Original Article

www.jmss.mui.ac.ir

Using Protein Interaction Database and Support Vector Machines to Improve Gene Signatures for Prediction of Breast Cancer Recurrence

Mohammad Reza Sehhati, Alireza Mehri Dehnavi, Hossein Rabbani, Shaghayegh Haghjoo Javanmard[1]

Departments of Biomedical Engineering, [1]Physiology, Isfahan University of Medical Sciences, Isfahan, Iran

Submission: 17-03-2013 Accepted: 07-04-2013

ABSTRACT

Numerous studies used microarray gene expression data to extract metastasis-driving gene signatures for the prediction of breast cancer relapse. However, the accuracy and generality of the previously introduced biomarkers are not acceptable for reliable usage in independent datasets. This inadequacy is attributed to ignoring gene interactions by simple feature selection methods, due to their computational burden. In this study, an integrated approach with low computational cost was proposed for identifying a more predictive gene signature, for prediction of breast cancer recurrence. First, a small set of genes was primarily selected as signature by an appropriate filter feature selection (FFS) method. Then, a binary sub-class of protein–protein interaction (PPI) network was used to expand the primary set by adding adjacent proteins of each gene signature from the PPI-network. Subsequently, the support vector machine-based recursive feature elimination (SVMRFE) method was applied to the expression level of all the genes in the expanded set. Finally, the genes with the highest score by SVMRFE were selected as the new biomarkers. Accuracy of the final selected biomarkers was evaluated to classify four datasets on breast cancer patients, including 800 cases, into two cohorts of poor and good prognosis. The results of the five-fold cross validation test, using the support vector machine as a classifier, showed more than 13% improvement in the average accuracy, after modifying the primary selected signatures. Moreover, the method used in this study showed a lower computational cost compared to the other PPI-based methods. The proposed method demonstrated more robust and accurate biomarkers using the PPI network, at a low computational cost. This approach could be used as a supplementary procedure in microarray studies after applying various gene selection methods.

Key words: *Breast cancer, feature selection method, protein–protein interaction, recurrence prediction, support vector machine*

INTRODUCTION

In recent years, a number of studies have been performed for biomarker discovery from the available datasets on the gene expression data of breast cancer tumors. However, applying different analytical procedures on different sample cohorts in independent studies led to an inconsistency between the introduced biomarkers. Moreover, previously introduced gene signatures could not assure the clinicians a good prognostic prediction on different datasets.[1,2] Researchers attribute the stability problem to different sources, such as, different technological platforms,[3] sensitivity of gene selection methods to the samples,[4,5] inadequate number of samples,[6] ignoring the gene interactions in traditional classification approaches,[7] and the limitations of working in high dimension.[8] In recent times, Haury *et al.* reported that filter feature selection (FFS) methods were more effective than other complex procedures, to overcome the high dimension problem in breast cancer datasets.[8] Wherever the platforms of datasets were the same, an appropriate feature selection method was used, primarily for dimension reduction. However, the problem of investigating the interactions between genes yet remains unsolved.[8]

On account of the high computational cost and limitation of resources, considering the interactions between all the genes in a gene selection step was not an applicable solution. Moreover, based on the previous studies and literature, only a small group of genes were interacted to make a clinical outcome. Meanwhile, interacting partners of human proteins have tight relation to proper biological activity.[9] Therefore, interactions between human proteins, which were obtained experimentally,[9] were used as supplementary data to overcome the stability problem. For this purpose the gene expression profiles integrated with the protein–protein interaction (PPI) network in different ways to improve the performance of prognostic prediction.[7,10-12] Chuang *et al.*[7] mapped all the gene expression profiles to the corresponding

Address for correspondence:
Dr. Alireza Mehri Dehnavi, Department of Biomedical Engineering, Medical Image & Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: mehri@med.mui.ac.ir

proteins in the PPI network, where they searched for subnetwork signatures. Subsequently, the mean of the expression levels of the genes in each resultant subnetwork was used for the prediction of metastasis in breast cancer patients. Later, another methodology with better accuracy was proposed by Taylor *et al.,*[10] who used the correlation of gene expression profiles among highly connected proteins in the PPI network. More recently, Zhang *et al.*[11] investigated disruptions between the functional blocks of proteins in the PPI network, called domains,[13] which they believed were the main cause of the cancer outcome. More recently, the study of Jahid and Ruan[12] introduced intermediate proteins between differentially expressed genes in the PPI network as being highly probable biomarkers.

Previous studies demonstrated that the PPI network could be a very useful biological database in the discrimination of cancer outcomes. However, considering the entire PPI network, with several ten thousands of nodes and relations, takes the problem of working in high dimension to a worse state. Our approach overcame this problem by using an appropriate feature selection technique for pruning the non-informative genes at the first step.

After integrating the primary gene signatures with the PPI network, we utilized the support vector machine-based recursive feature elimination (SVMRFE), which is one of the leading feature selection methods and has been originally proposed for cancer classification.[14] The SVMRFE, which is an embedded-class method, is robust against noisy data and has been shown to perform well in microarray data expression analysis,[15] especially when applied with a nonlinear Gaussian kernel.[16] On account of the computational cost of this method for applying on datasets with a large number of features, it is very promising to use this method after the primary pruning of features.

The remainder of this article is organized as follows: In Section 2, a brief introduction to feature selection approaches is presented, followed by details of our procedure. Utilized datasets and experimental results are reported in Section 3. In Section 4 we discuss the obtained results, and then conclude our study and outline some directions for a future study.

MATERIALS AND METHODS

Feature Selection Procedure

Choosing the appropriate machine learning method in different stages of a biomarker selection procedure has great impact on the performance of the final constructed gene signatures.[8] In this study, we have presented an approach for improving the performance of gene signatures that are selected from the microarray data by different feature selection methods. Depending on the interaction mechanism of the feature selection methods with the classifier, they can be classified into three main categories namely; filter method, wrapper technique, and embedded approach.[17] In the filter method, all features must be ranked using a scoring criterion, independent of a classifier. Low computational cost, better performance for biomarker discovery from the breast dataset,[8] and ease of theoretical design are the advantageous characteristics of this approach.[18] In the wrapper approach the accuracy of a specific classifier is used to score various possible subsets of features. Due to the necessity of the training process of the classifier for every subset candidate, this method has high computational complexity. In the embedded method, the search procedure was integrated with the classifier construction for selecting the optimal subset of features. From various points of view, such as, complexity, computational cost, and an overfitting problem, the filter method was the best among other approaches.[8,17] However, the main traditional weakness of this technique is that they do not take the interaction between features into account.

In this study two types of feature selection methods were used in two different stages of the whole procedure. In addition to the expressed advantages of the FFS methods, Haury *et al.,*[8] declared them as the most effective approaches to overcome the high dimensionality problem in breast cancer datasets. Therefore, at the first stage, an FFS method was used for pruning the vast majority of non-informative genes. Then the resultant signature was expanded by the described strategy in section 3. Finally we applied a successful embedded method to the expanded set that is expressed in subsections 2-3. Figure 1 illustrates a schematic overview of the overall proposed approach.

Extending Primary Gene Signatures

According to the scoring criterion of a given FFS method, the genes that have the highest predictive power will be selected.[18] Therefore, it is rational that the currently selected genes have a meaningful expression level for prediction of cancer recurrence. However, the predictive performance of the combination of genes is not satisfactory. Therefore, we proposed to perform a more comprehensive search within the selected genes and their interacting partners in the PPI network.

Proteins are the executive agents of the biological processes in a cell. They facilitate gene expression, cell growth, and other cancer-related processes. As the proper function of a majority of proteins depends on their interaction with other proteins, they should be investigated in the context of their mutual partners. The PPI-network that is derived from the human protein reference database (HPRD) is a well-established and widely used tool in bioinformatics that depicts the experimentally characterized interaction between proteins.[9,19] There are many biochemical, biophysical, and

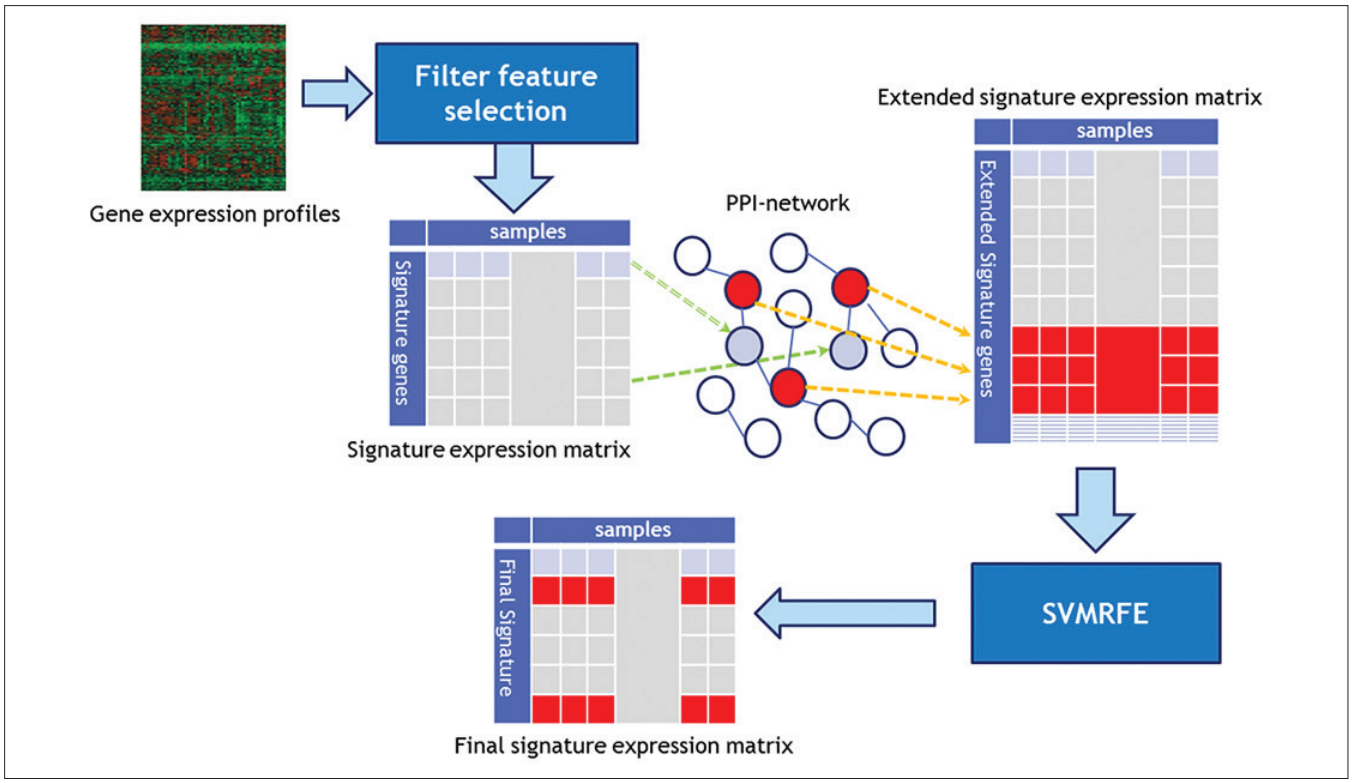


Figure 1: A schematic view of our approach

theoretical methods for investigating the interaction between proteins. The HPRD utilized here contains about 40,000 real binary interactions between human proteins, which have been obtained using *in vivo, in vitro*, and yeast two-hybrid experiments. In the binary sub-class of this network, we first find the nodes of the network that correspond to the genes that are selected by the feature selection method. Then we expand the primary gene set by adding nodes from the PPI-network that has a direct interaction with them (neighbor proteins). Using this procedure we have obtained a new gene set that could be about ten times bigger in size in comparison to the primary selected signature. Using this strategy we have only investigated a small part (about 1%) of the whole PPI network looking for biomarkers.

Extracting Final Biomarkers

Although the embedded and wrapper methods follow a more comprehensive procedure for feature selection, they generally do not outperform simple filter methods. A simple justification for this surprising result is 'the statistical issue of working in a high dimension with a few samples'.[8] In the previous sections we have described our strategy for constructing a smaller and more informative set of genes. It is promising to apply a complex feature selection method on this pruned set rather than on all the representatives in the microarray.

Therefore, we applied the SVMRFE to the expanded set, for choosing the best candidate as a gene signature. The SVMRFE followed a backward feature elimination strategy to consecutively eliminate the features from an initial set that contains all genes. The gene that obtained the smallest weight in a SVM classifier, which trained with the current subset for target prediction, was removed from the subset at each step. Using SVMRFE, all the features in the set that were expanded by the PPI-network would be ranked according to their corresponding weight, which was assigned by an SVM classifier, with a nonlinear Gaussian kernel to the features. Finally, a signature set was constructed by the genes with the highest rank, with a minimal appropriate size.[20]

RESULTS

Utilized Datasets

We utilized human breast cancer tumor microarray datasets of four independent studies including more than 800 samples that were publicly available from the Gene Expression Omnibus (GEO) database.[21] These datasets will be referenced later in this article by their GEO series code (GSExxx), as described in Table 1. All the utilized datasets were prepared in the same platform (HG-U133A or GPL96).

Table Samples were classified into two groups, high and low risk, according to the time to metastasis using a threshold of five years.

On the HG-U133A (GPL96) platform there are 22,283 probe sets that map to 12,172 genes. We used a one-to-one mapping method to select a single representative probe set for each gene.[26] Thus, 12,172 probes with the best score have been selected from this platform. Then all samples that are censored before the five-year follow ups and those that relapsed after five years have been removed from the datasets. After these pruning steps, all expression data were log2 transformed and then normalized, initially upon sample vectors and then across the feature vectors independently, in each dataset. The normalization procedure is made up of subtracting the mean overall expression values and dividing them by the corresponding standard deviation. Subsequently, the results pass through a mathematical function of the type f(x) = arctan(x), to uniformly limit the range of all values and reduce the importance of the outliers.[14] Finally, a metadata has been constructed by combining all the utilized datasets.

Experimental Results

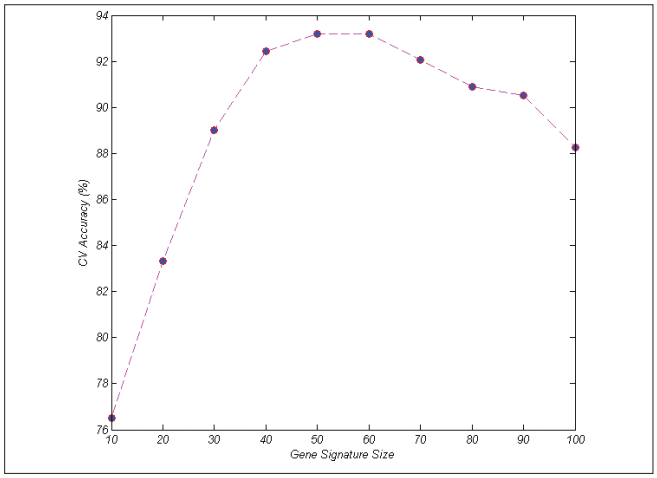
For evaluating the performance of the proposed method, we selected five FFS methods, with successful usage in the literature,[27] and assessed the accuracy of the signatures obtained by them, before and after applying our approach. The accuracy refers to the prediction performance of signatures that can be reached by a classifier trained on the genes. We chose the support vector machine (SVM) as a predictor model, which was a leading classifier method.[28] In this regard we utilized the LIBSVM package,[29] for implementing both the classifier model and SVMRFE. We followed the authors' practical guide for tuning the parameters of SVM. Penalty parameter C and Gamma of Gaussian kernel are two parameters that were tuned with regard to reaching the best five-fold cross-validation (CV) accuracy in 500 random subsamples of metadata. Figure 2 showed the mean accuracies obtained for different values of parameters C and Gamma over 500 subsamples at size 200. According to Figure 2 a range of values for these parameters (1 to 130 for C and 0.01 to 0.6 for Gamma) reached the same acceptable accuracy.

We utilized five successful filter methods for the evaluation, namely; joint mutual information (JMI),[30] mutual information maximization (MIM),[31] minimum redundancy-maximum relevancy (mRMR),[32] *t*-test, and Wilcoxon.[33] Following the five-fold CV test procedure, we randomly divided the dataset into five subsets of equal size. Consecutively one subset was tested using the SVM classifier, which trained on the remaining four subsets. The CV process was then repeated *five* times in such a way that all the samples were used for both the training and test, and each sample was used for testing exactly once. Thus, the CV accuracy was the percentage of all data that were correctly classified when they were chosen as the test subset. In this experiment a five-fold CV process was performed to evaluate the accuracy of gene signatures with different sizes over the Wang dataset [Supplemental Figure 1]. Based on the obtained results, the signature with size 50 reached the maximum accuracy. Thereafter, the CV accuracy of 50-gene signatures obtained by each feature selection method, were evaluated over different datasets. Figure 3 shows the CV accuracy of the primary extracted signatures and finally obtained signatures by our approach over the Wang dataset. We demonstrated the evaluation results over the Wang dataset for comparison with other similar PPI-based studies that also used this dataset. The

Table 1: Summary of breast cancer microarray datasets

Dataset	#Samples	#High-risk	#Low-risk	Source
GSE2034	286	95	169	Wang *et al*., 2005[22]
GSE7390	198	54	100	Desmedt *et al*., 2007[23]
GSE6532	244	65	123	Loi *et al*., 2007[24]
GSE3494	236	37	158	Miller *et al*., 2005[25]
Metadata	800	251	549	

#Number of; GSE – GEO series



Supplemental Figure 1: CV accuracy versus gene signature size. For size 50 we reached the max accuracy over the Wang dataset

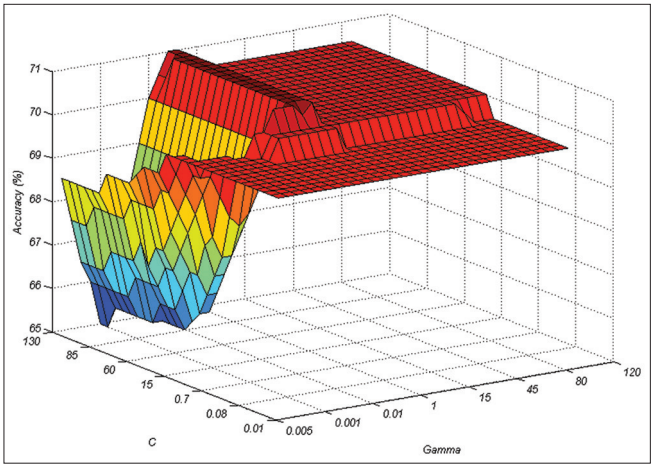


Figure 2: Tuning of SVM parameters to obtain the best accuracy

obtained results over the other three datasets which are not shown here are in agreement with Figure 3.

We also used the area under the receiver operator curve (AUC) as the classification performance measure [Supplemental Figure 2]. The comparative results of evaluation using the AUC are in concordance with the CV results, which are shown in Figure 3.

As mentioned before, the main challenge of gene selection was instability of the introduced biomarkers, which are extracted from different datasets. We showed that our approach obtained more robust signatures when utilizing a previously introduced gene signature as the primary signature. Evaluations of the CV accuracy of the obtained signatures from each dataset for the prediction on the other datasets proved the robustness issue. In this regard the Wilcoxon method was used for selecting primary signatures, based on the better prediction accuracy of this method among other approaches. Table 2 illustrates the five-fold CV accuracies for 50-gene signatures, which are extracted primarily from source datasets (row labels) and tested over the destination datasets (column labels).

For more obvious demonstration of the achieved improvment in robustness we illustrated this in Figure 4. In this figure, the horizontal axis indicates the source datasets that were used for extracting the primary signatures by the Wilcoxon method. The height of each bar, which two of them presented for every dataset, indicates the mean CV accuracy obtained over the other three datasets using the primary signature and the final modified one.

Furthermore, for deriving a hypothesis about breast cancer at a genomic level, we reported a summary of the best 100 selected biomarkers from the Wang dataset, with a brief biological description about them [Supplemental Table 1].

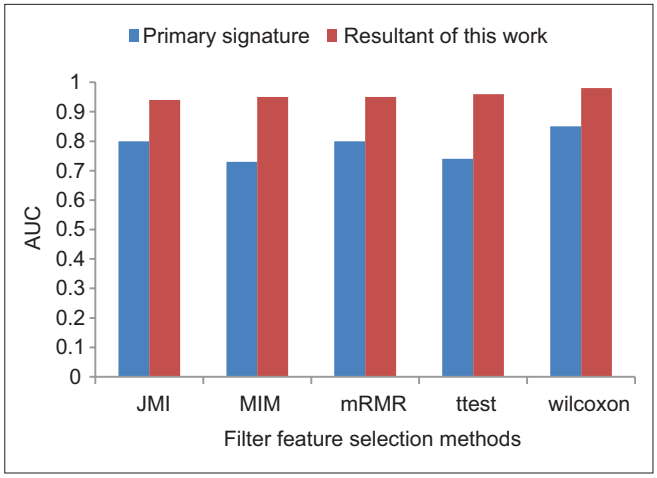
DISCUSSION

In this study, we presented an integrated approach that used the binary subclass of the PPI network for identification of more predictive genes from a microarray data. At the first step an arbitrary FFS method was used for selecting a small gene signature from the breast cancer datasets. Then a new set was constructed by adding the genes from the PPI network, which had direct connection with the primary genes. Afterward, an embedded method with a backward elimination strategy, called SVMRFE, was used to select a new signature set for prediction of breast cancer recurrence. On account of the computational cost and high dimensional

Table 2: Robustness of gene signature performances. Each cell indicates the five-fold CV accuracy of a signature that was extracted by Wilcoxon, primarily from the source dataset and then expanded and tested over the destination dataset

Destination source	GSE2034	GSE7390	GSE6532	GSE3494	Metadata
GSE2034	93.18	87.66	88.82	92.3	78.5
GSE7390	81.43	94.15	80.85	93.84	77.5
GSE6532	87.87	92.2	90.95	93.84	78.4
GSE3494	85.6	88.31	81.38	95.38	79.15

CV – Cross-validation; GSE – GEO series



Supplemental Figure 2: Improvement in accuracy (using AUC as classification measure) after applying our approach to different FFS methods.

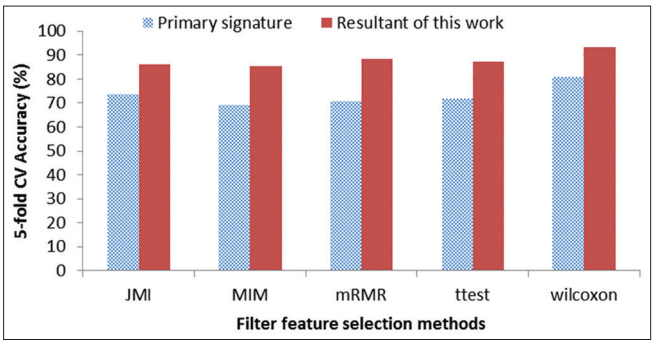


Figure 3: Classification performance (five-fold CV) in the Wang dataset before and after applying our approach to five FFS methods

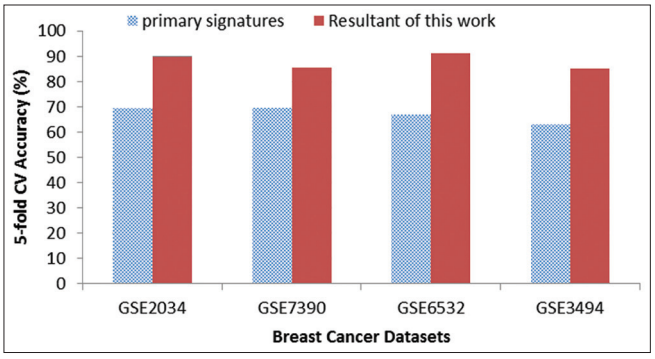


Figure 4: Comparison of robustness among different datasets for primary signatures and the final result of our approach. The horizontal axis indicates the source datasets that are used for extracting primary signatures by the Wilcoxon method. Bar height indicates the mean CV accuracy obtained over the other three datasets using the primary signatures

problem it is very promising to use such a complex and powerful method after the primary pruning of features.

We compared our approach with other similar recent studies that used the PPI network for improving the accuracy of metastasis prediction in breast cancer patients. Chuang *et al.*, presented the first method, which joined the gene expression profiles with the PPI network for classification of breast cancer metastasis. This method showed 10% improvement in the prediction accuracy when tested on the dataset of Wang *et al.*, (GSE2034) that reached an accuracy of 62% using a pattern-based approach. Later Zhang *et al.,* proposed a different strategy and obtained a better accuracy of 81.7% on the same dataset. More recently Jahid and Ruan[12] proposed another new procedure for using the PPI network and obtained a precision of 77% over the Wang dataset. The analytical methods that were used in all of these studies were computationally expensive, due to simultaneously investigating tens of thousands of nodes and connections in the PPI network. Therefore, the first major advantage of our approach in comparison with theirs is using the PPI network with lower computations, because we only investigated a small part of this network when looking for biomarkers. In this study we applied the FFS methods to the microarray dataset to prune the non-informative genes and limit the domain of investigations in the PPI network. A quick look at Figure 3 proves that we obtained more than 85% accuracy for prediction in the Wang dataset, using different FFS methods. As shown in Figure 3, our approach improved the accuracy of prediction at least by 13% for all methods and reached 93% accuracy using the Wilcoxon method.

We also investigated the prediction performance of each dataset signature over the other independent datasets. Table 2 and Figure 4 showed that our approach reached a high accuracy among various datasets and we can conclude that the final signatures were robust. According to Table 2, the lower accuracy (still acceptable) obtained for the metadata can be attributed to the divergence of data samples, which disabled the SVMRFE, for obtaining a more predictive gene signature.

Meanwhile significant improvement was achieved independent of the chosen FFS method. Choosing an appropriate feature selection technique has a great impact on the stability and prediction power of the final signature set and should be investigated in an independent study. This statement also includes the SVMRFE that was chosen based on literature.

It should be noted that discussion about the biological aspects of the final extracted gene signatures and the overlap between them are very important issues that should be considered consecutively. We also believe that utilizing other biological databases such as gene ontology and known biological pathways can improve the accuracy of the obtained gene indicators even further. These open discussions are the most important related issues to this study that should be considered in future studies.

REFERENCES

1. Li J, Lenferink AE, Deng Y, Collins C, Cui Q, Purisima EO, *et al.* Identification of high-quality cancer prognostic markers and metastasis network modules. Nat Com 2010;13:1-34.
2. Zhao X, Rødland EA, Sørlie T, Naume B, Langerød A, Frigessi A, *et al.* Combining gene signatures improves prediction of breast cancer survival. Plos One 2011;6:e17845.
3. Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, *et al.* Adjustment of systematic microarray data biases. Bioinformatics 2004;20: 105-14.
4. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: Is there a unique set? Bioinformatics 2005;21:171-8.
5. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: A multiple random validation strategy. Lancet 2005;365:488-92.
6. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proc Natl Acad Sci USA 2006;103:5923-8.
7. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol 2007;3:140-50.
8. Haury AC, Gestraud P, Vert JP. The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. Plos One 2011;6:e28210.
9. Golemis E. Protein-protein interactions: A molecular cloning manual. Cold Spring Harbor. NY: Cold Spring Harbor Laboratory Press; 2002.
10. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, *et al*. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 2009;27:199-204.
11. Zhang KX, Ouellette BF. CAERUS: Predicting Cancer outcomes Using Relationship between Protein Structural Information, Protein Networks, Gene Expression Data, and Mutation Data. PLoS Comput Biol 2011;7:e1001114.
12. Jahid MJ, Ruan J. A Steiner tree-based method for biomarker discovery and classification in breast cancer metastasis. BMC Genomics 2012;13:S8.
13. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. Science 2003;300:445-52.
14. Guyon I, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389-422.
15. Mundra PA, Rajapakse JC. SVM-RFE with MRMR filter for gene selection. IEEE Trans Nanobioscience 2010;9:31-7.
16. Liu Q, Chen C, Zhang Y, Hu Z. Feature selection for support vector machines with RBF kernel. Artif Intell Rev 2011;36:99-115.
17. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157-82.
18. Duch W. Feature Extraction: Foundations and Applications, Studies in Fuzziness and Soft Computing. Springer: Berlin Heidelberg New York; 2006. p. 89-117.
19. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, *et al*. Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 2003;13:2363-71.
20. Mehridehnavi A, Ziaei L. Minimal gene selection for classification and diagnosis prediction based on gene expression profile. Adv Biomed Res 2013;2:26.
21. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, *et al.* NCBI GEO: Archive for high-throughput functional genomic data. Nucleic Acids Res 2009;37:885-90.
22. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F,

et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancers. Lancet 2005;365:671-9.
23. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, *et al.* Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multi-center independent validation series. Clin Cancer Res 2007;13:3207-14.
24. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, *et al.* Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007;25:1239-46.
25. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci USA 2005;102:13550-5.
26. Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: Selecting the optimal microarray probe set to represent a gene. BMC Bioinformatics 2011;12:474.
27. Brown G, Pocock A, Zhao MJ. Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. J. Mach Learn Res 2012;13:27-66.
28. Platt, J. Fast training of support vector machines using sequential minimal optimization. Cambridge, MA, USA: MIT Press; 1999. p. 185-208.
29. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans Int Syst Technol 2011;2:1-27.
30. Meyer PE, Schretter C, Bontempi G. Information-theoretic feature selection in microarray data using variable complementarity. IEEE J Sel Top Signal Process 2008;2:261-74.
31. Lewis DD. Feature selection and feature extraction for text categorization. In Proc of the workshop on Speech and Natural Language. NJ, USA: Association for Computational Linguistics Morristown; 1992. p. 212-7.
32. Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE T Pattern Anal 2005;27:1226-38.
33. Hwang T, Sun CH, Yun T, Yi GS. FiGS: A filter-based gene selection workbench for microarray data. BMC Bioinformatics 2010;11:50-5.

How to cite this article: Sehhati MR, Dehnavi AM, Rabbani H, Javanmard SH. Using Protein Interaction Database and Support Vector Machines to Improve Gene Signatures for Prediction of Breast Cancer Recurrence. J Med Sign Sens 2012;3:87-93.
Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHIES

Mohammad Reza Sehhati was born in Isfahan, Iran, in 1981. He received the B.S and M.S degree in Biomedical Engineering from Shahed University and University of Tehran, Tehran, Iran, respectively. He is now Ph.D Candidate of Biomedical Engineering in Isfahan University of Medical Sciences, Isfahan, Iran. His research interests include Bioinformatics, Machine Learning, Data Mining, Image Processing, and Hospital Information Systems

E-mail: sehhati@resident.mui.ac.ir

Alireza Mehri Dehnavi was born in Isfahan province at 1961. He had educated in Electronic Engineering at Isfahan University of Technology at 1988. He had finished Master of Engineering in Measurement and Instrumentation at Indian Institute of Technology Roorkee (IIT Roorkee) in India at 1992. He has finished his PhD in Medical Engineering at Liverpool University in UK at 1996. He is an Associate Professor of Medical Engineering at Medical Physics and Engineering Department in Medical School of Isfahan University of Medical Sciences. He is currently visiting at School of Optometry and Visual Science at University of Waterloo in Canada. His research interests are medical optics, devices and signal processing

E-mail: mehri@med.mui.ac.ir

Hossein Rabbani is an Associate Professor at Isfahan University of Medical Sciences, in Biomedical Engineering Department also Medical Image & Signal Processing Research Center. Involved research topics include medical image/volume processing, noise reduction and estimation problem, image enhancement, blind deconvolution, video restoration, probability models of sparse domain's coefficients especially complex wavelet coefficients. He is a member of IEEE, Signal Processing Society, Engineering in Medicine and Biology Society, and Circuits and Systems Society

E-mail: h_rabbani@med.mui.ac.ir

Shaghayegh Haghjoo Javanmard is an Associate Professor at Isfahan University of Medical Sciences, in Physiology Department also Physiology Research Center. Involved research topics include Clinical Cancer Research, Bioinformatics, and Applied Physiology.

E-mail: shaghayeghhaghjoo@yahoo.com