70.09	69.26	68.51	**70.96**	69.45	68.77	70.18	70.53	69.33	69.64
7	64.23	59.50*	62.83	63.53	61.23*	62.40	62.97	63.87	**64.27**	63.50	60.90*	60.00*
8	**97.24**	91.10*	92.51*	86.46*	90.51*	94.80*	90.17*	89.64*	90.21*	87.68*	91.96*	90.47*
9	**93.94**	86.82*	90.72*	91.86*	91.03*	91.43*	89.68*	90.53*	87.35*	87.48*	85.68*	87.39*
10	**95.42**	90.73*	94.68*	93.22*	93.41*	94.05*	93.02*	92.88*	92.50*	91.07*	91.48*	91.43*
11	**98.46**	96.93*	96.73*	94.22*	95.92*	97.24*	96.39*	95.59*	95.69*	94.70*	95.29*	93.34*
12	**93.32**	84.72*	91.12*	90.99*	91.98*	90.63*	90.93*	90.84*	91.08*	90.81*	85.96*	88.75*
13	89.91	89.27	90.16	90.44	90.07	89.26	89.57	89.18	87.35*	87.92*	**92.19***	88.56*
14	**99.23**	91.70*	95.92*	95.61*	96.97*	96.33*	98.49*	96.44*	96.05*	94.45*	85.38*	92.88*
15	96.53	96.64	97.21*	97.04*	96.93*	96.45	95.69*	**97.25***	97.24*	96.87*	94.19*	96.63
Avg	**85.71**	82.03	84.36	83.69	83.91	84.40	84.00	83.75	83.20	83.01	81.81	81.82
Win/Tie/Loss	–	0/4/11	2/5/8	1/6/8	2/3/10	1/6/8	0/7/8	1/6/8	2/3/10	1/3/11	2/2/11	0/4/11

Table 4
Comparison of sensitivity (%) on SVM with the linear kernel function.

	FS-RRC	MIFS	mRMR	CMIM	ReliefF	FCBF	PGVNS	MCRMCR	MCRMICR	RCDFS	SAFE	SVM-RFE
1	94.04	93.25*	92.95*	93.07*	92.75*	**94.61***	93.64	88.55*	89.05*	88.56*	89.29*	87.47*
2	69.09	68.09	69.05	69.23	66.27*	**70.36**	68.59	67.35	64.72*	67.51	65.19*	66.55*
3	**73.69**	68.67*	72.97	72.05	71.08	70.56*	72.26	67.86*	69.66*	68.35*	66.24*	67.90*
4	84.85	83.90	86.10	85.15	**86.70***	86.45*	84.15	82.89*	84.74	82.48*	85.83	84.45
5	92.53	80.84*	84.21*	86.32*	84.63*	83.16*	**93.26**	89.70*	79.73*	78.92*	73.09*	77.26*
6	67.82	65.36	68.09	68.91	67.91	70.55	66.73	**71.74***	71.21*	68.84	69.89	69.82
7	73.90	69.90*	71.85	73.20	73.20	71.55	72.65	**73.95**	73.02	69.79*	69.57*	71.10*
8	**95.64**	80.73*	83.27*	79.82*	76.91*	88.73*	77.27*	86.61*	91.59*	81.08*	88.65*	80.91*
9	**93.67**	86.06*	90.39*	91.64*	90.55*	91.18*	89.08*	91.69*	87.71*	88.17*	86.52*	86.79*
10	94.24	85.20*	**94.32**	90.64*	88.72*	92.32*	91.04*	93.17	91.84*	79.18*	91.06*	85.04*
11	**99.19**	95.78*	95.60*	92.36*	93.62*	96.49*	95.21*	95.79*	97.22*	94.28*	92.19*	93.01*
12	84.94	73.24*	83.06*	82.12*	85.12	81.06*	81.94*	**92.90***	92.25*	85.50	84.23	80.06*
13	88.73	89.31	89.46	90.23*	90.04*	88.65	88.58	88.75	86.72*	89.15	**92.54***	88.81
14	**99.39**	92.28*	96.59*	96.32*	97.46*	97.11*	98.74*	98.92*	98.34*	97.64*	84.30*	93.60*
15	94.00	94.25	95.04*	94.69*	94.61*	93.78	93.02*	**97.30***	97.06*	97.11*	93.02*	94.13
Avg	**87.05**	81.79	84.86	84.38	83.97	85.10	84.41	85.81	84.99	82.44	82.10	81.79
Win/Tie/Loss	–	0/5/10	1/7/7	2/5/8	3/4/8	2/5/8	0/8/7	3/4/8	3/2/10	1/4/10	1/3/11	0/4/11

deviation of the classification accuracy rate is shown in Table S2 in the supporting file. Moreover, FS-RRC significantly outperforms MIFS, mRMR, CMIM, ReliefF, FCBF, PGVNS, MCRMCR, MCRMICR, RCDFS, SAFE and SVM-RFE over eleven, eight, eight, ten, eight, eight, eight, ten, eleven, eleven and eleven datasets by at most 8.60%, 4.73%, 10.78%, 6.73%, 3.64%, 7.07%, 7.6%, 7.03%, 9.56%, 13.85% and 8.72%, respectively. The "Avg" in Table 3 shows that FS-RRC is the best method in terms of classification accuracy rate.

Fig. 3 shows that FS-RRC obtains the maximal average classification accuracy rates in the fifteen datasets based on SVM with linear kernel function, SVM with RBF kernel function, NB and *k*NN. Therefore, combining the relevance, redundancy and complementarity could select a more powerful feature subset in most cases.

We also calculated the sensitivity and specificity based on SVM with

Table 5
Comparison of specificity (%) on SVM with the linear kernel function.

	FS-RRC	MIFS	mRMR	CMIM	ReliefF	FCBF	PGVNS	MCRMCR	MCRMICR	RCDFS	SAFE	SVM-RFE
1	65.85	65.22	64.66	64.90	67.16	64.69*	66.42	76.80*	75.64*	76.60*	**79.52***	65.91
2	57.21	52.18*	55.82	**57.82**	52.85*	53.27*	56.79	55.90	54.68	55.82	55.35	54.79
3	38.76	36.38	37.71	37.43	**44.10***	38.00	37.24	41.98	42.30	41.55	37.94	43.52*
4	69.64	**75.18***	72.18	70.64	74.09*	71.09	67.55	71.71	73.64*	72.02	75.10	66.82
5	**96.79**	92.90*	94.83*	94.34*	93.41*	95.00*	95.34*	96.15	93.07*	95.37*	92.19*	90.24*
6	71.18	69.36	72.00	69.45	69.18	71.27	**72.09**	67.99*	70.15	69.35	69.50	69.64
7	44.90	38.70*	44.80	44.20	37.30*	44.10	43.60	46.12	**46.54**	45.76	40.10*	37.80*
8	**97.93**	95.26*	96.30*	89.19*	96.07*	97.26	95.41*	92.23*	92.83*	91.06*	93.64*	94.37*
10	**96.04**	93.66*	94.85*	94.60*	95.91	94.98*	94.09*	94.76*	94.15*	93.38*	91.85*	94.85*
12	**97.90**	91.03*	95.52*	95.84*	95.74*	95.87*	95.84*	90.79*	91.41*	91.16*	86.89*	93.52*
13	91.08	89.20*	90.84	90.64	90.08	89.84	90.56	88.32*	86.95*	87.75*	**91.86***	88.28*
15	98.04	98.06	**98.49***	98.44*	98.31*	98.03	97.27*	97.02*	96.95*	96.78*	94.87*	98.11
Avg	**77.11**	74.76	76.50	75.62	76.18	76.12	76.02	76.65	76.53	76.38	75.73	74.82
Win/Tie/Loss	–	1/4/7	1/7/4	1/7/4	3/4/5	0/7/5	0/7/5	1/5/6	2/4/6	1/5/6	2/4/6	1/5/6

linear kernel function (see Tables 4 and 5). The standard deviation of the sensitivity and specificity is shown in Table S3 and Table S4 (see supporting file). The sensitivity measures the proportion of true positives. The specificity measures the proportion of true negatives. Table 4 shows that the sensitivities of FS-RRC are the highest in five of the fifteen datasets, and Table 5 shows that the specificities of FS-RRC are the highest in four of the twelve datasets. The average sensitivity and specificity of FS-RRC were the highest in all the datasets. Additionally, the sensitivity and specificity of FS-RRC are significantly higher than those of the other feature selection methods in most cases.

4.3. Comparison of stability

In addition to the classification accuracy, sensitivity and specificity, the stability of feature selection algorithms is also important. The stability of an algorithm is defined as its sensitivity to the data variation [40]. If the sample variation considerably changes the selected feature subset, the stability of the corresponding feature selection algorithm is poor. We adopted the average percentage of overlapping genes-related (APOGR) to measuring the stability of feature selection algorithms [41]. APOGR represents the consistency of two feature subsets based on the number of overlapping and related features and is defined as follows [41]:

$$APOGR = \frac{\sum_{1\le i<j\le k^*t} \frac{1}{2}\left(\frac{|Sub_i \cap Sub_j| + Cor_{ij}}{|Sub_i|} + \frac{|Sub_i \cap Sub_j| + Cor_{ji}}{|Sub_j|}\right)}{C^2_{k^*t}}, \quad (7)$$

where Sub_i is the i-th selected feature subset, Cor_{ij} represents the number of the selected features in Sub_i that are not the same but significantly positively correlated with at least one feature in Sub_j, k is the fold of cross-validation, and t is the running times for a feature selection method.

A large APOGR value means that the feature selection method is good and the variation in the training data has a small influence on the selected feature subset. Fig. 4 shows the comparison in average stability among FS-RRC, MIFS, mRMR, CMIM, ReliefF, FCBF, PGVNS, MCRMCR, MCRMICR, RCDFS, SAFE and SVM-RFE in the fifteen datasets. The APOGR of FS-RRC ranks second, and it is slightly lower than that of mRMR. mRMR obtains the maximal APOGR value. Therefore, FS-RRC is more stable than the other ten algorithms except for mRMR.

4.4. Comparison of time complexity

Time complexity is an important criterion for measuring the performance of feature selection algorithms. A good feature selection algorithm has not only higher classification accuracy but also less time complexity [1]. When the datasets are high-dimensional features or large samples, it is much important to consider the time complexity.

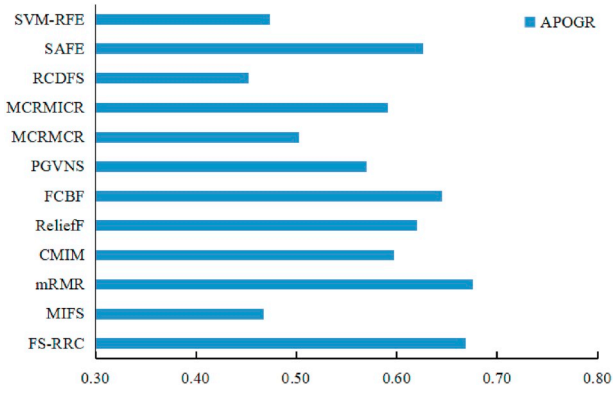

Fig. 4. Comparison of stability.

Table 6 gives the time complexity of twelve feature selection algorithms.

In Table 6, m is the number of samples in the dataset, n is the number of features, I is the number of iterations for PGVNS, and k is the number of selected features. From Table 6, we can conclude that FS-RRC, FCBF and SAFE are the fastest of the twelve algorithms. The time complexity of SVM-RFE is the highest of the twelve methods. Table 6 also illustrates that the computational complexities of filters are lower than those of the wrappers.

5. Discussion

Table 7 lists the characteristics of each feature selection method, which are selecting relevant features (SRF), elimination redundant features (ERF), considering complementary features (CCF), avoiding high-dimensional MI estimation (AHMI), without setting the number of selected features in prior (SNP), and no other parameters (OP) [42].

Table 7 shows that FS-RRC is parameter-free, which does not set the number of selected features in prior and other parameters. Additionally, FS-RRC avoids high-dimensional MI estimation, which ensures precision in computing MI. Selecting relevant features is the main goal for feature

Table 6
Comparison of time complexity.

	FS-RRC	MIFS	mRMR	CMIM	ReliefF	FCBF
Complexity	$O(n^2)$	$O(mn^2)$	$O(mn^2)$	$O(mn^2)$	$O(nm^2)$	$O(n^2)$
	PGVNS	MCRMCR	MCRMCIR	RCDFS	SAFE	SVM-RFE
Complexity	$O(Imn^2)$	$O(k^3mn)$	$O(k^2mn)$	$O(kmn)$	$O(n^2)$	$O(\max(m, n)n^2)$

Table 7
Characteristics comparison.

	SRF	ERF	CCF	AHMI	SNP	OP
FS-RRC	Yes	Yes	Yes	Yes	No	No
MIFS	Yes	Yes	No	Yes	Yes	Yes
mRMR	Yes	Yes	No	Yes	Yes	No
CMIM	Yes	Yes	No	Yes	Yes	No
ReliefF	Yes	Yes	No	Yes	Yes	Yes
FCBF	Yes	Yes	No	Yes	No	No
PGVNS	Yes	Yes	No	Yes	No	Yes
MCRMCR	Yes	Yes	No	Yes	Yes	Yes
MCRMICR	Yes	Yes	No	Yes	Yes	Yes
RCDFS	Yes	Yes	Yes	Yes	Yes	No
SAFE	Yes	Yes	Yes	Yes	No	No
SVM-RFE	Yes	Yes	No	Yes	No	Yes

selection methods. Eliminating redundant features aims at not selecting feature subsets with high redundancy. Considering complementary features in the procedure of feature selection aims at keeping the features that have strong relevance with class label by combining with their complementary features. FS-RRC combines relevance, redundancy and complementarity to select the feature subset, which can help to define more important information from the complex biological data.

6. Conclusions

In complex biological systems, molecules relate to each other, and they work together to reflect specific physiological and pathological changes. This study focuses on feature cooperation as well as feature relevance and redundancy and proposes a new feature selection algorithm FS-RRC. While removing irrelevant and redundant features, FS-RRC can select complementary features. The experiment on the two synthetic datasets illustrated the effectiveness of FS-RRC. The experiment on the fifteen public biological datasets also showed that FS-RRC could select more powerful feature subsets than MIFS, mRMR, CMIM, ReliefF, FCBF, PGVNS, MCRMCR, MCRMICR, RCDFS, SAFE and SVM-RFE in most cases, and its sensitivity to the change in training samples is low. Hence, feature synergy contains meaningful information related to different pathological states. Combining feature relevance, redundancy and complementarity can help to define more important information from the complex biological data for studying disease diagnosis and mechanisms.

Author contributions

X. Lin and C. Li conceived and designed the experiments; C. Li, X. Luo, Y. Qi and Z. Gao searched the datasets and performed the experiments. C. Li drafted the manuscript and X. Lin revised the manuscript.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.103667.

References

[1] J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: a new perspective, Neurocomputing 300 (2018) 70–79, https://doi.org/10.1016/j.neucom.2017.11.077.

[2] X. Lin, F. Yang, L. Zhou, P. Yin, H. Kong, W. Xing, X. Lu, L. Jia, Q. Wang, G. Xu, A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information, J. Chromatogr., B 910 (2012) 149–155, https://doi.org/10.1016/j.jchromb.2012.05.020.

[3] J. Wang, Z. Feng, N. Lu, J. Luo, Toward optimal feature and time segment selection by divergence method for EEG signals classification, Comput. Biol. Med. 97 (2018) 161–170, https://doi.org/10.1016/j.compbiomed.2018.04.022.

[4] L. Wolf, A. Shashua, Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weight-based approach, J. Mach. Learn. Res. 6 (2005) 1855–1887.

[5] D.H. Peluffo, J.A. Lee, M. Verleysen, J.L. Rodríguez-Sotelo, G. Castellanos - Domínguez, Unsupervised relevance analysis for feature extraction and selection. A distance-based approach for feature relevance, in: International Conference on Pattern Recognition, Applications and Methods - ICPRAM 2014, 2014.

[6] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (2007) 2507–2517, https://doi.org/10.1093/bioinformatics/btm344.

[7] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1997) 273–324, https://doi.org/10.1016/S0004-3702(97)00043-X.

[8] A. Wang, A. Ning, J. Yang, G. Chen, L. Li, G. Alterovitz, Wrapper-based gene selection with Markov blanket, Comput. Biol. Med. 81 (2017) 11–23, https://doi.org/10.1016/j.compbiomed.2016.12.002.

[9] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422, https://doi.org/10.1023/A:1012487302797.

[10] Y. Liang, C. Liu, X.Z. Luan, K.S. Leung, T.S. Chan, Z.B. X, H. Zhang, Sparse logistic regression with a $L_{1/2}$ penalty for gene selection in cancer classification, BMC Bioinf. 14 (2013) 198, https://doi.org/10.1186/1471-2105-14-198.

[11] H. Jiang, Y. Liang, The $L_{1/2}$ regularization network Cox model for analysis of genomic data, Comput. Biol. Med. 100 (2018) 203–208, https://doi.org/10.1016/j.compbiomed.2018.07.009.

[12] A.J. Ferreira, M.A.T. Figueiredo, Efficient feature selection filters for high-dimensional data, Pattern Recogn. Lett. 33 (2012) 1794–1804, https://doi.org/10.1016/j.patrec.2012.05.019.

[13] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Trans. Neural Network. 5 (1994) 537–550, https://doi.org/10.1109/72.298224.

[14] 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 (2005) 1226–1238, https://doi.org/10.1109/TPAMI.2005.159.

[15] F. Fleuret, Fast binary feature selection with conditional mutual information, J. Mach. Learn. Res. 5 (2004) 1531–1555, https://doi.org/10.1007/s10182-011-0155-4.

[16] H. Zhou, Y. Zhang, Y. Zhang, H. Liu, Feature selection based on conditional mutual information: minimum conditional relevance and minimum conditional redundancy, Appl. Intell. 49 (2019) 883–896, https://doi.org/10.1007/s10489-018-1305-0.

[17] L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, J. Mach. Learn. Res. 5 (2004) 1205–1224.

[18] M. García-Torres, F. Gómez-Vela, B. Melián-Batista, J.M. Moreno-Vega, High-dimensional feature selection via feature grouping: a variable neighborhood search approach, Inf. Sci. 326 (2016) 102–118, https://doi.org/10.1016/j.ins.2015.07.041.

[19] Z. Chen, C. Wu, Y. Zhang, Z. Huang, B. Ran, M. Zhong, N. Lyu, Feature selection with redundancy-complementariness dispersion, Knowl. Base Syst. 89 (2015) 203–217, https://doi.org/10.1016/j.knosys.2015.07.004.

[20] S. Singha, P.P. Shenoy, An adaptive heuristic for feature selection based on complementarity, Mach. Learn. 107 (2018) 2027–2071, https://doi.org/10.1007/s10994-018-5728-y.

[21] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: a data perspective, ACM Comput. Surv. 50 (2018) 94, https://doi.org/10.1145/3136625.

[22] D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. Mcvean, P. J. Turnbaugh, E.S. Lander, M. Mitzenmacher, P.C. Sabeti, Detecting novel associations in large data sets, Science 334 (2011) 1518–1524, https://doi.org/10.1126/science.1205438.

[23] H. Lyu, M. Wan, J. Han, R. Liu, C. Wang, A filter feature selection method based on the Maximal Information Coefficient and Gram-Schmidt Orthogonalization for biomedical data mining, Comput. Biol. Med. 89 (2017) 267–274, https://doi.org/10.1016/j.compbiomed.2017.08.021.

[24] J.R. Vergara, P.A. Estévez, A review of feature selection methods based on mutual information, Neural Comput. Appl. 24 (2014) 175, https://doi.org/10.1007/s00521-013-1368-0.

[25] U.M. Fayyad, K.B. Irani, Multi-interval discretization of continuous-valued attributes for classification learning, in: R. Bajcsy (Ed.), Proceedings of the 13th International Joint Conference on Artificial Intelligence, Morgan Kaufmann-Publishing Inc, San Mateo, 1993, pp. 1022–1027.

[26] G. Brown, A. Pocock, M.J. Zhao, M. Luján, Conditional likelihood maximisation: a unifying framework for information theoretic feature selection, J. Mach. Learn. Res. 13 (2012) 27–66.

[27] J. Budczies, S.F. Brockmöller, B.M. Müller, D.K. Barupal, C. Richter - Ehrenstein, A. Kleine - Tebbe, J.L. Griffin, M. Orešič, M. Dietel, C. Denkert, O. Fiehn, Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism, J. Proteom. 94 (2013) 279–288, https://doi.org/10.1016/j.jprot.2013.10.002.

[28] L.J. Van't Veer, H. Dai, M.J. van de Vijver, Y.D. He, A.A.M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen, G.J. Schreiber, R. M. Kerkhoven, C. Roberts, P.S. Linsley, R. Bernards, S.H. Friend, Gene expression profiling predicts clinical outcome of breast cancer, Nature 415 (2002) 530–536, https://doi.org/10.1038/415530a.

[29] S.L. Pomeroy, P. Tamayo, M. Gaasenbeek, L.M. Sturla, M. Angelo, M.E. Mclaughlin, J.Y.H. Kim, L.C. Goumnerova, P.M. Black, C. Lau, J.C. Allen, D. Zagzag, J.M. Olson, T. Curran, C. Wetmore, J.A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D.N. Louis, J.P. Mesirov, E.S. Lander, T.R. Golub, Prediction of central nervous system embryonal tumour outcome based on gene expression, Nature 415 (2002) 436–442, https://doi.org/10.1038/415436a.

[30] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D.M.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6745–6750, https://doi.org/10.1073/pnas.96.12.6745.

[31] A. Statnikov, I. Tsamardinos, Y. Dosbayev, C.F. Aliferis, Gems: a system for automated cancer diagnosis and biomarker discovery from microarray gene expression data, Int. J. Med. Inf. 74 (2005) 491–503, https://doi.org/10.1016/j.ijmedinf.2005.05.002.

[32] C.W. Tseng, C.C. Lin, C.N. Chen, H.C. Huang, Integrative network analysis reveals active microRNAs and their functions in gastric cancer, BMC Syst. Biol. 5 (2011) 99, https://doi.org/10.1186/1752-0509-5-99.

[33] N. Iizuka, M. Oka, H. Yamada-Okabe, M. Nishida, Y. Maeda, N. Mori, T. Takao, T. Tamesa, A. Tangoku, H. Tabuchi, K. Hamada, H. Nakayama, H. Ishitsuka, T. Miyamoto, A. Hirabayashi, S. Uchimura, Y. Hamamoto, Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection, Lancet 361 (2003) 923–929, https://doi.org/10.1016/S0140-6736(03)12775-4.

[34] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537, https://doi.org/10.1126/science.286.5439.531.

[35] A.A. Alizadeh, M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J.I. Powell, L. Yang, G.E. Marti, T. Moore, J. Hudson Jr., L. Lu, D.B. Lewis, R. Tibshirani, G. Sherlock, W.C. Chan, T. C. Greiner, D.D. Weisenburger, J.O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J.C. Byrd, D. Botstein, P.O. Brown, L.M. Staudt, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, Nature 403 (2000) 503–511, https://doi.org/10.1038/35000501.

[36] A. Rakotomamonjy, Variable selection using SVM-based criteria, J. Mach. Learn. Res. 3 (2003) 1357–1370, https://doi.org/10.1162/153244303322753706.

[37] D. Singh, P.G. Febbo, K. Ross, D.G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A.V. D'Amico, J.P. Richie, E.S. Lander, M. Loda, P.W. Kantoff, T. R. Golub, W.R. Sellers, Gene expression correlates of clinical prostate cancer behavior, Canc. Cell 1 (2002) 203–209, https://doi.org/10.1016/s1535-6108(02)00030-2.

[38] J. Khan, J.S. Wei, M. Ringnér, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, P.S. Peterson, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nat. Med. 7 (2001) 673–679, https://doi.org/10.1038/89044.

[39] A. Asuncion, D. Newman, UCI repository of machine learning datasets. http://archive.ics.uci.edu/ml/datasets.html.

[40] Z. Wang, M. Li, J. Li, A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure, Inf. Sci. 307 (2015) 73–88, https://doi.org/10.1016/j.ins.2015.02.031.

[41] M. Zhang, L. Zhang, J. Zou, C. Yao, H. Xiao, Q. Liu, J. Wang, D. Wang, C. Wang, Z. Guo, Evaluating reproducibility of differential expression discoveries in microarray studies by considering correlated molecular changes, Bioinformatics 25 (2009) 1662–1668, https://doi.org/10.1093/bioinformatics/btp295.

[42] B. Brycki, I. Kowalczyk, J. Werner, T. Borowiak, I. Wolska, Information-Theoretic feature selection in microarray data using variable complementarity, IEEE J. Sel. Top. Signal Process. 2 (2008) 261–274, https://doi.org/10.1109/JSTSP.2008.923858.