

Profiling of mass spectrometry data for ovarian cancer detection using negative correlation learning

Shan He, Huanhuan Chen, Xiaoli Li, and Xin Yao

The Centre of Excellence for Research in Computational Intelligence and Application (Cercia)
School of Computer Science
University of Birmingham
Birmingham, B15 2TT, United Kingdom

Abstract. This paper proposes a novel Mass Spectrometry data profiling method for ovarian cancer detection based on negative correlation learning (NCL). A modified Smoothed Nonlinear Energy Operator (SNEO) and correlation-based peak selection were applied to detected informative peaks for NCL to build a prediction model. In order to evaluate the performance of this novel method without bias, we employed randomization techniques by dividing the data set into testing set and training set to test the whole procedure for many times over. The classification performance of the proposed approach compared favorably with six machine learning algorithms.

Key words: negative correlation learning, bioinformatics, proteomics, data mining

1 Introduction

Ovarian cancer is aggressive: it is rarely detected in early stage and when detected in late stages, e.g., stage III and beyond, the 5-year survival rate is only approximately 15% [6]. Detection of early-stage ovarian cancer can reduce the death rate significantly. For example, the reported 5-year survival rate is about 90% for those women detected in stage I. Cancer antigen 125 (CA125) has been introduced for cancer diagnosis [21]. However, the accuracy for early-stage cancer diagnosis is very low (about 10%) and is prone to large false positive rate.

In recent year, Mass Spectrometry (MS) as a proteomics tool is applied for early-stage ovarian cancer diagnosis. This new proteomics tool is simple, inexpensive and minimally invasive [20]. The first application of MS to the early-stage ovarian cancer diagnosis was done by Petricoin [17]. The author employed genetic algorithms (GAs) coupled with clustering analysis to generate diagnosis rule sets to predict ovarian cancer. The study was based on the SELDI-TOF (Surface-enhanced Laser Desorption/Ionization Time-Of-Flight) low-resolution MS data. With the advance of the mass spectrometry technology, high-resolution SELDI-TOF was employed and studied by the same authors to discriminate ovarian cancer from normal tissue. This dataset is collected with extensive quality control and assurance (QC/QA) analysis which are supposed to have superior classification patterns when compared to those collected with low-resolution instrumentation [6]. In their paper, the sensitivity and specificity were

claimed to be both almost 100%. However, a reproducing study done by Jerries [12] shows that the performance of the best prediction model generated by their GA only achieved 88% accuracy at 25th percentile and 93% accuracy at 75th percentile.

Recently, in attempt to improve the accuracy of identifying cancer on the high-resolution SELDI-TOF ovarian cancer data, Yu et.al. [19] proposed a method that consists of Kolmogorov-Smirnov (KS) test, wavelet analysis and Support Vector Machine (SVM). The average sensitivity and specificity are 97.38% and 93.30%. Before the classification using SVM, the proposed method selected 8094 m/z values via KS test and further compressed to a 3382-dimensional vector of approximation coefficients with Discrete Wavelet Transformation (DWT). Although the accuracy achieved by the procedure was improved, the biological interpretability was greatly sacrificed since the 3382-dimensional DWT coefficient vector for classification is not biologically meaningful.

In this paper, we propose a novel MS data profiling method based on a novel ensemble learning technique, Negative Correlation Learning (NCL) for ovarian cancer detection, which can generate more accurate and biologically meaningful results. We firstly employed MS data preprocessing techniques for signal denoising, peak detection and selection proposed in [10]. The selected peaks will be used for NCL to build a prediction model. We compared the classification performance of NCL with Support Vector Machines (SVM), Prediction Analysis for Micro-arrays (PAM), Bagging, and Random Forests (RF). We also compared the proposed method with Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).

The paper is organized as follows: Section 2 describes details of the proposed methods, followed by the detailed setting of our experiments, control parameter selection and experimental results in Section 3. Finally, Section 4 concludes the paper.

2 Peak Detection and Classification Algorithms

The proposed profile method consists of the two major steps: data preprocessing and NCL classification model. In the data preprocessing step, there are five components: 1). data preprocessing; 2). SNEO based peak detection; 3). peak calibration; 4). correlation-based peak selection; 5). peak qualification. In the following subsections, we give details of each step.

2.1 Data Preprocessing

Data preparation As the m/z data points of each original MS spectrum are different, in order to compare different spectra under the same reference and at the same resolution, we homogenize the m/z vector using a resampling algorithm in the MATLAB Bioinformatics Toolbox.

We correct the baseline caused by the chemical noise in the matrix or by ion overloading using the following procedure: 1). estimated the baseline by calculating the minimum value within the width of 50 m/z points for the shifting window and a step size of 50 m/z points; 2). regresses the varying baseline to the window points using a spline approximation; and 3). subtract the resulting baseline from the spectrum. Finally,

each spectrum was normalized by standardizing the area under the curve (AUC) to the median of the whole set of spectrum. The dataset is split for training and testing as detailed in Section 3. We illustrate the baseline correction result in Figure ??

Modified SNEO based Peak Detection Algorithm Smoothed Non-linear Energy Operator (SNEO), or also known as the Smoothed Teager Energy Operator, has been used to detect hidden spikes in EEG and ECG biomedical signal. The method is sensitive to any discontinuity in the signal. It was shown by [16] that the output of SNEO is the instantaneous energy of the high-pass filtered version of a signal. For MS data, true peaks can be regarded as instantaneous changes in the signal. Therefore, the SNEO is ideal for the detection peaks in MS data because of its instantaneous nature. The generalized SNEO Ψ_s is defined as [16]:

$$\Psi_s[x(n)] = \Psi[x(n)] \otimes w(n) \quad (1)$$

$$\Psi[x(n)] = x^2(n) - x(n+j)x(n-j) \quad (2)$$

where \otimes is the convolution operator and $w(n)$ is a smoothing window function; in this study, bartlett window function is used. Usually, the step size j is set to be 1 which gives us a standard SNEO. For the high-resolution MS data, we selected the step size $j = 3$, which gives the best classification results.

Conventionally after applying SNEO to pre-emphasize peaks in the signal, potential peaks are detected using a threshold approach [16]. However, our research indicated that the threshold method is not suitable for MS peak detection since the background noise, e.g., false peaks, is non-stationary and there is no precise knowledge on the energy distribution of the true peaks and background noise. In this study, we replace the threshold detection method by a naive peak finding algorithm: we firstly calculate the first derivatives in the SNEO emphasized signals, then those local maximum will be detected as peaks. Obviously, this naive peak finding algorithm detects true peaks as well as a large number of false peaks. We therefore employ a filter-based peak selection algorithm as detailed in Section 2.1 to filter out false peaks.

Peak calibration After the peak detection, it is necessary to calibrate the peaks in order to alleviate the impact of the m/z axis shifting problem. We divide each spectrum into windows with an equal number of m/z values. The selection of an optimal number of m/z values is done by experiments as detailed in 3.3. At each m/z point in each window, the total number of peaks across all the spectra is calculated. The m/z point that has the highest number of peaks within the window is set as the calibration m/z value. Then the peaks in all spectra within the window are moved to this calibrated m/z point. For each spectrum, if there is more than one peak in the m/z window, only the highest peak will be moved to the calibrated m/z point, all the other lower intensity peaks will be removed. We plot a MS spectrum of a cancer sample and peaks detected and calibrated by our method in Figure 1.

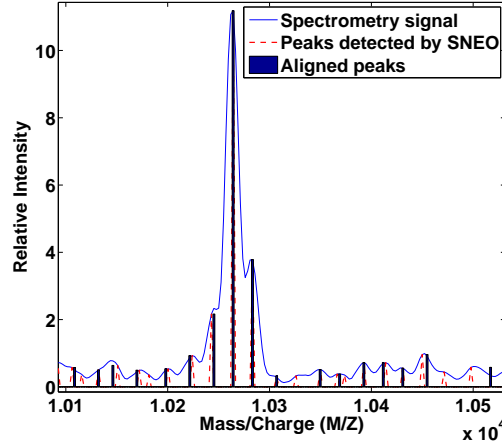


Fig. 1. MS spectrum of an ovarian cancer sample, peaks detected and calibrated by our SNEO peak detection and calibration method.

Correlation-based Peak Selection The correlation-based feature selection [8] uses a correlation based heuristic to determine the usefulness of feature subsets. The usefulness is determined by measuring the “merit” of each individual feature for predicting the class label as well as the level of inter-correlation among them. First, an evaluation function is defined as:

$$G_s = \frac{k\bar{r}_{ci}}{\sqrt{k + k(k-1)\bar{r}_{ii}}} \quad (3)$$

where k is the number of features in the subset; \bar{r}_{ci} is the mean feature correlation with the class, and \bar{r}_{ii} is the average feature intercorrelation.

Equation (3) is the core of the feature selection algorithm. With this evaluation function, heuristic search algorithm then can be applied to search the feature subset with the best merit as measured in Equation (3). In the implementation, a best first heuristic search strategy was used to search the feature subset space in reasonable time. The peak selection algorithm starts from the empty set of features and uses a forward best first search to search an optimal subset. The stopping criterion of five consecutive fully expanded non-improving subsets was used [8].

There are two approaches to measure the correlation between features and the class (r_{ci}), and between features (r_{ii}). One is based on classical linear correlation and the other is based on information theory. The correlation-based feature selection employed here used the information theory based approach since it can capture correlations that are either linear or nonlinear. For details of implementation, please refer to [8].

Peak Qualification After applying the correlation-based peak selection to the detected peaks from the training set, a small set of peaks then can be generated. This set of peaks

will be used as inputs for NCL to build a prediction model. Based on the selected peak set from training set, we construct m/z windows using the same width as used in the calibration step. For each spectrum in test set, peaks detected by SNEO peak detection algorithm is qualified by the constructed m/z windows, that is, only those peaks within the m/z windows will be used as inputs in testing. If there are more than one peak in a m/z window in a spectrum, only the highest peak will be retained.

2.2 Negative Correlation Learning

Algorithm Negative Correlation Learning (NCL)
Input: the training set $\mathbf{D} = \{\mathbf{x}_n, y_n\}_{n=1}^N$, integer M specifying size of ensemble, the learning rate η in backpropagation (BP) algorithm and integer T specifying the number of iterations.
For $t = 1, \dots, T$ **do:**

1. Calculate $f_{ens}(\mathbf{x}_n) = \frac{1}{M} \sum_{i=1}^M f_i(\mathbf{x}_n)$.
2. For each network from $i = 1$ to M do: for each weight $w_{i,j}$ in network i , perform a desired number of updates,

$$e_i = \sum_{n=1}^N (f_i(\mathbf{x}_n) - y_n)^2 - \lambda \sum_{n=1}^N (f_i(\mathbf{x}_n) - f_{ens}(\mathbf{x}_n))^2,$$

$$\frac{\partial e_i}{\partial w_{i,j}} = 2 \sum_{n=1}^N (f_i(\mathbf{x}_n) - y_n) \frac{\partial f_i}{\partial w_{i,j}} - 2\lambda \sum_{n=1}^N (f_i(\mathbf{x}_n) - f_{ens}(\mathbf{x}_n)) \left(1 - \frac{1}{M}\right) \frac{\partial f_i}{\partial w_{i,j}},$$

$$\Delta w_{i,j} = -2\eta \left\{ \sum_{n=1}^N (f_i(\mathbf{x}_n) - y_n) \frac{\partial f_i}{\partial w_{i,j}} - \lambda \sum_{n=1}^N (f_i(\mathbf{x}_n) - f_{ens}(\mathbf{x}_n)) \left(1 - \frac{1}{M}\right) \frac{\partial f_i}{\partial w_{i,j}} \right\}.$$

Output: NCL ensemble

$$f(\mathbf{x}) = \frac{1}{M} \sum_i f_i(\mathbf{x}).$$

Fig. 2. Negative Correlation Learning Algorithm

Ensemble of multiple learning machines, i.e. a group of learners that work together as a committee, has attracted a lot of research interests in the machine learning community because it is considered as a good approach to improve the generalization ability [9].

Negative Correlation Learning (NCL) [14, 13] is a successful ensemble technique and it has shown a huge number of empirical applications [15, 11, 3, 7, 5]. NCL introduces a correlation penalty term into the cost function of each individual network so that each neural network minimizes its MSE error together with the correlation of the ensemble.

Given the training sets $\{\mathbf{x}_n, y_n\}_{n=1}^N$, NCL combines M neural networks $f_i(\mathbf{x})$ to constitute the ensemble.

$$\bar{f}(\mathbf{x}_n) = \frac{1}{M} \sum_{i=1}^M f_i(\mathbf{x}_n). \quad (4)$$

In training network f_i , the cost function e_i for network i is defined by

$$e_i = \sum_{n=1}^N (f_i(\mathbf{x}_n) - y_n)^2 + \lambda p_i, \quad (5)$$

where λ is a weighting parameter on the penalty term p_i :

$$\begin{aligned} p_i &= \sum_{n=1}^N \left\{ (f_i(\mathbf{x}_n) - \bar{f}(\mathbf{x}_n)) \sum_{j \neq i} (f_j(\mathbf{x}_n) - \bar{f}(\mathbf{x}_n)) \right\} \\ &= - \sum_{n=1}^N (f_i(\mathbf{x}_n) - f_{ens}(\mathbf{x}_n))^2. \end{aligned} \quad (6)$$

The first term in the right-hand side of (5) is the empirical training error of network i . The second term p_i is a correlation penalty function. The purpose of minimizing p_i is to negatively correlate each network's error with errors for the rest of the ensemble. The λ parameter controls a trade-off between the training error term and the penalty term. With $\lambda = 0$, we would have an ensemble with each network training with plain back propagation, exactly equivalent to training a set of networks independently of one another. If λ is increased, more and more emphasis would be placed on minimizing the penalty. The algorithm is summarized in Figure 2.

For the NCL model, we used radial basis function (RBF) networks as base classifiers. The training of RBF network is separated into two steps. In the first step, the means μ_k are initialized with randomly selected data points from the training set and the variances σ_k are determined as the Euclidean distance between μ_k and the closest μ_i ($i \neq k, i \in \{1, \dots, K\}$). Then in the second step we perform gradient descent in the regularized error function (weight decay)

$$\min e = \frac{1}{2} \sum_{n=1}^N (y_n - f(\mathbf{x}_n))^2 + \alpha \sum_{k=1}^K w_k^2. \quad (7)$$

In order to fine-tune the centers and widths, we simultaneously adjust the output weights, the RBF centers and variances. Taking the derivative of Equation (7) with respect to RBF means μ_k and variances σ_k^2 we obtain

$$\frac{\partial e}{\partial \mu_k} = \sum_{n=1}^N (f(\mathbf{x}_n) - y_n) \frac{\partial f(\mathbf{x}_n)}{\partial \mu_k}, \quad (8)$$

with $\frac{\partial f(\mathbf{x}_n)}{\partial \mu_k} = w_k \frac{\mathbf{x}_n - \mu_k}{\sigma_k^2} \phi_k(\mathbf{x}_n)$ and

$$\frac{\partial e}{\partial \sigma_k} = \sum_{n=1}^N (f(\mathbf{x}_n) - y_n) \frac{\partial f(\mathbf{x}_n)}{\partial \sigma_k}, \quad (9)$$

with $\frac{\partial f(\mathbf{x}_n)}{\partial \sigma_k} = w_k \frac{\|\mathbf{x} - \mu_k\|^2}{\sigma_k^3} \phi_k(\mathbf{x}_n)$. These two derivatives are employed in the minimization of Equation (7) by scaled conjugate gradient descent, where we always compute the optimal output weights in every evaluation of the error function. The optimal output weights \mathbf{w} can be computed in closed form by

$$\mathbf{w} = (\Phi^T \Phi + \alpha I)^{-1} \Phi^T \mathbf{y}, \quad (10)$$

where $\mathbf{y} = (y_1, \dots, y_n)^T$ denotes the output vector, and I an identity matrix.

3 Numerical Experiments

3.1 Datasets

The SELDI-TOF high resolution ovarian cancer dataset OC-WCX2-HR was collected by NCI-FDA using a hybrid quadruple time-of-flight spectrometer with extensive quality control and assurance (QC/QA) analysis, which are supposed to have superior classification patterns when compared to those collected with low-resolution instrumentation. The dataset consists of 216 samples, of which 95 control and 121 cancer. Each spectrum contains 350,000 m/z values. We condensed the spectrum into 7064 m/z values following [6].

3.2 Experimental Setting

The objective of our experiments is to assess the classification performance of our proposed method. In order to evaluate the classification performance of the proposed method with minimal bias, we employed the same experimental setting used in [12], which is also similar to the “proportional validation” in [19] by randomly splitting all the datasets into a training set and a test set. 52 control samples and 53 cancer samples from the OC-WCX2-HR dataset were selected for training data, the rest 43 control samples and 68 cancer samples were set aside for evaluation as a blind dataset. These same settings were used in [6] and [12].

For comparison purpose, we ran experiments on the dataset with Bagging [1], RF [2], SVM [4], PAM [18], LDA and QDA. We investigated the classification performance of these machine learning algorithms in comparison with NCL on the same peak set extracted by SNEO peak detection and correlation based peak selection algorithms.

We first carried out preliminary experiments to select optimal parameters for each algorithm as detailed in Section 3.3. Then based on the optimal parameters, we executed the experiments to evaluate the classification performance.

3.3 Parameter Selection

For the SNEO peak detection method, there is no tunable parameter. However, in the peak calibration step, the calibration window width is adjustable. Following [10], we selected calibration window width of 10, which generated the best results.

Table 1. Test set accuracy (%) percentiles of the dataset OC-WCX2-HR from 50 runs of NCL and 6 other machine learning algorithms. The 6 machine learning algorithms used the peaks selected by the proposed SNEO peak detection and correlation based selection method.

| Algorithm | Test set accuracy 25th | | | Test set accuracy 75th | | |
|-----------|------------------------|--------------|--------------|------------------------|--------------|--------------|
| | Overall | Sensitivity | Specificity | Overall | Sensitivity | Specificity |
| NCL | 93.69 | 93.83 | 92.48 | 97.39 | 98.30 | 97.39 |
| PAM | 90.99 | 86.78 | 94.67 | 94.60 | 92.30 | 97.99 |
| RF | 91.89 | 89.28 | 91.07 | 96.40 | 94.64 | 98.27 |
| SVM | 92.79 | 94.83 | 87.75 | 95.50 | 98.31 | 96.15 |
| Bagging | 90.54 | 93.10 | 87.23 | 95.04 | 96.28 | 94.23 |
| LDA | 88.29 | 82.76 | 92.45 | 93.68 | 95.00 | 98.21 |
| QDA | 91.89 | 89.47 | 90.90 | 95.49 | 96.61 | 97.92 |

For Random Forests classifier, we tuned the following parameters: the number of candidate variables for each split and the minimum size of terminal nodes. We search the grid $\{1, 2, \dots, 8\} \times \{100, 200, \dots, 500\}$ on the training data.

The shrinkage parameter of PAM was selected by 10-fold cross-validation as suggested by [18].

For SVM we used C -classification with RBF kernel. In order to select optimal parameters for SVM, e.g., kernel parameter σ and cost C , we executed grid search on $\{2^{-10}, 2^{-9}, \dots, 2^5\} \times \{2^{-5}, 2^{-4}, \dots, 2^{10}\}$ by 10-fold cross-validation.

In the Bagging algorithm, 100 classification and regression trees are grown in each Bagging ensemble.

We use the traditional Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). In matlab, the LDA and QDA are called by the classify function.

The number of hidden nodes in RBFs of the NCL model is randomly selected but restricted in the range of 5 to 15. The ensemble consists of 100 RBF networks.

3.4 Experimental Results

In total, 128 peaks were detected and selected as biomarkers. These peaks were then used for NCL to build a prediction model.

The overall average accuracy obtained by our method from 50 runs is 95.16% with a standard deviation of 2.75%. The average sensitivity and specificity are 96.21% with a standard deviation of 2.83% and 93.96% with a standard deviation of 3.9%, respectively.

The results obtained from 6 other machine learning methods on the same training set and test set are presented in Table 1. It can be seen from the table that the SVM and RF algorithms generated better results than the other machine learning algorithms but still worse than the results generated by NCL.

Jeffries et al. [12] employed GA coupled with clustering analysis on the same dataset, the overall average accuracy from 50 runs of the GA was only 88% and 93% at 25 and 75 percentiles, respectively, which is far worse than the results obtained by our

proposed method. Apart from its poorer accuracy, the GA based method actually fall into the whole-spectrum method since the method treated each m/z point in the spectrum as a separate test. The outputted biomarkers were a set of significant m/z values but were not necessarily a peak set. Therefore, the biological interpretation of their results is not guaranteed.

In [19], the average sensitivity and specificity were improved to 97.47% and 93.35% in 1000 independent 2-fold proportional validation using SVM. These results are slightly better than the results generated by our method. However, in their study, KS test and Discrete Wavelet Transform (DWT) were employed to reduce the dimensionality of the data. The biological interpretability of this method was greatly sacrificed since the features used for classification were a set of coefficients of DWT, which are even less biologically meaningful than a set of m/z values.

4 Conclusion

In this study, we have propose a novel MS data profiling method for ovarian cancer detection based on a novel ensemble method, Negative Correlation Learning (NCL). To our best knowledge, it is the first time NCL have been applied to proteomics.

In order to assess the classification performance of the proposed method, we evaluated our method on one high resolution ovarian cancer dataset using the same experimental settings used in [6] and [12]. We compared the proposed profiling method with six machine learning algorithms. The experimental results show that the proposed method can generate excellent classification accuracy. Results from our method are also better than most of the results in the literature, even some whole-spectrum methods. The most notable merit of our proposed method is that, besides its excellent classification performance, it obtains more biologically meaningful results, e.g., a parsimonious set of peaks, for further study and validation.

Acknowledgment

This work is supported by the Leverhulme Trust Early Career Fellowship (ECF/2007/0433) awarded to Dr. Shan He.

References

1. L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.
2. L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.
3. G. Brown, J. Wyatt, and P. Tiño. Managing diversity in regression ensembles. *Journal of Machine Learning Research*, 6:1621–1650, 2005.
4. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining And Knowledge Discovery*, 2:121–167, 1998.
5. H. Chen and X. Yao. Evolutionary random neural ensemble based on negative correlation learning. In *Proceedings of The 2007 IEEE Congress on Evolutionary Computation (CEC'07)*, pages 1468–1474, 2007.

6. T.P. Conrads, V.A. Fusaro, S. Ross, D. Johann, V. Rajapakse, B.A. Hitt, S.M. Steinberg, E.C. Kohn, and D.A. Fishman. High-resolution serum proteomic features for ovarian cancer detection. *Endocr. Relat. Cancer*, 11(2):163–178, 2004.
7. N. García, C. Hervás, and D. Ortiz. Cooperative coevolution of artificial neural network ensembles for pattern classification. *IEEE Transactions on Evolutionary Computation*, 9(3):271–302, 2005.
8. M.A. Hall. *Correlation-based Feature Selection for Machine Learning*. PhD thesis, The University of Waikato, 1999.
9. L. K. Hansen and P. Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):993–1001, 1990.
10. S. He and X.L. Li. Profiling of high-throughput mass spectrometry data for ovarian cancer detection. In *8th Int.Conf.on Intelligent Data Eng. ,Automated Learning*, volume 4881 of *Lecture Notes in Computer Science*, pages 860–869. Springer, 2007.
11. M. M. Islam, X. Yao, and K. Murase. A constructive algorithm for training cooperative neural network ensembles. *IEEE Transaction on Neural Networks*, 14(4):820–834, 2003.
12. N.O. Jeffries. Performance of a genetic algorithm for mass spectrometry proteomics. *BMC Bioinformatics*, 5(1):180, 2004.
13. Y. Liu and X. Yao. Ensemble learning via negative correlation. *Neural Networks*, 12(10):1399–1404, 1999.
14. Y. Liu and X. Yao. Simultaneous training of negatively correlated neural networks in an ensemble. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 29(6):716–725, 1999.
15. Y. Liu, X. Yao, and T. Higuchi. Evolutionary ensembles with negative correlation learning. *IEEE Transaction on Evolutionary Computation*, 4(4):380–387, 2000.
16. S. Mukhopadhyay and G.C. Ray. A new interpretation of nonlinear energy operator and its efficacy in spike detection. *IEEE Transactions Biomedical Engineering*, 45(2):180–187, 1998.
17. E.F. Petricoin, A. Ardekani, B. Hitt, P. Levine, V. Fusaro, S. Steinberg, G. Mills, C. Simone, D. Fishman, and E. Kohn. Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet*, 359:572–577, 2002.
18. R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. USA*, 99:6567–6572, 2002.
19. J.S. Yu, S. Ongarello, R. Fieldler, X.W. Chen, G. Toffolo, C. Cobelli, and Z. Trajanoski. Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data. *Bioinformatics*, 21(10):2200–2209, 2005.
20. X. Zhang, D. Wei, Y. Yap, L. Li, S. Guo, and F. Chen. Mass spectrometry-based “omics” technologies in cancer diagnostics. *Mass Spectrometry Reviews*, 26:403–431, 2007.
21. V.R. Zurawski, H. Orjaseter, A. Andersen, and E. Jellum. Elevated serum ca 125 levels prior to diagnosis of ovarian neoplasia: relevance for early detection of ovarian cancer. *Int. J. Cancer*, 42(5):677–680, 1988.