Integrating Discrete Wavelet Transform and Neural Networks for Prostate Cancer Detection Using Proteomic Data

Grace J. Hwang[1] Chuan-Ching Huang[1] Ta Jen Chen[1] Jack C. Yue[2]
Yuan-chin Ivan Chang[3] Bao-Ling Adam[4]
[1]*Department of Computer Science and Information Engineering, Fu Jen University, Taiwan*
[2]*Department of Statistics, National Chengchi University, Taiwan*
[3]*Institute of Statistical Science, Academia Sinica, Taiwan*
[4]*Center for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta, Georgia*
Email : jihwang@csie.fju.edu.tw, sense92@csie.fju.edu.tw, a8851002@stmail.fju.edu.tw, csyue@nccu.edu.tw
ycchang@stat.sinica.edu.tw, badam@mcg.edu

ABSTRACT: An integrated approach for prostate cancer detection using proteomic data is presented. Due to the high-dimensional feature of proteomic data, the discrete wavelet transform (DWT) is used in the first-stage for data reduction as well as noise removal. After the process of DWT, the dimensionality is reduced from 43,556 to 1,599. Thus, each sample of proteomic data can be represented by 1599 wavelet coefficients. In the second stage, a voting method is used to select a common set of wavelet coefficients for all samples together. This produces a 987-dimension subspace of wavelet coefficients. In the third stage, the Autoassociator algorithm reduces the dimensionality from 987 to 400. Finally, the artificial neural network (ANN) is applied on the 400-dimension space for prostate cancer detection. The integrated approach is examined on 9 categories of 2-class experiments, and also 3- and 4-class experiments. All of the experiments were run 10 times of ten-fold cross-validation (*i.e.* 10 partitions with 100 runs). For 9 categories of 2-class experiments, the average testing accuracies are between 81% and 96%, and the average testing accuracies of 3- and 4-way classifications are 85% and 84%, respectively. The integrated approach achieves exciting results for the early detection and diagnosis of prostate cancer.

1 INTRODUCION

The advent of mass spectrometry (MS) has driven the recent progress in the proteomic pattern analysis. With the serum proteomic spectra generated from MS technology, it is now possible to utilize the proteomic pattern for the early diagnosis of diseases such as prostate cancer. The detection of human tumors using mass spectral data is a challenging task, one of the major difficulties is the high-dimensional characteristic of proteomic data, and the other problem is the noise appeared in the data spectra. Since wavelets are being utilized in many applications, including data compression and noise removal. This leads to our motivation to apply the wavelet transform to conquer the difficulties.

For data classification, artificial neural network (ANN) is one of the basic techniques and has been successfully applied to problems such as pattern recognition, speech recognition, and handwriting recognition. ANN is a powerful method for the analysis of complex data; in particular, when data are high-dimensional continuous values. This makes it highly suitable for the study of proteomic data. Therefore, a multi-layer ANN with the Backpropagation algorithm is used for the classification of prostate cancer.

In this study, we propose two integrated approaches for prostate cancer detection. One is a two-stage approach, which is a hybrid method of DWT and ANN. The other one is a four-stage approach, which refines the two-stage one. In the first stage, we applied DWT on MS data for data preprocessing. Next, we use a voting method to select common wavelet coefficients from all samples in the set. Third, we utilize the Autoassociator algorithm for further data reduction. Finally, we apply ANN on the resulting data to classify different disease status of prostate cancer.

The reminder of this paper is organized as follows: in next section, a brief description of the prostate cancer dataset and its past usage are given. The use of DWT, Autoassociaor algorithm, and ANN as an integrated approach for prostate cancer detection is explained in section 3. The computational experiments to verify the integrated approach are reported in section 4. Finally, some concluding remarks and the future work are presented in section 5.

2 DATA SET

2.1 Description of dataset

The prostate cancer data was obtained from Eastern Virginia Medical School, Norfolk, Virginia, and the data was generated using a Surface Enhanced Laser Desorption/Ionization Time-of Flight Mass spectrometer (SELDI TOF MS). The data set contains four groups of patients (see Table 1) with the following characteristics.

- There are 652 spectra from 326 patients (each patient has two replicates) with different disease status: Age matched Normal (NO), Prostate cancer stage A & B (CAB; early stage), Prostate Cancer Stage C& D (CCD; late stage), and Benign prostate hyperplasia (BPH).

Group	Sample size	Description
NO	82	Age matched Normal (ages > 50)
CAB	84	Early stage of cancer
CCD	83	Late stage of cancer
BPH	77	Benign prostate hyperplasia

Table 1: Description of patient groups in the prostate cancer data set

- There are approximately 48,000 (x, y) data points in each spectrum, where x is the protein mass (Dalton) and y is the relative intensity of the ion measurement. This is the data set used for our integrated approach study. We referred it as raw data.
- The range of x is generally between 0 to 200,000 Daltons (1 Dalton = 1 atomic mass unit).
- A pre-processed data set is also available. Spectra in the set were manually preprocessed and stored as vectors of dimension 779. This was the data set we used for ANN analysis only without using DWT and the results of the study was published in [1]. It is referred as pre-processed data.

2.2 Previous work

The dataset was first used in [2], Adam et al. developed a decision tree on the pre-processed data and used a single training/testing split to verify the algorithm. A sensitivity (true positive ratio) of 83%, a specificity (true negative ratio) of 97% and a positive predict value of 96% were achieved when comparing prostate cancer (CAB + CCD) versus non-cancer (BPH + NO).

Lillien et al. developed an algorithm, Q5, which used principal component analysis and linear discriminant analysis followed by probabilistic classification [3]. The authors applied Q5 on the pre-processed data for 2-class experiments (CAB + CCD vs. BPH + NO) and 3-class experiments (CAB + CCD / BPH / NO). Since Q5 is a probabilistic classification, there is a trade off between confidence in classification and the number of samples classified. For example, Q5 was able to classify 85.6% of the samples with a positive predictive value of 94.3%, a sensitivity of 91.3%, and a specificity of 93.0% in the 2-class experiments. Q5 was also used in 3-class experiments and its positive predictive value was 96.1% when 92.0% of the samples were classified.

Qu et al. [4] presented a method of data reduction, which used a wavelet transform in discriminant analysis for very high dimensionality data. The method was applied to the raw dataset for prostate cancer detection. The computational experiment only used a subset of the raw dataset (without BPH patients) for binary classification (CAB + CCD vs. NO), and the result showed a sensitivity of 97% and specificity 100% using a separate test set.

Yasui et al. [5] presented a data-analytic strategy for the raw dataset. After the pre-analysis processing of the data, the boosting algorithm was applied to distinguish cancer and benign hyperplasia (CAB + CCD + BPH) from normal (NO). For a stopping rule of 100% sensitivity and specificity in the training data, the empirical sensitivity and specificity on a separate test data were 97.8% and 100%, respectively. However, the performance was not satisfactory for the cancer (CAB + CCD) vs. BPH classification. For a stopping rule of $\geq$ 90% sensitivity and specificity in the training data, the empirical sensitivity and specificity in the data set were 93.3% and 46.7%, respectively.

Our research group compared the performance of several classification algorithms for MS data analysis [1]. These algorithms included ANN, support vector machine (SVM), and the classification and regression tree (CART). The algorithms were applied to the pre-processed dataset. We found that ANN outperformed other algorithms in the analysis of MS data. For each binary classification between two different groups, the average testing accuracies of 100 simulations were between 92% and 97% except the case of CCD vs. CAB, which was 85%. In the 4-class experiment, the average testing accuracy was 91%. All the simulations used 90% data for training and the remaining 10% data for testing.

3 METHODS

The data analysis approach here is a hybrid method, which integrates the techniques of DWT, Autoassociator, and ANN. DWT is utilized in the first step for data compression and noise removal, and followed by Autoassociator algorithm to do further dimension reduction. After the data preprocessing of DWT and Autoassociator, ANN is finally applied for classifying disease status of prostate cancer.

3.1 Discrete wave transform

Due to the characteristic and limitation of the SELDI technology, the data points below 2 KDa and above 20 KDa were eliminated. The data points below 2 KDa consists the matrix peaks consistently and the laser energy used to generate the spectra were not strong enough to ionize the proteins efficiently above 20 KDa. Therefore, There are approximately 43,556 features from each spectrum were used for the study. For each mass spectrum, the 43,556 features can be represented by a set of nonzero wavelet coefficients, selected by a threshold. For different samples, sets of nonzero wavelet coefficients may appear on different dimensions of the wavelet space. We then need to union the wavelet dimensions for each set of sample. In this study, we obtained the dimensionality of 1,599 after the union. Note that, each spectrum now is represented by 1,599 wavelet coefficients. For those dimensions were not originally belong to the sample, the corresponding wavelet coefficients are all set zeros. Thus, the dimensionality of data space reduced from 43,556 to 1,599 after the DWT processing. It is a dramatic data reduction with a compression ratio of 3.7% (1,599 / 43,556).

3.2 Autoassociators

As we will see in section 4 of the computational experiments, when ANN applies on the 1,599-dimension space, the results of binary classifications are quite good. However, the 3- and 4-way experiments can not converge in the ANN stage. This indicates that a further data reduction may be needed. We therefore propose another data reduction approach, autoassociator [6, 7], on a subspace of 1,599-dimension space of wavelet coefficients.

Autoassociator is also called autoencoders, bottlenecks or n-h-n neural network, which is a two-layer (one-hidden layer) perceptron with n inputs, h ($< n$) hidden units and n outputs (see Figure 1), trained with backpropagation typically. When an n x h x n network is trained to replicate the input in the output layer minimizing the squared sum of errors over all of output units, the hidden layer learns to represent intermediate features that are useful for the learning the target values and that are not explicit in the network inputs. For example [6], consider an 8 x 3 x 8 network (see Figure 1), which can be trained to learn the identity function, using the training examples of eight

distinct binary strings shown in the input of Table 2. The three hidden node values encode the eight distinct strings using the encoding shown on the right. We note that the result is the standard binary encoding for eight distinct values, if we round the encoded values to zero or one.

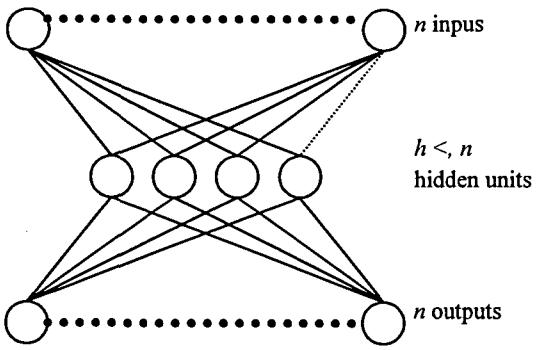

Figure 1: Topology of the autoassociator; n x h x n neural network

Input		Hidden Values				Output
10000000	→	.89	.04	.08	→	10000000
01000000	→	.15	.99	.99	→	01000000
00100000	→	.01	.97	.27	→	00100000
00010000	→	.99	.97	.71	→	00010000
00001000	→	.03	.05	.02	→	00001000
00000100	→	.01	.11	.88	→	00000100
00000010	→	.80	.01	.98	→	00000010
00000001	→	.60	.94	.01	→	00000001

Table 2: An illustrative example of the Autoassociator

3.3 Artificial neural networks

The last stage of our approach is using ANN to classify the disease status of each sample. ANN is a well-known approach for pattern recognition, and mass spectrometry classification is actually an example of pattern recognition. The ANN is also suited to problem in which input corresponds to complex data with noise. It is also applicable to problems have real-valued input. This leads us to use ANN for disease classification. In this study, the Backpropagation algorithm is applied for a two-layer (one hidden layer) ANN learning.

4 EXPERIMENTAL RESULTS

In this section, we report our computational results for two different integrated approaches. The first integrated approach is a two-stage approach which combines DWT and ANN only. Thus we apply DWT on MS data with 43,556 dimensions, and then utilize ANN for classification on the resulting 1,599-dimension space of wavelet coefficients. The second integrated approach is a 4-stage approach, which contains the steps of DWT, voting method, Autoassociator algorithm, and ANN. The first stage is exactly the same as that of the first integrated approach. Thus, after the DWT process, each mass spectrum can be represented by 1,599 wavelet coefficients, in which some of them are zero. If there are quite a number of samples with zero coefficients in some specific dimension, we may delete it without losing too much information. We therefore use a voting method to delete less common dimensions. In our study, we deleted a dimension if there are more than 90% of data samples appearing zero coefficients in that dimension. After this process, we obtain 987 common dimensions. In other words, we use the voting method to choose a 987-dimension subspace of 1,599-dimension space in stage 2. In the following stage, we apply Autoassociator algorithm to do further dimensionality reduction. We use a 987 x 400 x 987 network to reduce the dimension from 987 to 400. In the last stage, ANN is applied on the 400-dimension space to classify disease status of prostate cancer.

To verify these two integrated approaches, 2-, 3-, and 4-way classifications are performed using a ten-fold cross-validation. In each ten-fold cross-validation, the data is divided randomly into ten parts, each part is held out in turn and the learning scheme trained on the remaining nine-tenths, then its test accuracy rate is calculated in the holdout set. Thus the procedure is executed a total of ten times, on different training subsets. Finally, the ten accuracy estimates are averaged to yield an overall accuracy estimate. In our computational tests, all of the experiments are run 10 times of ten-fold cross-validation (*i.e.* 10 partitions with 100 runs).

	Classification	Training Accuracy	Testing Accuracy	Testing Sensitivity	Testing Specificity
1	BPH vs. NO	92.2%±10.3%	91.5%±10.8%	90.3%±19.7%	92.9%±18.7%
2	CAB vs. NO	96.2%±0.1%	89.5%±0.1%	89.9%±0.1%	89.5±0.1%
3	CCD vs. NO	96.5%±2.6%	90.8%±5.9%	90.9%±11.5%	95.2%±12.1%
4	CAB vs. BPH	88.6%±3.8%	76.5%±8.1%	79.5%±13.9%	74.8%±18.3%
5	CCD vs. BPH	88.2%±5.2%	78.2%±8.1%	78.1%±16.7%	78.1%±11.6%
6	CCD vs. CAB	90.1%±2.0%	81.9%±7.7%	79.5%±12.4%	83.5%±10.4%
7	(CAB + CCD) vs. NO	93.1%±1.2%	87.7%±6.4%	91.8%±4.8%	81.3%±12.6%
8	(CAB + CCD) vs. (NO + BPH)	74.3%±4.1%	76.4%±5.5%	74.4%±9.6%	78.6%±10.9%
9	(BPH + CAB + CCD) vs. NO	87.4%±1.2%	85.8%±4.4%	84.6%±2.9%	86.2%±3.2%

Table 3: Examination of two-stage integrated approach using raw data set with cross-validation. Training and testing accuracies (mean ± standard deviation) are calculated from 100 runs; 90% of data used for training of each run

	Classification	Training Accuracy	Testing Accuracy	Testing Sensitivity	Testing Specificity
1	BPH vs. NO	97.6%±9.6%	95.8%±8.6%	97.6%±7.9%	93.9%±9.1%
2	CAB vs. NO	98.9%±2.4%	94.3%±3.6%	96.2%±3.1%	92.3%±3.8%
3	CCD vs. NO	99.1%±3.1%	93.8%±5.2%	95.4%±6.9%	92.1%±4.3%
4	CAB vs. BPH	96.6%±4.6%	82.9%±7.8%	84.9%±7.4%	81.7%±6.9%
5	CCD vs. BPH	94.3%±6.7%	83.3%±6.0%	85.1%±3.1%	82.0%±10.3%
6	CCD vs. CAB	95.8%±1.8%	84.4%±4.4%	85.2%±7.1%	82.5%±2.2%
7	(CAB + CCD) vs. NO	97.9%±2.9%	93.2%±7.8%	94.1%±6.4%	92.4%±8.2%
8	(CAB + CCD) vs. (NO + BPH)	86.7%±6.0%	81.2%±4.1%	84.3%±8.7%	78.1%±7.4%
9	(BPH + CAB + CCD) vs. NO	93.6%±4.1%	87.4%±7.4%	86.1%±7.1%	87.8%±7.7%

Table 4: Examination of four-stage integrated approach using raw data set with cross-validation. Training and testing accuracies (mean ± standard deviation) are calculated from 100 runs; 90% of data used for training of each run

3-way classification							
Training				Testing			
	NO	BPH	(CAB+CCD)		NO	BPH	(CAB+CCD)
NO	135.84 (91.6%)	3.96 (2.9%)	12.38 (4.1%)	NO	13.35 (85%)	0.89 (5.1%)	2.15 (6.7%)
BPH	5.34 (3.6%)	125.99 (92.3%)	16.31 (5.4%)	BPH	1.22 (7.8%)	15.1 (86.3%)	2.91 (9.1%)
CANCER	7.12 (4.8%)	6.55 (4.8%)	273.31 (90.5%)	CANCER	1.13 (7.2%)	1.51 (8.6%)	26.94 (84.2%)
Average accuracy		91.2%±4.9%		Average accuracy		85.0%±6.3%	

Table 5: Confusion matrix of 3-way classification; Examination of four-stage integrated approach using raw data set with cross-validation. Training and testing accuracies (mean ± standard deviation) are calculated from 100 runs; 90% of data used for training of each run

4-way classification								
Training					Testing			
	NO	BPH	CAB	CCD		NO	BPH	CAB	CCD
NO	132.27 (90.1%)	1.10 (0.8%)	3.68 (2.4%)	4.34 (2.9%)	NO	14.67 (85.3%)	0.46 (2.7%)	0.54 (3.7%)	1.05 (6.4%)
BPH	2.79 (1.9%)	122.75 (89.6%)	7.82 (5.1%)	5.38 (3.6%)	BPH	0.72 (4.2%)	14.65 (86.2%)	1.23 (8.4%)	0.84 (5.1%)
CAB	6.31 (4.3%)	4.24 (3.1%)	134.38 (87.6%)	7.93 (5.3%)	CAB	0.97 (5.6%)	0.58 (3.4%)	12.06 (82.6%)	1.08 (6.6%)
CCD	5.43 (3.7%)	8.91 (6.5%)	7.52 (4.9%)	131.95 (88.2%)	CCD	0.84 (4.9%)	1.31 (7.7%)	0.77 (5.3%)	13.43 (81.9%)
Average accuracy		88.8%±8.2%			Average accuracy		84.1%±7.4%		

Table 6: Confusion matrix of 4-way classification; Examination of four-stage integrated approach using raw data set with cross-validation. Training and testing accuracies (mean ± standard deviation) are calculated from 100 runs; 90% of data used for training of each run

The results of binary classifications for two-stage integrated approach are shown in Table 3. The sensitivity and specificity are measured with the ratios of true positive (illness) and true negative (health). In all of the 9 categories of binary classifications, we put the "positive" status first and then the "negative" one. For example, the sixth test of CCD vs. CAB, we consider CCD is the "positive" status and CAB is the "negative" one. We can observe that the top 5 of testing accuracies are rows 1, 3, 2, 7, and 9, which of them all contain the group of normal controls. It is clear that the group of healthy people is easier to distinguish with the group of ill ones.

As we mentioned earlier, the two-stage integrated approach could not converge for 3- or 4-way experiments in ANN classification learning. We then propose a 4-stage integrated approach to refine it. The experimental tests of 9 categories of binary classifications are then repeated for this refined approach, and the results are shown in Table 4. It is very obvious that the 4-stage approach outperforms the two-stage method and achieves excellent performance. All of testing accuracies and sensitivities are above 81% and 84%, respectively. Moreover, the testing accuracies, sensitivities, and specificities are all above 92% on rows 1, 2, 3, and 7.

In the 3- and 4-way experiments, the results are shown in Tables 5 and 6. The left half part of each table is a confusion matrix for training, and the other half on the right is for testing. A confusion matrix contains information about actual and predicted classifications done by a classification system. The entries in the diagonal are classified correctly, and those entries not on the diagonal are classified incorrectly. The class labels listed on the top row are actual class levels and those listed on the first column are the predicated ones. For example, the top left entry in Table 5 is 135.84, which is the number of correct predictions that a sample is "NO". This is an average of 100 runs and therefore the number is not an integer. The value 91.6% in parentheses is the accuracy rate (135.84 / 148 = number of samples classified as "NO" / number of "NO" samples in training set). For another example, the middle left entry in Table 5 is 5.34, which is the number of samples classified as "BPH" that a sample is "NO". The value 3.6% in the parentheses is the error rate (5.34 / 148). The total average accuracy of all class levels is listed in the bottom row of the table. For example, the average training accuracy of 100 runs in Table 5 is 91.2%. From Tables 5 and 6, we observe that the average testing accuracies of 3- and 4-way classifications are 85% and 84%, respectively. This demonstrates the effectiveness of the integrated method.

5 DISCUSSION ANS CONCLUSION

We have presented two integrated approaches for prostate cancer detection. The first one is a two-stage method which integrates DWT and ANN only. These two stages can be considered as the data pre-processing and classification learning. The proteomic data is a high-dimensional mass spectrometry data, and DWT has the capability to extract information from complex data. This leads us to apply the wavelet representation in our first stage of the integrated approach. By using a wavelet transform and a selected threshold, the dimensionality of MS data reduces from 43,556 to 1,599 in our DWT process, which is a dramatic data reduction with a compression ratio of 3.7% (1,599 / 43,556). ANN has been successfully applied in pattern recognition and classification, and medical diagnosis can be considered as an example of pattern recognition. This motivates us to utilize ANN in the second stage for classifying different disease status. Moreover, ANN achieved the best performance compared to SVM, and CART in our previous study for MS data analysis [1]. We examined those three classification algorithms on the manually pre-processed data in which each sample contains 779 features which is a much reduced dimensions.

The two-stage approach obtains satisfactory results in the 2-way experiments. This shows the effectiveness of using DWT and ANN for data pre-processing and disease classification on the raw dataset. However, the ANN training did not converge in the 3- and 4-way experiments. In contrast, in our previous study of using the pre-processed data, the ANN achieved 91% of the average testing accuracy in the 4-way experiment. This gives us a clue that a further data reduction may be needed. Since the dimensionality of the pre-processed data is 779, which is about a half of the 1,599, where the features we have after the DWT process on the raw dataset.

A four-stage approach is then proposed to refine the two-stage method in data reduction. There exists several data reduction techniques, and principal component analysis (PCA) is possibly the most widely used one. We therefore chose PCA for data reduction followed by DWT. However, we use ANN to extract principal components instead of finding eigenvalues and eigenvectors from the covariance matrix. Several neural network architectures exist for extracting principal components [8], and we chose the *n-h-n* network (or called Autoassociator) for its simplicity. However, some difficulties arisen when we applied the Autoassociator algorithm. Due to the characteristic of *n-h-n* network, n is dimension of the input space as well as the output space. The dimensionality after the DWT process is 1,599, and therefore a 1,599 x h x 1,599 network is trained with Backpropagation. The training usually did not converge because of the large value of n (=1,599). Hence, the voting method is proposed to select the common wavelet coefficients followed by DWT, and the number of features n is then reduced to 987. Although the large value of n had been solved, there was another problem for the choice of h. We then used an empirical test to find 400 as an optimal value. The improved approach for the two-stage method has then become a 4-stage integrated approach through the processes of DWT, voting, Autoassociator, and ANN. The experiments of the four-stage integrated approach show the excellent performance in 2-, 3-, and 4-way classifications.

Although this prostate cancer dataset was investigated by several authors [1, 2, 3, 4, 5], there were only two of them worked on the raw dataset [4, 5]. Yuasi et al. studied two categories of 2-way experiments, and Qu et al. only worked on one binary classification. In addition, Qu et al. used a subset of the samples and without considering the group of BPH patients. In contrast to previous work, we have investigated 9 categories of 2-way experiments. Moreover, 3- and 4- way experiments have also studied.

Although the integrated approach was tested on the prostate cancer dataset only, it is reasonable to hypothesize that the method may be effective to classifying disease status of other cancers using proteomic data. This leads us to the direction of our future work.

REFERENCES

[1] C. C. Huang, C. W. Huang, T. J. Chen, J. C. Yue, G. Hwang, Y. C. Chang, B.L. Adam. "Comparison of Classification Algorithms for Prostate Cancer Detection Using Mass Spectrometry Data". In Processing of 2005 International Conference on Intelligent Technologies and Applied Statistics, Taipei, Taiwan, June 2005.

[2] BL. Adam, Y.Qu, J. Davis, M. Ward, M. Clements, L. Cazares, OJ. Semmes, P. Schellhammer, Y. Yasui, Z. Feng, and G. Wright Jr. "Serum Protein Fingerprinting Coupled with a Pattern-Matching Algorithm Distinguishes Prostate Cancer From Benign Prostate Hyperplasia and Healthy Men. Cancer Research, 62:3609-3614, 2002.

[3] R. Lilien, H. Farid, and B. Donald. "Probabilistic Disease Classification of Expression-Dependent Proteomic Data from Mass Spectrometry of Human Serum". Journal of Computational Biology, 2003.

[4] Y. Qu, BL. Adam, M. Thornquist, J. Potter, M. Thompson, Y. Yasui, J. Davis, P. Schelhammer, L. Cazares, M. Clements, G. Wright, Jr., and Z. Feng. "Data Reduction using a Discrete Wavelet Transform in Discriminant Analysis of Very High Dimensionality Data". Biometrics, 59, 143-151, 2003.

[5] Y. Yasui, M. Pepe, M. Thompson, BL. Adam, G. Wright, Jr., Y. Qu, J. Potter, M. Winget, M. Thornquist, and Z. Feng. "A Data-Analytic Strategy for Protein Biomarker Discovery: Profiling of High-Dimensional Proteomic Data for Cancer Detection". Biostatistics, 4, 449-463, 2003.

[6] T. M. Mitchell. Machine Learning. McGraw-Hill, 1997.

[7] M.A. Carrerira-Peripinan. "A Review of Dimension Reduction Techniques". Technical Report CS-96-09, Dept. of Computer Science, University of Sheffield, 1997.

[8] K. I. Diamantaras and S. Y. Kung, Principal Component Neural Networks. Theory and Applications, Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, New York, London, Sydeny, 1996.