Journal of Chromatography B, 910 (2012) 149–155

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

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

A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information☆

Xiaohui Lin[a], Fufang Yang[a], Lina Zhou[b], Peiyuan Yin[b], Hongwei Kong[b], Wenbin Xing[c], Xin Lu[b], Lewen Jia[d], Quancai Wang[a], Guowang Xu[b,*]

[a] School of Computer Science and Technology, Dalian University of Technology, 116024 Dalian, China
[b] CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China
[c] The Sixth People's Hospital, Dalian 116001, China
[d] Department of Nephrology, The First Affiliated Hospital of Dalian Medical University, Dalian 116011, China

ARTICLE INFO

Article history:
Received 9 February 2012
Received in revised form 11 May 2012
Accepted 14 May 2012
Available online 24 May 2012

Keywords:
Artificial contrast variables
Mutual information
SVM-RFE
Liver diseases
Metabolomics

ABSTRACT

Filtering the discriminative metabolites from high dimension metabolome data is very important in metabolomics study. Support vector machine-recursive feature elimination (SVM-RFE) is an efficient feature selection technique and has shown promising applications in the analysis of the metabolome data. SVM-RFE measures the weights of the features according to the support vectors, noise and non-informative variables in the high dimension data may affect the hyper-plane of the SVM learning model. Hence we proposed a mutual information (MI)-SVM-RFE method which filters out noise and non-informative variables by means of artificial variables and MI, then conducts SVM-RFE to select the most discriminative features. A serum metabolomics data set from patients with chronic hepatitis B, cirrhosis and hepatocellular carcinoma analyzed by liquid chromatography–mass spectrometry (LC–MS) was used to demonstrate the validation of our method. An accuracy of 74.33 ± 2.98% to distinguish among three liver diseases was obtained, better than 72.00 ± 4.15% from the original SVM-RFE. Thirty-four ion features were defined to distinguish among the control and 3 liver diseases, 17 of them were identified.

1. Introduction

Metabolomics, as a branch of systems biology, quantitatively measures the metabolic response of living systems to environment stimuli or genetic modifications [1]. It studies the biological process with a different view from that of genomics, proteomics and transcriptomics. The change of metabolite concentrations could reflect the healthy state of a living system [2], and tell us what has happened. In recent years, metabolomics has shown promising applications in many fields, such as disease diagnosis [3], drug research and development [4], etc. Cancer is one of the major diseases which bother the human. Many metabolomic studies have been conducted on liver cancer [5], lung cancer [6], prostate cancer [7], bladder cancer [8], etc.

Liquid chromatography–mass spectrometry (LC–MS) is one of the main analytical techniques in metabolomics study. LC–MS metabolomic data can bring researchers very rich information. However, it also bothers the researchers for the high dimension. There exist noisy variables besides the problem related ones people are really interested in. To filter out the noises from the high-dimensional data, many multivariate analytical techniques have been adopted.

Feature selection technique is an efficient tool to select the meaningful information from the metabolome dataset. It can be organized into three categories: filter, wrapper and embedded methods, depending on how it combines the feature selection procedure with the construction of the learning model [9]. Support vector machine-recursive feature elimination (SVM-RFE) as a popular embedded method was firstly introduced in 2002 to do gene selection for cancer classification [10]. It is much more robust to data over-fitting than other feature selection techniques [10,11] and has shown its power in many fields, for example, genomics [12], proteomics [13], metabolomics [14], etc. During performance, it removes one feature with the smallest weight iteratively to a feature rank until all the features have been removed.

Removing only one feature at one time is quite time consuming, some options have been proposed to delete a number of features with the lowest weights in each iteration, such as E-RFE [15] and SVM-RFE-annealing [16]. To improve the robustness of SVM-RFE

☆ This paper belongs to the Special Issue Chemometrics in Chromatography, Edited by Pedro Araujo and Bjørn Grung.

* Corresponding author. Tel.: +86 411 84379530; fax: +86 411 84379559.
E-mail address: xugw@dicp.ac.cn (G. Xu).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.05.020

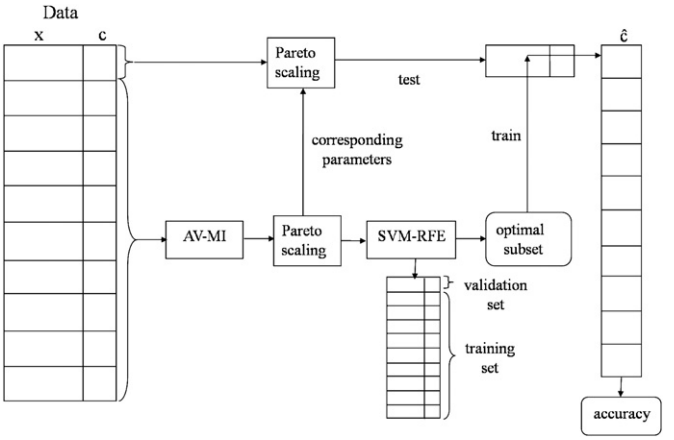

Fig. 1. Procedure of MI-SVM-RFE. x, c and ĉ represent the original data, original sample class marker and predictive sample class marker, respectively.

to noise and outliers, R-SVM [17] and F-SVM [18] were proposed. Though SVM-RFE was originally proposed for two class problems, now, it has been generalized to the analysis of multi-class problems [19,20].

SVM-RFE calculates the weight of each feature according to the support vectors of the current learning model. While the LC–MS metabolomic data usually contain noisy features which may affect the optimal hyper-plane constructed by SVM and influence the weights of the features. Therefore some meaningful features may not be evaluated correctly. Even worse, some meaningful features may be evaluated as non-meaningful due to the incorrect hyper-plane, and may be removed in early iterations. To select the most informative ones from the high dimension data, filtering out the noises before SVM-RFE is quite helpful. The artificial contrast variable is an autonomous variable selection method [21]. If the importance of a variable is significantly lower than its permuted counterpart, the variable is non-problem related and non-meaningful [22].

Mutual Information (MI) is a frequently used filter method to evaluate a feature's distinguishing ability. Here we combined artificial variables (AV) and MI to filter out non-informative variables firstly, and then conducted SVM-RFE to select the most discriminative ion features from the LC–MS data. A data set from serum LC–MS analysis of patients with chronic hepatitis B, cirrhosis and hepatocellular carcinoma was used to show the validation of our method.

2. Methods

The metabolome data usually contain hundreds, even thousands of variables, there may exist noisy variables, affecting the optimal hyper-plane constructed by SVM and the weights of the variables. If the original data contain such information, some meaningful features may be wrongly measured as non-meaningful. To obtain the most informative features, we propose a MI-SVM-RFE method which first filters out the noisy features by means of artificial variables and mutual information (AV–MI), and then selects the "best" feature subset by SVM-RFE. The whole procedure of the method is given in Fig. 1.

2.1. Mutual information

Mutual information as a statistic measurement can reflect how much information a feature contains about the class label or another feature. Let f denote a feature and C denote the class label, $I(f,C)$ denote the mutual information between f and C. $I(f,C)$ is defined by the following formula [23]:

$$I(f;C) = \sum_{j=1}^{m}\sum_{i=1}^{n} p(f_i, c_j)\log_2 \frac{p(f_i, c_j)}{p(f_i)p(c_j)}$$

where n and m are the number of different values of f and C, respectively, $p(f_i)$ is the probability of f's value which equals to f_i, $p(c_j)$ is the probability of C's value which equals to c_j and $p(f_i, c_j)$ is the joint probability.

The mutual information between a feature and the class label can reflect the feature's distinguishing ability. The larger the mutual information is, the more distinguishable the feature is.

2.2. Artificial contrast variable

By using the artificial contrast variables, non-problem related information could be excluded automatically [21]. For a feature f in the original data, its artificial variable *ar-f* is constructed by permuting f's values on the samples randomly. Hence the artificial variable and its corresponding original feature have the same value set but different distributions. Since the distribution of the value is rearranged with randomness, if the mutual information of the original variable with the class label C is smaller than that of the artificial contrast one, it is more likely to be a noise.

The artificial contrast variable may be affected by the randomness of the process because the values are rearranged randomly. In order to get a more accurate measurement, the permutation is run more than once [21]. Here we construct $t > 1$ artificial contrast variables for each feature and calculate the average mutual information of the t artificial variables with class label C. The feature whose MI is smaller than the average mutual information of its artificial contrast variables is more likely to be non-informative and could be filtered out.

2.3. SVM-RFE

After filtering out the noisy and non-related information, SVM-RFE is adopted to select the most discriminative ion features. First, SVM model is built on the current feature subset, the weight of each feature is computed, $r\%$ ($0 < r < 100$) ion features with the smallest weights are removed. If the number of $r\%$ current ion features is smaller than 1, only one ion feature with the smallest weight is removed. The iteration is continued until all the ion features have been removed. In each iteration, the current ion feature subset is evaluated by the k-fold cross-validation. The one with the highest cross-validation accuracy is kept as the final selected feature subset.

3. Experimental

3.1. Sample collection and sample preparation

Thirty aliquots of fasting sera were collected from patients with chronic hepatitis B (CHB), cirrhosis (CIR) and hepatocellular carcinoma (HCC), respectively, from the Sixth People's Hospital, Dalian. Each group of patients was the age and sex matched and there were another 30 aliquots of fasting sera from healthy volunteers as the healthy control. Each patient has taken liver function test, tumor markers test, ultrasonography and CT or MRI. About 1 ml serum sample was collected and stored at −80 °C.

For each sample, 100 μl serum was drawn into a centrifugal tube with 400 μl acetonitrile added for protein precipitation. After votexing for 30 s, the mixture was centrifuged at 15,000 × *g* for 10 min at 4 °C. The supernatant was directly analyzed by LC–MS.

Equal aliquot of supernatant from each sample was pooled as the quality control (QC) sample. The QC sample was continuously

run 10 times to equilibrate the column and then was run once after 10 samples to monitor the stability of LC–MS system during sample analysis.

3.2. Liquid chromatography–mass spectrometry analysis

An Agilent 1200 rapid resolution liquid chromatography system (Agilent, USA) used for liquid chromatographic separation was coupled to an Agilent 6510 Q-TOF mass spectrometer (Agilent, USA) with dual electrospray ionization (ESI) source used for data acquisition. The HILIC BEH column (100 mm × 2.1 mm, 1.7 μm) (Waters, USA) was used for the separation of polar compounds with column temperature set at 35 °C. The chromatographic flow rate was 0.35 ml/min. The mobile phase A was 80 mM ammonium formate modified with 0.1% (v/v) formic acid and mobile phase B was 0.1% formic acid in acetonitrile. The linear gradient started with 5% A, increasing to 9% A at 8 min, then to 15% A at 14 min, and then 50% A at 21 min, at last to 5% A at 24 min. Its total run time was 30 min including equilibration of 6 min. Data were acquired in the positive ion mode with the ion source temperature set at 300 °C. And other source parameters were set as our previous protocol [24]. The full mass accuracy calibration was exerted before analysis with residual error within 0.2 ppm. And a real-time mass calibration was performed to minimize the fluctuation of mass during the analysis. This ensures the stability of mass measurement during the duty cycle. The robustness of LC–MS system was ensured by the tight clustering of QC injections projected onto the plane of principal components in the principal component analysis (PCA) [25].

3.3. Data analysis

3.3.1. Data pretreatment

The deconvolution of raw total ion current spectra was performed by the Molecular Features Extraction program from Agilent. Then, the peak alignment was done with the mass window set as ± 0.03 Da and the retention time window set as ± 0.3 min. Then, an excel table was exported with retention time, *m*/*z* and peak intensity. If the peak area of a feature is equal to zero in more than 20% samples in each group, it was deleted [26]. The features with relative standard deviation (RSD) > 30% according to QC samples were also excluded. Finally, 196 variables were left. For the remaining missing values, we adopted the class-conditional mean imputation [27,28], that is, in a group, if the peak area of a variable is equal to zero in less or equal than 30% samples, we replaced the missing values of the variable in the group with the mean of nonzero values in that group.

3.3.2. Feature selection

Pareto scaling on the training data (Fig. 1) was adopted, the obtained parameters (mean value and standard deviation) were applied to the test data. The penalty factor, *r* and the kernel function in SVM-RFE is set to 1, 5, and *linear kernel function*, respectively. The implementation of SVM was from LIBSVM (http://www.csie.ntu.Edu.tw/~cjlin/libsvm), MI-SVM-RFE and SVM-RFE were written in C++.

To select the important ion features related to the liver diseases, we divide our research into five sub-objects: (1) a binary problem to distinguish between the normal group and the disease group (N *vs* M); (2) three binary problems to distinguish between every two of CHB, CIR and HCC (i.e., CHB *vs* CIR, CHB *vs* HCC, CIR *vs* HCC); (3) a three-class problem to distinguish among CHB, CIR and HCC (CHB *vs* CIR *vs* HCC) simultaneously.

Ten-fold cross-validation was applied to evaluate the performance of our method. In the training set, our method AV–MI is first performed to remove some noisy variables and *t* is set to 100. Then SVM-RFE is carried out to select the most discriminative ion features (Fig. 1). The performance of the method is tested by the test set. The ten-fold cross-validation runs 10 times to get a more reliable result.

PCA and partial least squares discriminant analysis (PLS-DA) were performed by SIMCA-P 11.5 (Umea, Sweden).

Table 1
Accuracy comparison of two methods.

Classification	SVM-RFE (%)	MI-SVM-RFE (%)
N *vs* M	100	100
CHB *vs* CIR	76.83 ± 3.09	82.67 ± 3.94
CHB *vs* HCC	79.83 ± 3.37	79.83 ± 3.19
CIR *vs* HCC	79.83 ± 4.87	85.17 ± 2.99
CHB *vs* CIR *vs* HCC	72.00 ± 4.15	74.33 ± 2.98

4. Results and discussion

4.1. Performance of MI-SVM-RFE

To demonstrate the performance of our MI-SVM-RFE method, we compared it with the original SVM-RFE. Ten-fold cross-validation was run 10 times for both methods. The average classification accuracy rate and the standard deviation were given in Table 1. It is clear that MI-SVM-RFE outperforms the original SVM-RFE in classification accuracy and the standard deviation. This means: (1) our method can select virtually more discriminative features; (2) filtering out noise by artificial variable and MI is quite meaningful; (3) the non-problem related variables in the high dimension metabolomic data could affect the weight calculated by SVM learning model. The accuracy rates of MI-SVM-RFE in distinguishing between CHB and CIR, CIR and HCC, and among the three diseases are 5.84%, 5.34% and 2.33% higher than those of the original SVM-RFE (see Table 1), respectively. Though two methods have the same accuracy rates in distinguishing between CHB and HCC, the standard deviation of MI-SVM-RFE is much lower than that of SVM-RFE.

Further, for binary classification cases we calculated the average sensitivity and specificity (Table 2). It can be observed that in most of the cases the sensitivity and specificity obtained in MI-SVM-RFE are higher than those in SVM-RFE.

4.2. Analysis of the selected features

Since the ten-fold cross-validation was run 10 times, there were 100 selected feature subsets in each research sub-object, the frequency of each feature in the 100 feature subsets was computed. And the features were ranked according to their frequency in a descending order. Thus the top ranked features are the most informative ones. Here 15 top ranked features were chosen in each classification described in Table 1. Totally 34 ion features were got from the 5 classifications, among which 33 ion features were got from 3 binary classifications to distinguish between every two of CHB, CIR and HCC, and the three-class problem to distinguish among CHB, CIR and HCC, simultaneously.

Fig. 2A and B shows the PCA score plots of the original data and the selected 34 ion features. Similarly, PCA plots of the three liver diseases are given in Fig. 3. From Fig. 2A and Fig. 3A, it can be seen that all the samples including normal ones and those of the three liver diseases are mixed together due to the noisy and non-informative variables. In Fig. 2B and Fig. 3B, different classes of the samples show a clear separated trend, it means that MI-SVM-RFE is quite effective in selecting the discriminative features. Fig. 4 shows the PLS-DA score plot of the three liver diseases based on the selected 33 ion features. Three different liver disease groups

Table 2
Comparison of sensitivity and specificity*

Classification	SVM-RFE sensitivity (%)	MI-SVM-RFE sensitivity (%)	SVM-RFE specificity (%)	MI-SVM-RFE specificity (%)
N *vs* M	100	100	100	100
CHB *vs* CIR	76.00 ± 4.66	81.33 ± 6.13	77.67 ± 5.22	84.00 ± 4.92
CHB *vs* HCC	79.00 ± 2.74	78.00 ± 3.91	80.67 ± 5.84	81.67 ± 4.23
CIR *vs* HCC	78.67 ± 5.92	83.00 ± 5.76	81.00 ± 5.45	87.33 ± 5.16

* In each run of ten-fold cross-validation, every sample is used as a test sample once and gets a predicted class label based on which the sensibility and specificity are calculated. As the ten-fold cross-validation runs 10 times, the average specificity and its standard deviation is calculated.

are clearly separated. R^2-intercept and Q^2-intercept are 0.196 and −0.316, respectively, implying that the PLS-DA model does not overfit the data.

Based on our previous method [29], of the 34 ion features, 17 ions were identified (see Table 3) including two phosphatidylethanolamines (PEs), five acyl-carnitines (acyl-CNs), carnitine, lysophosphatidylcholine (18:2) (lysoPC (18:2)) and sphingomyelin (d18:0/22:2 (OH)) (SM (d18:0/22:2 (OH)). Among these identified compounds, 11 compounds are differentially expressed in CIR and CHB. PC (18:2/18:3), pimelylcarnitine and acetylcarnitine etc. contribute to the discrimination of HCC from CHB. And SM (d18:0/22:2 (OH)), pimelylcarnitine and carnitine etc. contribute to the discrimination of HCC from CIR. The compounds identified above involved in fatty acid oxidation, phospholipid metabolism which are closely associated with the development of chronic liver diseases (CHB and CIR) and HCC [30].

Long-chain acyl-CNs are responsible for the transportation of fatty acids into mitochondria for β-oxidation. The accumulation of C16:1-CN (Fig. 5A) and C18:1-CN (Fig. 5B) in CIR and HCC compared to CHB and healthy control demonstrate that there are significant metabolic changes involved in fatty acid oxidation in more severe liver diseases (CIR and HCC). And the alteration of these two serum long-chain acyl-CNs are consistent with the observation in our previous investigation using reversed-phase liquid chromatography separation method [31].

Carnitine (Fig. 5C) is essential for fatty acid β-oxidation in mitochondria [32]. And urinary carnitine has been reported elevated in HCC cases in comparison to healthy control and cirrhosis, which may result from tumor overproduction of carnitine for mitochondria β-oxidation to support elevated metabolic activity and high cell-turnover in HCC [33,34]. And it is also reported that carnitine can protect mitochondria from oxidative damage [35,36] and a long-term administration of L-carnitine has been observed to inhibit hepatitis and subsequent hepatocellular carcinoma in Long–Evans Cinnamon rats [37]. So, the lower level of carnitine in CHB and CIR cases compared to healthy control may imply a severe liver mitochondria injury. By contrast, the serum carnitine increases in HCC cases compared to CIR, this may be

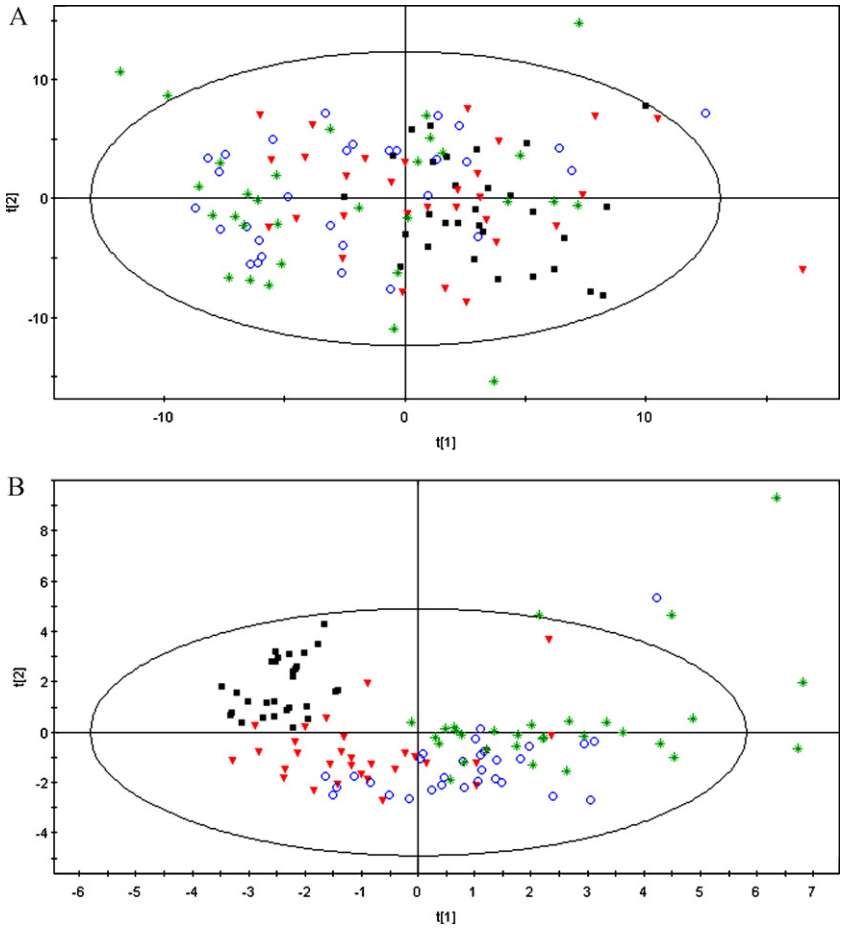

Fig. 2. Score plots of PCA based on (A) original data and (B) 34 selected ion features. N (*box*), CHB (*inverted triangle*), CIR (*circle*) and HCC (*star*) are displayed.

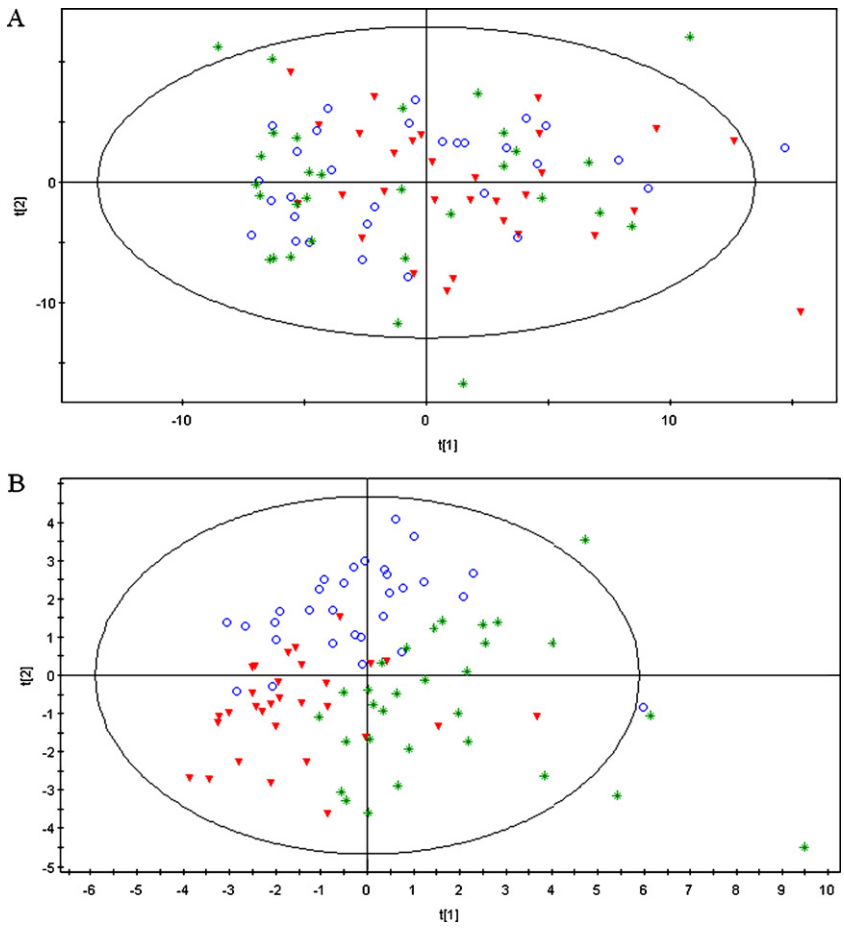

Fig. 3. Score plots of PCA based on (A) original data and (B) 33 selected ion features. CHB (*inverted triangle*), CIR (*circle*) and HCC (*star*) are displayed.

explained by the need of overproduction of carnitine by tumor. Pimelylcarnitine is elevated in HCC compared to CIR, with significant differences compared with the contents in CIR and CHB. Pimelylcarnitine (Fig. 5D) is an ester form of pimelic acid, a dicarboxylic acid [38]. It is reported that most of the carbon atoms of essential biotin are derived from pimelic acid. The detailed relationship of pimelylcarnitine level to cancer metabolism needs further research.

Phospholipids are important ingredients of cell membrane. The alteration of phospholipids content may influence the membrane fluidity, whose increase may link to hepatocyte regeneration and carcinogenesis [39,40]. In comparison to healthy control, the level of PE (22:6/16:0) (Fig. 5E) and PE (20:4/18:0) (Fig. 5F) in HCC are significantly increased. The increase of PE and cholesterol in cell membrane has been reported mainly causing the increase of membrane fluidity in regenerating liver and hepatocyte nodules of rats [41]. And the increase of polyunsaturated fatty acid in PE may affect apoptosis through attenuating free polyunsaturated fatty acid level, while the increase in fatty acid unsaturation of PE doesn't affect membrane fluidity directly [41].

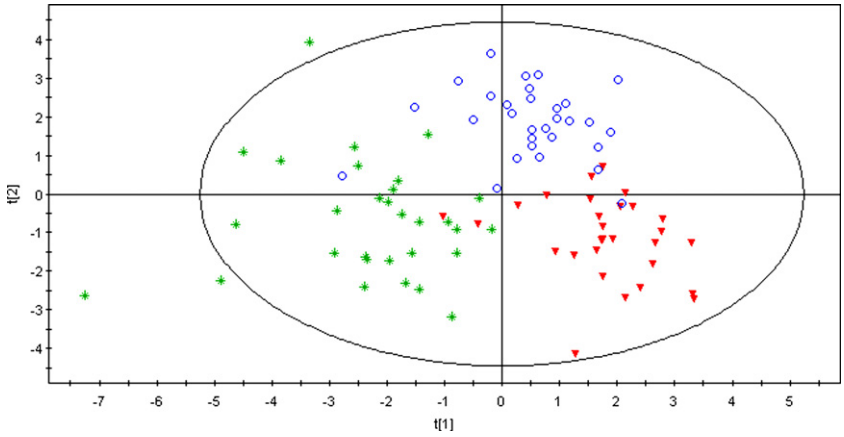

Fig. 4. Score plot of PLS-DA based on 33 selected ion features. CHB (*inverted triangle*), CIR (*circle*) and HCC (*star*) are displayed.

Table 3
17 identified ion features and *p* values in different classes.

tR (min)	MASS	Metabolite	N *vs* M	CHB *vs* CIR	CHB *vs* HCC	CIR *vs* HCC
1.019	362.21	Cortisol	3.18e−01	7.41e−06**	5.35e−01	6.44e−03**
7.781	626.53	PE (20:4/18:0) fragment	2.65e−01	2.97e−02*	5.65e−01	1.08e−02*
7.849	763.52	PE (22:6/16:0)	2.17e−03**	2.40e−03**	3.32e−01	1.02e−01
10.592	447.33	C18:1-CN $[M+Na]^{+1}$	1.67e−02*	3.56e−03**	7.98e−02	3.48e−01
11.042	397.33	C16:1-CN	5.78e−02	6.89e−07**	5.16e−03**	5.50e−01
11.475	817.51	PC (18:2/18:3) $[M+K]^{+1}$	4.79e−01	4.76e−09**	1.12e−06**	1.62e−01
11.509	765.56	PC (P-16:0/20:4)	1.85e−01	4.13e−03**	2.17e−02*	7.71e−02
11.852	769.51	PC (16:1/16:0) $[M+K]^{+1}$	5.02e−08**	6.71e−06**	7.67e−04**	7.29e−02
11.889	733.56	PC (16:0/16:0)	7.26e−09**	3.85e−05**	2.16e−05**	9.55e−01
12.018	705.53	PC (16:0/14:0)	2.45e−08**	3.92e−02*	2.12e−01	3.18e−01
12.146	256.17	C10-CN fragment	6.27e−09**	8.60e−04**	4.63e−02*	1.24e−01
13.676	798.66	SM (d18:0/22:2(13Z,16Z)(OH))	8.48e−02	3.89e−01	5.25e−02	1.11e−02*
14.545	541.31	LysoPC (18:2) $[M+Na]^{+1}$	1.21e−11**	5.03e−01	4.25e−01	1.73e−01
15.116	303.20	Pimelylcarnitine	1.83e−03**	3.48e−01	3.29e−05**	3.14e−04**
16.209	143.09	L-Acetylcarnitine fragment	4.39e−01	2.68e−01	1.08e−04**	3.24e−03**
16.919	102.03	L-Carnitine fragment	9.97e−03**	5.51e−02	1.58e−01	5.56e−03**
16.922	101.08	L-Carnitine fragment	9.75e−03**	2.78e−02*	1.36e−01	2.58e−03**

* *t*-test $p < 0.05$.
** *t*-test $p < 0.01$.

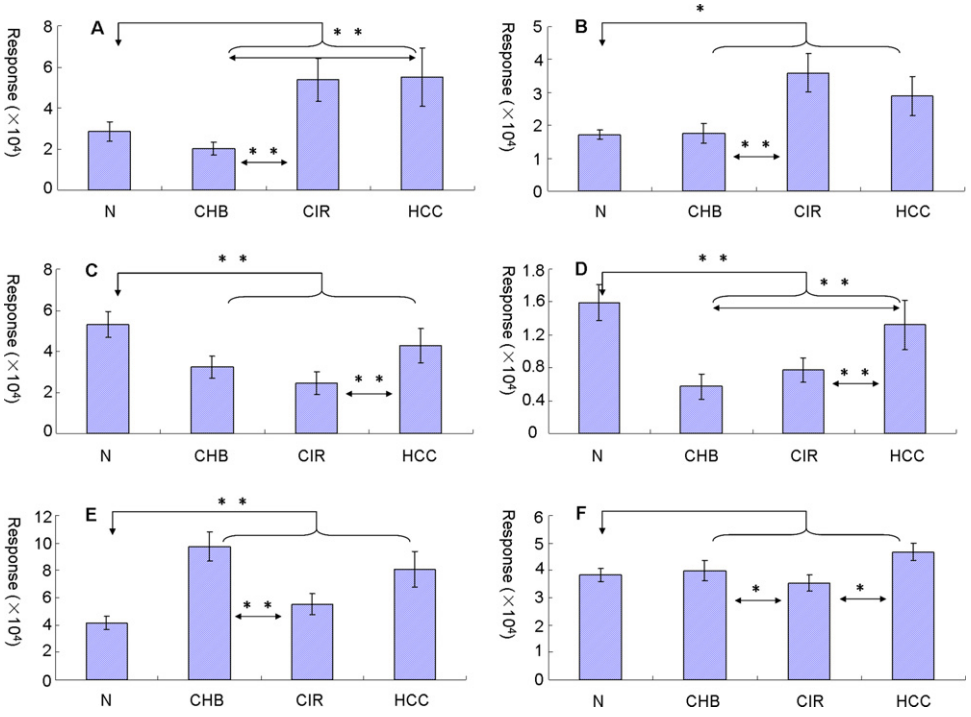

Fig. 5. Six important ion features selected. (A) C16:1-CN; (B) C18:1-CN $[M+Na]^{+1}$; (C) Carnitine fragment; (D) Pimelylcarnitine; (E) PE(22:6/16:0); (F) PE(20:4/18:0) fragment. The error bar was ±SEM of the mean. The significance level was labbled with one asterisk ($p < 0.05$) or two asterisks ($p < 0.01$).

5. Conclusion

In this paper, we proposed a MI-SVM-RFE method which combines the artificial variables and mutual information to filter out the noisy variables from the high dimension metabolome data, and then selects the most discriminative ion features by SVM-RFE. A liver disease metabolome data set from LC–MS was used to validate our new method. The ten-fold cross validation showed that after removing the apparent non-related information, the weight of the features could be calculated more accurately, the average accuracy, sensitivity and specificity were improved.

Acknowledgments

The study has been supported by the State Key Science & Technology Project for Infectious Diseases (2012ZX10002011), the key foundation (No. 20835006) and the creative research group project (No. 21021004) from National Natural Science Foundation of China.

References

[1] O. Fiehn, Plant Mol. Biol. 48 (2002) 155.
[2] R. Madsen, T. Lundstedt, J. Trygg, Anal. Chim. Acta 659 (2010) 23.
[3] J.R. Marchesi, E. Holmes, F. Khan, S. Kochhar, P. Scanlan, F. Shanahan, I.D. Wilson, Y.L. Wang, J. Proteome Res. 6 (2007) 546.
[4] H.C. Keun, T.J. Athersuch, Pharmacogenomics 8 (2007) 731.
[5] J. Chen, W.Z. Wang, S. Lv, P.Y. Yin, X.J. Zhao, X. Lu, F.X.Zhang, G.W. Xu, Anal. Chim. Acta 650 (2009) 3.
[6] T.W.M. Fan, L.L. Bandura, R.M. Higashi, A.N. Lane, Metabolomics 1 (2005) 325.
[7] C.A. Shen, M.M. Shen, Nature 457 (2009) 799.
[8] K.K. Pasikanti, K. Esuvaranathan, P.C. Ho, R. Mahendran, R. Kamaraj, Q.H. Wu, E. Chiong, E.C.Y. Chan, J. Proteome Res. 9 (2010) 2988.
[9] Y. Saeys, I. Inza, P. Larrañaga, Bioinformatics 23 (2007) 2507.
[10] I. Guyon, J. Weston, S. Barnhill, Mach. Learn. 46 (2002) 389.
[11] M. Pirooznia, J.Y. Yang, M.Q. Yang, Y.P. Deng, BMC Genomics 9 (2008) s13.

[12] Y.C. Tang, Y.Q. Zhang, Z. Huang, IEEE/ACM Trans. Comput. Biol. Bioinform. 4 (2007) 365.
[13] M.West-Nielsen, E.V.H. Øgdall, E. Marchiori, C.K.H. Øgdall, C. chou, N.H.H. Heegaard, Anal. Chem. 77 (2005) 5114.
[14] S. Mahadevan, S.L. Shah, T.J. Marrie, C.M. Slupsky, Anal. Chem. 80 (2008) 7562.
[15] C. Furlanello, M. Serafini, S. Merler, G. Jurman, BMC Bioinformatics 4 (2003) 54.
[16] Y.Y. Ding, D. Wilkins, BMC Bioinformatics 7 (2006) s12.
[17] X.G. Zhang, X. Lu, Q. Shi, X.Q. Xu, H.E. Leung, L.N. Harris, J.D. Iglehart, A. Miron, J.S. Liu, W.H. Wong, BMC Bioinformatics 7 (2006) 197.
[18] X.H. Lin, Y. Zhang, G.Z. Ye, X. Li, P.Y. Yin, Q. Ruan, G.W. Xu, J. Sep. Sci. 34 (2011) 3029.
[19] H. Chai, C. Domeniconi, Proceedings of the Second European Workshop on Data Mining and Text Mining in Bioinformatics, Pisa, 2004, p. 7.
[20] X. Zhou, D.P. Tuck, Bioinformatics 23 (2007) 1106.
[21] E. Tuv, A. Borisov, K. Torkkola, International Joint Conference on Neural Networks, Vancouver, 2006, p. 2181.
[22] V.A. Huynh-Thu, L. Wehenkel, P. Geurts, JMLR: Workshop and Conference Proceedings, vol. 4, 2008, p. 61.
[23] H.W. Liu, J.G. Sun, L. Liu, H.J. Zhang, Pattern Recognit. 42 (2009) 1330.
[24] P.Y. Yin, D.F. Wan, C.X. Zhao, J. Chen, X.J. Zhao, W.Z. Wang, X. Lu, S.L. Yang, J.R. Gu, G.W. Xu, Mol. BioSyst. 5 (2009) 868.
[25] H.G. Gika, G.A. Theodoridis, J.E. Wingate, I.D. Wilson, J. Proteome Res. 6 (2007) 3291.
[26] A.K. Smilde, M.J. van der Werf, S. Bijlsma, B.J.C. van der Werff-van-der Vat, R.H. Jellema, Anal. Chem. 77 (2005) 6729.
[27] P.J. García-Laencina, J.L. Sancho-Gómez, A.R. Figueiras-Vidal, Neural Comput. Appl. 19 (2010) 263.
[28] R. Latkowski, M. Mikołajczyk, Lect. Notes Comput. Sci. 3100 (2004) 299.
[29] J. Chen, X.J. Zhao, J. Fritsche, P.Y. Yin, P. Schmitt-Kopplin, W.Z. Wang, X. Lu, H.U. Haring, E.D. Schleicher, R. Lehmann, G.W. Xu, Anal. Chem. 80 (2008) 1280.
[30] U.N. Das, N. Madhavi, G. Sravan Kumar, M. Padma, P. Sangeetha, Prostaglandins. Leukot. Essent. Fatty Acids 58 (1998) 39.
[31] L.N. Zhou, Q.C. Wang, P.Y. Yin, W.B. Xing, Z.M. Wu, S.L. Chen, X. Lu, Y. Zhang, X.H. Lin, G.W. Xu, Anal. Bioanal Chem. 403 (2012) 203.
[32] F.M. Vaz, R.J.A. Wanders, Biochem. J. 361 (2002) 417.
[33] M.I.F. Shariff, N.G. Ladep, I.J. Cox, H.R.T. Williams, E. Okeke, A. Malu, A.V. Thillainayagam, M.E. Crossey, S.A. Khan, H.C. Thomas, S.D. Taylor-Robinson, J. Proteome Res. 9 (2010) 1096.
[34] M.I.F. Shariff, A.I. Gomaa, I.J. Cox, M. Patel, H.R.T. Williams, M.M.E. Crossey, A.V. Thillainayagam, H.C. Thomas, I. Waked, S.A. Khan, S.D. Taylor-Robinson, J. Proteome Res. 10 (2010) 1828.
[35] T. Furuno, T. Kanno, K. Arita, M. Asami, T. Utsumi, Y. Doi, M. Inoue, K. Utsumi, Biochem. Pharmacol. 62 (2001) 1037.
[36] G. Therrien, C. Rose, J. Butterworth, R.F. Butterworth, Hepatology 25 (1997) 551.
[37] B.J. Chang, M. Nishikawa, S. Nishiguchi, M. Inoue, Int. J. Cancer 113 (2005) 719.
[38] S. Lin, R.E. Hanson, J.E. Cronan, Nat. Chem. Biol. 6 (2010) 682.
[39] S.M. Mahler, P.A. Wilce, B.C. Shanley, Int. J. Biochem. 20 (1988) 613.
[40] S.M. Mahler, P.A. Wilce, B.C. Shanley, Int. J. Biochem. 20 (1988) 605.
[41] S. Abel, C.M. Smuts, C. de Villiers, W.C.A. Gelderblom, Carcinogenesis 22 (2001) 795.