

# Probabilistic Neural Networks for Multi-class Tissue Discrimination with Gene Expression Data

Rui Xu and Donald C. Wunsch II  
Applied Computational Intelligence Laboratory  
Dept. of Electrical and Computer Engineering  
University of Missouri - Rolla  
Rolla, MO 65409-0249 USA  
rxu@umr.edu, dwunsch@ece.umr.edu

**Abstract** – With the emergence and rapid advancement of DNA microarray technologies, construction of gene expression profiles for different cancer types has already become a promising means for cancer diagnosis and treatment. Most previous research has focused on binary classification. Here, we use a probabilistic neural network (PNN) for multi-classification of cancer data. The experimental results demonstrate the effectiveness of the PNN in addressing gene expression data.

## I INTRODUCTION

With the emergence and rapid advancement of DNA microarray technologies [1-2], cancer classification through identifying the corresponding gene expression profiles has already attracted numerous efforts from a wide variety of research communities [3-16]. Cancer classification is important to the subsequent diagnosis and treatment. Without the correct identification of cancer types, it is rarely possible to provide useful therapies and achieve expecting effects. Traditional classification methods are largely dependent on morphological appearance of tumors, and their applications are limited by the existing uncertainties [3]. Tumors with similar appearance may have quite different origins and therefore respond differently for the same treatment therapy. For example, in diffuse large B-cell lymphoma (DLBCL), almost half of clinical cases fail to the treatment due to the existence of unknown subtypes that cannot be discriminated by their morphologic parameters [4]. DNA microarray technologies offer cancer researchers a new way to investigate the pathologies of cancer from the molecular angle, and further, to make more accurate predictions in prognosis and treatment.

There exist different types of microarray technologies based on the nature of the attached DNA (cDNA with length varying from several hundred to thousand bases or oligonucleotides containing 20-30 bases). For cDNA technologies, a microarray consists of a solid substrate to which a large amount of cDNA clones are attached according to some certain order [1]. Fluorescently labeled cDNA, obtained from RNA samples of interest (e.g. tumor samples) through the process of reverse transcription, is hybridized with the array. A reference sample (e.g. normal samples) with a different fluorescent label is also required for the purpose of comparison. Image analysis techniques are then used to

measure the fluorescence of each dye after the genes are washed off. The resulting ratio reflects relative levels of gene expression. For high-density oligonucleotide microarray, oligonucleotides are fixed on a chip through techniques like photolithography and solid-phase DNA synthesis [2]. With a wealth of gene expression data at hand, researchers have more opportunities, while inevitably facing new challenges. Gene expression data sets usually have features like high dimensionality, high redundancy and inherent noise, which ask for the computational analysis methods to have corresponding mechanism to deal with them. Research on gene expression data is summarized in three levels, according to the task complexity [17]. The bottom level investigates the activities of single genes under different conditions or tissues. The second level focuses on the relations and interactions among genes and conditions. The top level attempts to infer the whole genetic network that finally determines all the patterns we observed. The tumor classification researches based on gene expression data can be classified into the intermediate level, which explores the relations between types of tumors and gene markers.

This kind of research has already been reported in the literature with promising results [3-16]. Golub et al. deemed cancer classification as two challenges: class discovery and class prediction and used several strategies, including weighted voting, neighborhood analysis and self-organizing feature maps (SOFMs) to discriminate two types of human acute leukemias (ALL vs. AML) [3]. According to their results, two subsets of acute lymphoblastic leukemia (ALL), with different origin of lineage, are also well separated. Alizadeh et al. distinguished two molecularly distinct subtypes of diffuse large B-cell lymphoma by their gene expression profiles [4]. Alon et al. performed a two-way clustering for both colon tissues and genes and revealed the potential relations between them [5]. Other explorations include ovarian cancer [6], breast cancer [7], cutaneous melanoma [8], and so on. For most of the researches aforementioned, hierarchical clustering (HC) is employed. Although HC has the advantage such as informative visualization of the clustering results and the versatility, it lacks robustness and does not have favorable scalability properties. We have used a new family of neural network architecture – Ellipsoid ART and ARTMAP (EA/EAM) to

analyze several publicly accessible data sets [11]. EA/EAM has the properties of fast, stable and finite learning and can create hyper-ellipsoidal clusters with complex nonlinear boundaries. Other examples include graph theory-based methods [10], singular value decomposition with Bayesian models [9], partial least squares combined with logic discrimination and quadratic discriminant analysis [14] and support vector machines (SVMs) [16].

In practice, it is common to discriminate more than two types of cancers. Ramaswamy et al. divided the multi-class problem as a series of binary classification sub-problems through either one-versus-all or all-pairs approach [18-19]. SVMs, weighted voting and  $k$ -nearest-neighbors algorithm were then used to perform binary classification and the final label was decided according to some confidence values. Khan trained perceptrons to categorize small round blue-cell tumors (SRBCTs) with 4 subclasses [12]. A nearest shrunken centroid method was proposed by Tibshirani et al. and was tested on the SRBCT data set with 100% accuracy [13]. Furthermore, Scherf et al. constructed a gene expression database to study the relationship between genes and drugs for 60 human cancer cell lines originating from 10 different tumors, which provides an important criterion for therapy selection and drug discovery [15].

In our study, we use the probabilistic neural network (PNN) [21] to address the problem of multiple tumor classification without the need to divide it into binary sub-problems. As a powerful tool developed to approximate the Bayesian decision rule, PNN has already shown appealing performance in a large number of applications [22-24]. Here, we further demonstrate the potential and effectiveness of PNN in addressing the challenges of gene expression data analysis with promising results based on several publicly accessible data sets on cancer researches.

The paper is organized as follows. Section II presents a brief introduction to PNN. Section III describes the data sets and experimental methods. The experimental results are presented and discussed in section IV and section V concludes the paper.

## II. PROBABILISTIC NEURAL NETWORKS

Probabilistic neural networks were first introduced by Specht [21] as an implementation of nonparametric Pazen window estimation with feed-forward neural network architecture. A typical PNN architecture is illustrated in Fig. 1, which consists of three layers, known as input layer, pattern layer and category layer [20, 24]. The input layer works as a distribution mechanism and receives input components from the data set. Therefore, the number of nodes in this layer is equal to the dimension of the input vector. All of these nodes are fully connected with the nodes in the pattern layer, which is considered as the key of PNN. The PNN requires  $n$  pattern nodes if the total number of

training patterns is  $n$ , so that each pattern node can be regarded as corresponding to a training pattern. Different from link between input and pattern layer, the nodes of pattern and category are sparsely connected. Each pattern node is only connected to the category node that correctly indicates the its associated class.

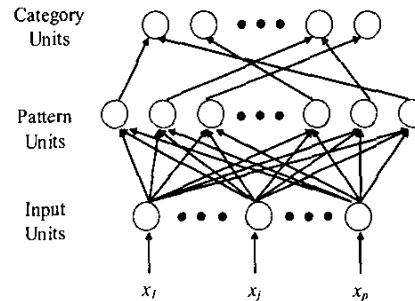


Fig. 1. PNN architecture. Each pattern node represents a pattern in the training set. The Bayesian posterior probability for each category is obtained as the output of the corresponding category node.

The PNN calculates the Bayesian posterior probability for each category. During the training phase, the weights connecting the input and pattern layer are simply set as the copy of input vectors, i.e.  $w_i = x_i$ , for  $i = 1, \dots, n$ . As mentioned before, each pattern node merely has connection with the category node representing the class of the corresponding pattern. The process is one of the fastest in known training strategies. However, the cost for the time efficiency is the storage complexity, one pattern unit is required for each pattern in the entire training set, which causes problems when dealing with large volume of data sets. One of potential solutions is to group training patterns into a series of clusters and only use the centroid of each cluster to supervise the network learning [25].

During the test or classification phase, each pattern node performs a dot product operation with a new pattern vector  $x$  and a weight vector  $w_i$ , expressed as  $P_i = x \cdot w_i$ . The final output of pattern layer is obtained via a nonlinear transformation. Usually a Gaussian activation function,  $\exp\left(-\frac{P_i}{\sigma^2}\right)$ , is used, though other alternatives are also available [21]. Here,  $\sigma$  is the smoothing parameter of the Gaussian kernel and is also the only one parameter that is dependent on the users to decide. It has great effect to the formation of resulting decision boundary and the strategy on how to select appropriate  $\sigma$  was discussed in details by Specht in [21]. Note that if both the training patterns (equivalent to weight vectors) and the new patterns are normalized to unit length, the output of pattern layer can be represented as

$$\begin{aligned}
& \exp\left(\frac{P_i - 1}{\sigma^2}\right) \\
& = \exp\left(-\frac{(\mathbf{x}^T \mathbf{x} + \mathbf{w}_i^T \mathbf{w}_i - 2\mathbf{x}^T \mathbf{w}_i) / 2\sigma^2}{\sigma^2}\right), \\
& = \exp\left(-\frac{(\mathbf{x} - \mathbf{w}_i)^T (\mathbf{x} - \mathbf{w}_i) / 2\sigma^2}{\sigma^2}\right)
\end{aligned}$$

which is identical to the Parzen window function [21]. In this sense, each pattern node provides the corresponding category node with the class conditional probability given the training pattern. These values are then summed up in the category layer for each category as the estimated probability for the new pattern. The label of the pattern can be predicted by just choosing the maximum probability.

Due to its design, PNN displays many desirable properties and characteristics in the context of pattern classification. Among all of those features, the fast training is most appealing. The training rule is very simple and can be used in online learning. Only one pass of the training data is required under the training scheme. Though PNN sacrifices the space efficiency for it, the cost can be decreased to certain extent with the introduction of clustering techniques [25]. Also, the structure of PNN makes it very easy to achieve parallel implementation. Another important feature of PNN is the fewer user-dependent parameters compared with other classifiers. There is only one Gaussian kernel width parameter  $\sigma$  relying on the selection of the users. According to [21] and our experiments, it is not difficult to find appropriate  $\sigma$  and the experimental results are not sensitive to the small changes of  $\sigma$ . Moreover, PNN puts the nonparametric statistical approach into a neural network framework and provides the output of the neural network with a new interpretation, i.e. through the form of conditional probability density function. The resulting probability for each category makes it possible to investigate the confidence of assigning a new pattern to a class. Finally, PNN has the ability to approximate Bayesian optimal decision surfaces that can be arbitrarily complex.

### III. DATA SETS AND EXPERIMENTS

We use two data sets to test PNN performance in multiple cancer classification.

**SRBCT data set.** This data set is on the diagnostic research of small round blue-cell tumors (SRBCTs) of childhood and consists of 83 samples from four categories, known as Burkitt lymphomas (BL), the Ewing family of tumors (EWS), neuroblastoma (NB) and rhabdomyosarcoma (RMS) [12]. 5 non-SRBCT samples are also included in the original data set for testing the ability of diagnosis rejection, but we do not use them in our study. Gene expression levels of 6567 genes were measured using cDNA microarray for each sample, 2308 of which passed the filter that requires the red intensity of a gene to be greater than 20 and were kept for further analyses. The relative red intensity (RRI) of a gene is defined as the ratio between the mean intensity of that

particular spot and the mean intensity of all filtered genes and the ultimate expression levels measure is the natural logarithm of RRI. The data are expressed as a matrix  $E = \{e_{i,j}\}_{83 \times 2308}$ , where  $e_{i,j}$  represents the expression level of gene  $j$  in tissue sample  $i$ .

**GCM14 data set.** This data set is available at <http://www-genome.wi.mit.edu/cgi-bin/cancer/datasets.cgi>. There are 14 different tumor types, consisting of breast, prostate, lung, colorectal, lymphoma, bladder, melanoma, uterus, leukemia, renal, pancreas, ovary, mesothelioma, and CNS cancer, with 218 samples. These samples are divided into three groups in the original research, 144 for training, 54 for testing, and the rest 20 as poorly differentiated (PD) tumors. Also, 90 normal tissue samples are included for the study of discriminating tumor and normal tissues. In our experiments, we only work on the 198 tumor samples in the original training and test sets. Gene expressions for 16,063 genes were measured using oligonucleotide microarrays. The final matrix is in the form of  $E = \{e_{i,j}\}_{218 \times 16063}$ .

**Gene selection.** From the above description of the two data sets used in the paper and other public data sets [3-8], it can be seen that one of the common feature is the overwhelming number of measures of gene expression levels compared with the number of samples. Not all of these genes are relevant to the discrimination of tumors and sometimes, only a small part of them is enough for correct classification [3, 11]. The existence of more genes that do not contribute to the distinction in the data sets not only increases the computational complexity, but impairs the effects of those relevant ones to some extent. Furthermore, cancer researches also require identifying the relation of tumors and their causes in the molecular level, which is imperative in determining appropriate therapy. Therefore, feature selection or extraction, also known as informative gene selection, is critically important in this context.

Principal component analysis (PCA) is a widely used tool for dimension reduction, which attempts to seek the projection that best interpret the variation of the data [20]. PCA has already been used in some applications on gene expression data [12]. But according to the experimental results in [26], PCA cannot always find the correct structure with just the first few principal components and therefore is not recommended under general cases. Several other methods based on ranking genes according to their expression differentiation under two different classes (represented as +1 and -1 here) have been proposed, examples including:

(1) Discrimination score [3]:

$$D(i) = \frac{\mu_+(i) - \mu_-(i)}{\sigma_+(i) + \sigma_-(i)},$$

where  $\mu_+(i)$  and  $\mu_-(i)$  are the mean values of gene  $i$  for the samples in class +1 and class -1, and  $\sigma_+(i)$  and

$\sigma_{-}(i)$  are the standard deviations of gene  $i$  for the samples in class +1 and -1.

(2)  $t$ -statistics score [14]:

$$T(i) = \frac{\mu_{+}(i) - \mu_{-}(i)}{\sqrt{\sigma_{+}^2(i)/n_{+} + \sigma_{-}^2(i)/n_{-}}},$$

where  $n_{+}$  and  $n_{-}$  are the sizes for samples in the two classes.

(3) TNoM score [10]:

$$TNoM(i, l) = \min_{d, t} Err(d, t | i, l),$$

where  $d$  is the class label parameter,  $t$  represents the threshold of the gene  $i$ ,  $l_k$  is the label of the  $k$ th sample, and

$$Err(d, t | i, l) = \sum_k 1\{l_k \neq \text{sign}(d(e_{ki} - t))\}$$

is the number of errors of a decision stump rule.

According to some experimental studies, most of these methods intend to choose the similar subsets of genes and do not greatly affect the performance of the classifiers. Thus, we employed the first criterion to select informative genes. The only change is that we just use the absolute value of the score. So it reflects the expression level difference between the two classes for each gene. Gene expresses itself most differently in two classes will have the highest score. Since our final goal is to classify multiple types of cancer, we utilize a one-versus-all strategy to seek gene predictors. In other words, for a  $C$ -class prediction problem, we compare a particular class with the other  $C-1$  classes that are considered as a whole. We can just select genes according to their contribution to distinguish each class, or the total score for each gene is summed over all  $C$  comparison and the top genes are selected with the highest scores. Of course, pairwise strategy can also be used, which performs  $\binom{C}{2}$  comparisons between each pair of classes.

**Experiment designs.** Since the data sets consist of only a small number of samples, it is better to use the jackknife approach, which is also called leave one out cross validation (LOOCV), to examine the performance of the classifier. For a data set with  $n$  samples, the classifier is trained  $n$  times. Each time, a different single sample is left out as the test point and the other  $n-1$  samples are used to train the classifier. Performance evaluation of the classifier is estimated by considering the average accuracy of the  $n$  cross-validation experiments.

We summarize our experimental procedure with the following steps. First, normalize the input patterns to unit length. Then, two strategies are used. For the first strategy, we rank and select a set of informative genes with the aforementioned criterion across all samples, and perform the LOOCV operation with one sample left out for test and the others for training in the learning and prediction phase. This method does not consider the effect of bias for gene selection

and is regarded as too optimistic [9]. The second strategy aims to overcome the bias by selecting informative genes at each step of LOOCV operation. Therefore, genes selected in the subset may be different for each stage. But generally, these gene subsets are highly overlapping, with only a small portion of difference observed. The experimental results are illustrated and depicted in the next section.

#### IV. RESULTS

Fig. 2 describes the classification accuracy for the SRBCT data set with the selection of different numbers of gene predictors. We illustrate the results for both the bias included and free strategy here. From the figure, we can see that there is just some minor decrease in the performance when bias free method was used. This shows that the effect of selection bias may not be very critical in the case of LOOCV procedure because of the high overlapping of gene subsets. We also can see the importance of informative genes selection in tumor classification. The PNN classifier can achieve 100% accuracy when only top 50 genes are chosen. If the subset is increased to include all genes in the data set or to another extreme, reduced to just comprise several genes, the accuracy will decrease in both cases, though not much. These suggest that too many or too few genes both deteriorate the performance of the classifier. Many genes are not related to the classification and including them in the data set will bring noise into the classification system. On the other hand, important information will be wrongly discarded with inadequate genes chosen. These results are consistent with those reported in [12] and [13].

Fig. 3 shows the effects of the smoothing parameter  $\sigma$  to the classification accuracy, when top 100 gene markers are selected with selection bias free strategy, or all genes are used. Obviously, it is not difficult to find a kernel width  $\sigma$  that can lead to a satisfying result. With the increase of the dimensionality, the effective  $\sigma$  tends to become larger in order to provide better interpolation.

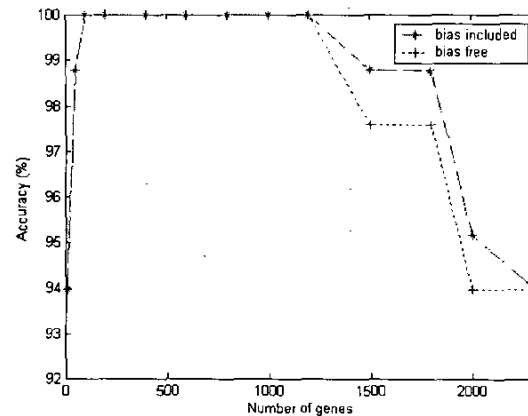


Fig. 2. The classification accuracy as a function of the number of informative genes.

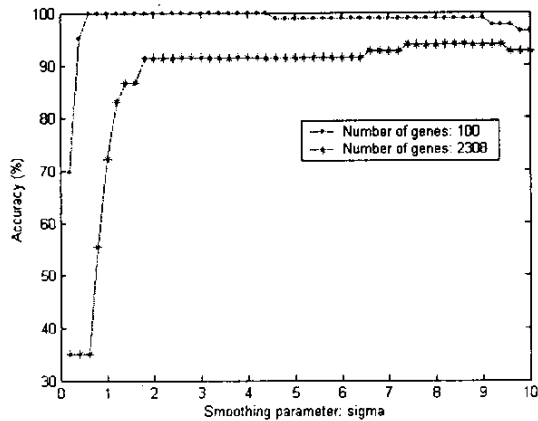


Fig. 3. The classification accuracy as a function of  $\sigma$ .

The best performance we obtain for the GCM14 data set is summarized in Table I, in which the numbers along the diagonal indicate the correct assignment of samples by PNN. For each one-versus-all comparison, top 10 genes are selected and the corresponding total number of genes is 138. The accuracy is around 75.3% and can be compared with the result reported in the original paper that constructs a multi-class prediction scheme consisting of 14 SVMs [18]. From the table, we also can observe that PNN can correctly classify most of cases for tumor types like colorectal, lymphoma, leukemia, mesothelioma and CNS. But for bladder, renal and ovary cancer, the classifier cannot effectively discriminate them from other tumor types. The analysis is also supported by other methods such as SOFMs and SVMs [18].

TABLE I. CONFUSION MATRIX OF PNN FOR 14 TUMOR TYPES: OVERALL ACCURACY IS 75.3%

		PNN Predicted Class													
		B R	P R	L U	C O	L Y	B L	M E	U T	L E	R E	P A	O V	M L	C N S
Actual Class	BR	7			1		3								1
	PR		1 0	1			1	1	1						
	LU	3		7			2								
	CO				1 0								1		1
	LY					2 2									
	BL	5			1		4								1
	ME	2					1	6	1						
	UT							2	7						1
	LE									2 9				1	
	RE	1						1	1		4	1	3		
	PA	2										8	1		
	OV					2		2			1	1	6		
	ML								1						1 0
	CNS														1 1 9

TABLE II. CLASSIFICATION CONFIDENCE

Cancer Type	Confidence			Percentage
	First	Second	Third	
BR	7	1	2	83.3%
PR	10	0	0	71.4%
LU	7	1	1	75%
CO	10	0	0	83.3%
LY	22	0	0	100%
BL	4	3	3	90.9%
ME	6	0	1	70%
UT	7	2	0	90%
LE	29	0	0	96.7%
RE	4	3	0	63.6%
PA	8	0	1	81.8%
OV	6	4	0	83.3%
ML	10	0	0	90.9%
CNS	19	1	0	100%

Since PNN has the mechanism that can estimate the confidence of the predictions, it provides us a way to evaluate the results by comprehensively considering these values. Table II lists the predicted results with the first three largest confidence values. Almost half of the errors are caused due to the reason that the confidence value for the correct category is ranked second or third by the classifier. This suggests that it is more effective and accurate to make predictions by combining the information of confidence values.

## V. CONCLUSIONS

Cancer classification is critically important for prognosis and treatment. Microarray technologies provide a new and effective avenue to address the problem, while bringing many challenges. Here, we utilized the probabilistic neural network to distinguish tumor tissues with more than two categories, by analyzing gene expression profiling. Because of some limitations in conditions like sample collections, almost all of the publicly accessible data sets merely include a small set of samples for each tumor type, in contrast to the rapidly and persistently increasing capability of gene chip technologies that also follow the Moore's law [27]. The initial experimental results demonstrate the potential of PNN, combined with feature selection technique, in extracting useful information from these high-dimensional data sets. More experiments will be performed for further evaluation with richer data available. In the meantime, new feature selection approaches are required in order to find informative genes that are more efficient in prediction and prognosis.

## Acknowledgment

The authors wish to thank Dr. Danil Prokhorov for helpful discussions. Partial support for this research from the National Science Foundation, and from the M.K. Finley Missouri endowment, is gratefully acknowledged.

## References

- [1] M. Eisen and P. Brown, "DNA Arrays for Analysis of Gene Expression," *Methods Enzymol*, vol. 303, pp. 179-205, 1999.
- [2] R. Lipshutz, S. Fodor, T. Gingeras, and D. Lockhart, "High Density Synthetic Oligonucleotide Arrays," *Nature Genetics*, vol. 21, pp. 20-24, 1999.
- [3] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander, "Molecular Classification of Cancer: Class Discovery And Class Prediction by Gene Expression Monitoring", *Science*, 286: 531-537, 1999.
- [4] A. Alizadeh, M. Eisen, R. Davis, C. Ma, I. Lossos, A. Rosenwald, J. Boldrick, H. Sabet, T. Tran, X. Yu, J. Powell, L. Yang, G. Marti, T. Moore, J. Hudson Jr, L. Lu, D. Lewis, R. Tibshirani, G. Sherlock W. Chan, T. Greiner, D. Weisenburger, J. Armitage, R. Warnke, R. Levy, W. Wilson, M. Grever, J. Byrd, D. Bostein, P. Brown, and L. Staudt, "Distinct Types of Diffuse Large B-cell Lymphoma Identified by Gene Expression Profiling", *Nature*, vol. 403, pp.503-511, 2000.
- [5] U. Alon, N. Barkai, D. Notterman, K. Gish, S. Ybarra, D. Mack, and A. Levine, "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor And Normal Colon Tissues Probed by Oligonucleotide Arrays", *Proc. Natl. Acad. Sci. USA* 96, pp.6745-6750, 1999.
- [6] M. Schummer, W. Ng, R. Bumgarner, P. Nelson, B. Schummer, D. Bednarski, L. Hassell, R. Baldwin, B. Karlan, and L. Hood, "comparative Hybridization of An Array of 21500 Ovarian cDNA for the discovery of Genes Overexpressed in Ovarian Carcinomas," *Gene*, vol. 238, pp. 375-385, 1999.
- [7] C. Perou, T. Sorlie, M. Eisen, M. Rijn, S. Jeffrey, C. Rees, J. Pollack, D. Ross, J. Johnsen, L. Akslen, Ø. Fluge, A. Pergamenschikov, C. Williams, S. Zhu, P. Lonning, A. Børresen-Dale, P. Brown, and D. Botstein, "Molecular Portraits of Human Breast Tumors," *Nature*, vol. 406, pp. 747-752, 2000.
- [8] M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, V. Sondak, N. Hayward, and J. Trent, "Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling," *Nature*, vol. 406, pp. 536-540, 2000.
- [9] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. Olson, J. Marks, and J. Nevins, "Predicting the Clinical Status of Human Breast Cancer by Using Gene Expression Profiles," *PNAS*, vol. 98, no. 20, pp. 11462-11467, 2001.
- [10] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and Z. Yakhini, "Tissue Classification with Gene Expression Profiles", *Proceedings of the Fourth Annual International Conference on Computational Molecular biology*, pp.598-583, 2000.
- [11] R. Xu, G. Anagnostopoulos and D. Wunsch II, "Tissue Classification Through Analysis of Gene Expression Data Using A New Family of ART Architectures", *IJCNN02*, vol. 1, pp. 300-304, 2002.
- [12] J. Khan, J. Wei, M. Ringnér, L. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. Antonescu, C. Peterson, and P. Meltzer, "Classification and Diagnostic Prediction of Cancers Using Gene Expression Profiling and Artificial Neural Networks," *Nature Medicine*, vol. 7, no. 6, pp. 673-679, 2001.
- [13] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Diagnosis of Multiple Cancer Types by Shrunk Centroids of Gene Expression," *PNAS*, vol. 99, no. 10, pp. 6567-6572, 2002.
- [14] D. Nguyen and D. Rocke, "Tumor Classification by Partial Least Squares using Microarray Gene Expression Data," *Bioinformatics*, vol. 18, no. 1, pp. 39-50, 2002.
- [15] Uwe Scherf, Douglas T. Ross, Mark Waltham, Lawrence H. Smith, Jac K. Lee, Lorraine Tanabe, Kurt W. Kohn, William C. Reinhold, Timothy G. Myers, Darren T. Andrews, Dominic A. Scudiero, Michael B. Eisen, Edward A. Sausville, Yves Pommier, David Botstein, Patrick O. Brown, and John N. Weinstein, "A Gene Expression Database for The Molecular Pharmacology of Cancer", *Nature Genetics*, 24(3), pp.236-44, 2000.
- [16] T. Furey, N. Cristianini, N. Duffy, D. Bednarski, M. Schummer, and D. Haussler, "support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data," *Bioinformatics*, vol. 16, no. 10, pp. 906-914, 2000.
- [17] P. Baldi, A. D. Long, "A Bayesian Framework for The Analysis of Microarray Expression Data: Regularized t-test And Statistical Inferences of Gene Changes", *Bioinformatics*, 17:509-519, 2001.
- [18] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. Mesirov, T. Poggio, W. Gerald, M. Loda, E. Lander, and T. Golub, "Multiclass Cancer Diagnosis Using Tumor Gene Expression Signatures," *PNAS*, vol. 98, no. 26, pp. 15149-15154, 2001.
- [19] C. Yeang, S. Ramaswamy, P. Tamayo, S. Mukherjee, R. Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov, and T. Golub, "Molecular Classification of Multiple Tumor Types," *Bioinformatics*, vol. 17, pp. s316-s322, 2001.
- [20] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, 2<sup>nd</sup> Ed., Wiley & Sons, New York, 2001.
- [21] D. Specht, "Probabilistic Neural Networks," *Neural Networks*, vol. 3, pp. 109-118, 1990.
- [22] E. Saad, D. Prokhorov, and D. Wunsch, "Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1456-1470, 1998.
- [23] D. Specht, "PNN: From Fast Training to Fast Running," In *Computational Intelligence: A Dynamic System Perspective*, IEEE Press, NY, pp. 246-258, 1995.
- [24] S. Hart, R. Shaffer, S. Rosch-pehrsson, and J. McDonald, "Using Physics-Based Modeler Outputs to Train Probabilistic Neural Networks for Unexploded Ordnance (UXO) Classification in Magnetometry Surveys," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 4, pp. 797-804, 2001.
- [25] D. Specht, "Enhancement to Probabilistic Neural Networks," *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 761-768, 1992.
- [26] K. Yeung and W. Ruzzo, "Principal Component Analysis for Clustering Gene Expression Data," *Bioinformatics*, vol. 17, no. 9, pp. 763-774, 2001.
- [27] S. Moore, "Making chips to probe genes," *IEEE Spectrum*, vol. 38, pp. 54-60, 2001.