International Journal of Scientific and Technical Advancements
IJSTA
ISSN: 2454-1532

Neural Network for Diagnosis of Ovarian Cancer based on Proteomic Patterns in Serum

Anjali Sharma[1], Satnam Singh[2]
[1, 2]Department of ECE, SSCET, Badhani, Punjab, India
Email address: [1]engineeranjalisharma@gmail.com

***Abstract*—** Ovarian Cancer detection is an emerging research area as it is the most common and fifth most common cause of death in women. The diagnosis of early stage ovarian tumour would significantly decrease the morbidity and mortality rate from this disease. In order to find the cure it is necessary to quickly diagnose the disease accurately and treat it based on the kind of symptoms appeared. This paper provides the way to diagnose the ovarian cancer using proteomic patterns in serum. The paper use neural network for classification of affected and not affected persons.

***Keywords*—** Neural network; ovarian tumour; CA-125; SELDI.

I. INTRODUCTION

The diagnosis of complex genetic diseases like tumor has conventionally been done based on the non-molecular characteristics like kind of tumor tissue, pathological characteristics and clinical phase. Ovarian tumor precedents to almost 27% of all mortalities, making it the leading cause of death in America and also around the world. Timely and exact detection of tumor is life-threatening to the comfort of patients. Examinations of gene expression data precedents to cancer recognition and categorization, which will make ease appropriate treatment selection and drug development. Recognition of the signals that are symptoms for the disease phenotype and its progression requires the use of hardy techniques.

The advancements in technology and modern diagnostic systems made possible the thorough investigation of ovary but there are still unsolved problems. Ovarian tumor has an unknown natural evolution, starting often insidiously, without specific symptoms; the diagnosis is put during a routine exam [5-6]. Although it was tried to associate precursor lesions to the disease, the results were not conclusive, cellular changes can be incriminated also in other non-tumor pathologies. We motivate towards this topic due to the alarming increase in the number of cases in the last 20 years and becoming the main cause of death from malignancy in gynecology.

Serum proteomic profiling, by using surfaced-enhanced laser desorption mass spectrometry is one of the most promising new techniques for cancer diagnostics. Exceptional sensitivities and specificities have been reported for some cancer types such as prostate, ovarian, breast, and bladder cancers [2]. These sensitivities/specificities are far superior to those obtained by using classical cancer biomarkers.

II. MASS SPECTROMETRIC BIOMARKER

This approach represents a paradigm shift in cancer diagnostics, based on complex mass spectrometric differences between proteomic patterns in serum between patients with or without cancer identified by bioinformatics. Their premise is that no matter what the nature of these molecules are, their potential to discriminate between these two conditions should be further exploited.

The central hypothesis of this approach is as follows: protein or protein fragments produced by cancer cells or their microenvironment may eventually enter the general circulation. Then, the concentration (abundance) of these proteins/fragments could be analysed by mass spectrometry and used for diagnostic purposes, in combination with a mathematical algorithm.

The vast majority of the currently available data have been produced by using the SELDI-TOF technology, marketed by Ciphergen Biosystems (Fremont, CA). Ciphergen claims that over 200 papers have already been published with this technology.

The types of cancers that have been examined include ovarian, prostate, breast, bladder, renal, and others, and the biological fluids analysed include serum, urine, cerebrospinal fluid, nipple aspirate fluid, etc. The apparent successes with this technology have been recently reviewed by many investigators. In general, it has been suggested that this technology can achieve much higher diagnostic sensitivity and specificity (approaching 100%) in comparison to the classical cancer biomarkers [6]. The technology's potential has been expanded to other diseases such as Alzheimer's disease, Creutzfeldt-Jakob disease, renal allograft rejection, etc.

The analytical procedure with this technology involves a few common steps. The biological fluid of interest is first interacted with a protein chip that incorporates some kind of an affinity separation between "non-informative" and "informative" proteins. After washing, the immobilized (and fortunately mostly informative) proteins can be studied by using SELDI-TOF mass spectrometry.

Two types of data have been reported in the literature: 1) discriminating peaks of unknown identity that are different in amplitude (increased or decreased) between normal individuals and patients with cancer; and 2) data in which at least some of these peaks have been positively identified (see below). Computer algorithms have been used to analyze these multidimensional data to demonstrate that a pattern consisting of several peaks (from tens to thousands) is sufficiently

Anjali Sharma and Satnam Singh, "Neural network for diagnosis of ovarian cancer based on proteomic patterns in serum ," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 2, pp. 25-27, 2016.

different between the two groups of subjects. This technology is now seen as the most promising way of diagnosing early cancer [7].

Clinical trials are now underway and will reveal, in a blinded fashion, if these data can be reproduced and if they are robust enough for clinical use. The use of SELDI-TOF technology as a cancer biomarker discovery tool (as opposed to a cancer diagnostic tool) is straightforward. The discriminatory peaks, if positively identified, may represent molecules that could be measured with simpler and cheaper techniques for the purpose of diagnosing cancer.

For example, some investigators postulate that such molecules may be routinely quantified by using enzyme-linked immune-sorbent assay (ELISA) technologies [8]. In practice, very few, if any, of the SELDI-TOF identified novel candidate biomarkers have been validated by using alternative technologies.

III. Implemented System

In this paper, we demonstrate using a neural network to detect cancer from mass spectrometry data on protein profiles. Serum proteomic pattern diagnostics can be used to differentiate samples from patients with and without disease. The goal is to build a classifier that can distinguish between cancer and control patients from the mass spectrometry data. The methodology followed is to select a reduced set of measurements or "features" that can be used to distinguish between cancer and control patients using a classifier. These features will be ion intensity levels at specific mass/charge values. we download and uncompress the raw mass-spectrometry data from the FDA-NCI web site. The new file contains variables Y, MZ and grp. Each column in Y represents measurements taken from a patient. Each row in Y represents the ion intensity level at a specific mass-charge value indicated in MZ. The variable 'grp' holds the index information as to which of these samples represent cancer patients and which ones represent normal patients.

This is a typical classification problem in which the number of features is much larger than the number of observations, but in which no single feature achieves a correct classification, therefore we need to find a classifier which appropriately learns how to weight multiple features and at the same time produce a generalized mapping which is not over-fitted.

A 1-hidden layer feed forward neural network with 5 hidden layer neurons is created and trained. The input and target samples are automatically divided into training, validation and test sets. The training set is used to teach the network. Training continues as long as the network continues improving on the validation set. The test set provides a completely independent measure of network accuracy. The input and output have sizes of 0 because the network has not yet been configured to match our input and target data. Performance is measured in terms of mean squared error, and shown in log scale. It rapidly decreased as the network was trained. Performance is shown for each of the training, validation and test sets. The version of the network that did best on the validation set is was after training.

The trained neural network can now be tested with the testing samples we partitioned from the main dataset. The testing data was not used in training in any way and hence provides an "out-of-sample" dataset to test the network on. This will give us a sense of how well the network will do when tested with data from the real world. The network outputs will be in the range 0 to 1, so we threshold them to get 1's and 0's indicating cancer or normal patients respectively.

Result

We measure how well the NN has fit the data is the confusion plot. Here the confusion matrix is plotted across all samples. The confusion matrix shows the percentages of correct and incorrect classifications. Correct classifications are the green squares on the matrices diagonal. Incorrect classifications form the red squares. If the network has learned to classify properly, the percentages in the red squares should be very small, indicating few misclassifications. If this is not the case then further training, or training a network with more hidden neurons, would be advisable.

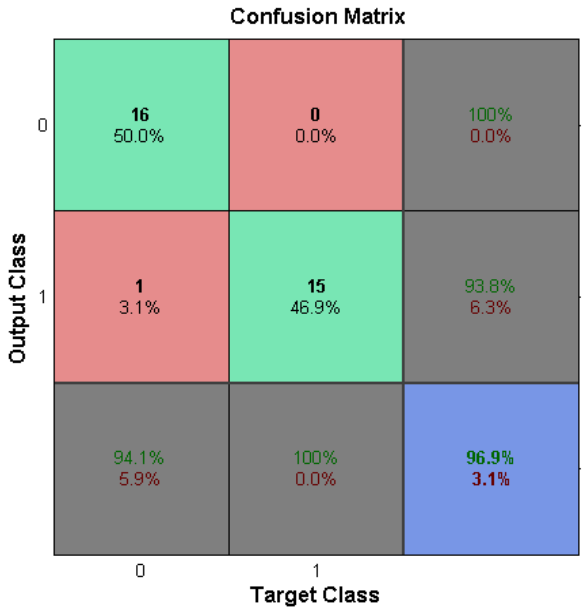

Fig. 1. Confusion matrix.

Here are the overall percentages of correct and incorrect classification.

Percentage Correct Classification: 96.87%

Percentage Incorrect Classification: 3.12%

IV. Conclusion

This paper describes ovarian tumor and its detection techniques. We illustrated how NN can be used as classifiers for cancer detection. The performance of the neural network is evaluated in this paper. We diagnose the ovarian cancer based on the blood mass-spectrum curve and identified the relevant points of the curve. The microarray gene data must be pre-processed for classification with good accuracy using the classifier. The neural networks based system gives high

Anjali Sharma and Satnam Singh, "Neural network for diagnosis of ovarian cancer based on proteomic patterns in serum ," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 2, pp. 25-27, 2016.

accuracy and good success results rate with 98% of performance for classification when compared to the conventional techniques.

REFERENCES

[1] Y. Ireaneus, A. Rejani, and Dr. S. T. Selvi, "Early detection of breast cancer using svm classifier technique," *International Journal on Computer Science and Engineering*, vol 1, issue 3, pp. 127-130, 2009.

[2] Y. Wang, I. V. Tetko, M. A. Hall, E. Frank, A. Facius, K. F. Mayer, and H. -W. Mewes, "Gene selection from microarray data for cancer classification —a machine learning approach," *Computational Biology and Chemistry*, vol. 29, issue 1, pp. 37-46, 2005.

[3] F. Chu and L. Wang, "Applications of support vector machines to cancer classification with microarray data," *International Journal of Neural Systems*, vol. 15, no. 6, pp. 475–484, 2005.

[4] H. Xiong and X. W. Chen, "Optimized kernel machines for cancer classification using gene expression data," *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1-7, 2005.

[5] L. Shen and E. C. Tan, "Dimension reduction-based penalized logistic regression for cancer classification using microarray data," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 2, no. 2, pp. 166-175, 2005.

[6] F. Chu and L. Wang, "Applying RBF neural networks to cancer classification based on gene expressions," *International Joint Conference on Neural Networks*, 2006.

[7] Zhang, G. B. Huang, N. Sundararajan, and P. Saratchandran, "Multi category classification using an extreme learning machine for microarray gene expression cancer diagnosis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 3, pp. 485–495, 2007.

[8] R. Xu, G. C. Anagnostopoulos, and D. C. I. I. Wunsch, "Multi class cancer classification using semi supervised ellipsoid artmap and particle swarm optimization with gene expression data," *IEEE/ACM Transactions on Computational Biology And Bioinformatics*, vol.4, no.1, pp. 65-77, 2007.

[9] L. Wang, F. Chu, and W. Xie, "Accurate cancer classification using expressions of very few genes," *IEEE/ ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, pp. 40-52, 2007.

[10] X. Wang and O. Gotoh, "Cancer classification using single genes," *Genome Informatics*, vol. 23, pp.179-188, 2009.

[11] X. Hang, "Cancer classification by sparse representation using microarray gene expression data," *IEEE International Conference on Bioinformatics and Biomedicine Workshops*, pp. 174-177, 2008.

[12] M. Rangasamy and S. Venketraman, "An efficient statistical model based classification algorithm for classifying cancer gene expression data with minimal gene subsets," *International Journal of Cyber Society and Education*, vol. 2, no. 2, pp. 51-66, 2009.

Anjali Sharma and Satnam Singh, "Neural network for diagnosis of ovarian cancer based on proteomic patterns in serum ," *International Journal of Scientific and Technical Advancements*, Volume 2, Issue 2, pp. 25-27, 2016.