19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Neural Network Techniques for Cancer Prediction: A Survey

Shikha Agrawal, Jitendra Agrawal

Department of Computer Science and Engineering, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India

Abstract

Cancer is a dreadful disease. Millions of people died every year because of this disease. It is very essential for medical practitioners to opt a proper treatment for cancer patients. Therefore cancer cells should be identified correctly. Neural networks are currently a burning research area in medical science, especially in the areas of cardiology, radiology, oncology, urology and etc. In this paper, we are surveying various neural network technologies for classification of cancer. The main aim of this survey in medical diagnostics is to guide researchers to develop most cost effective and user friendly systems, processes and approaches for clinicians**.**

Keywords: Artificial Neural Network; Cancer Prediction; Classification;

1. Introduction

Cancer is a major root cause of disease among human deaths in many developed countries. Cancer classification in medical practice trusted on clinical and histopathological facts may produce incomplete or misleading results .The DNA microarray is very useful to determine the expression level of thousands of genes simultaneously in a cell mixture [1]. DNA microarray technology has been applied to find out the accurate prediction and diagnosis of cancer. Molecular level diagnostics with gene expression profiles can offer the methodology of accurate and systematic cancer classification. It's very important for treatment of cancer to classify tumor accurately. Since the gene expression data generally comprise of huge number of genes, several scholars have been scrutinizing the problems of cancer classification using data mining approaches, statistical methods and machine learning algorithms to effectually evaluate these data [2].Various machine learning techniques are used for cancer classification, such as support vector machine, k-nearest neighbor and neural network techniques etc. Neural network techniques are very useful for detection and monitoring of cancer. Artificial neural network is a robust tool recently used as either clustering or classification of gene expression data. Supervised models are used for classification while unsupervised models are used for clustering. Neural network model has been successfully implemented in various classification problems. Classification is crucially important for cancer diagnosis and treatment. A classification problem occurs when an object needs to be allotted into a predefined group or class based on a number of observed attributes related to that object. Clustering can be considered as unsupervised learning problem, it deals with finding a structure in collection of unlabeled data. Artificial neural network had been proven a very effective method for pattern recognition. This made them very useful for diagnosis of cancer disease at very early stages.

2. Neural Network

An Artificial Neural Network (ANN)[7] , usually known as a neural network, is basically a mathematical model motivated by biological nervous systems like brain processes information. A neural network comprises an interconnected group of simulated neurons and it uses connectionist approach to process information for computation. Neural network works like an adaptive system, which changes its structure in learning phase. Simple and complex relationships can be easily modeled using neural networks. They are also used to find out patterns and clusters in data. An ANN can be designed for a particular application, such as data classification and pattern categorization through a learning process. There are several types of neural network configurations. The arrangement of neurons to form layers and the connection pattern formed within and between layers is called the network structures. An ANN can be classified into following categories:

- Single layer feed-forward network
- Multilayer feed- forward network
- Single node with its own feedback
- Multilayer recurrent network

The main property of an ANN[2] is its ability of learning. Training or learning is a procedure of parameter tuning by which a neural network adapts itself to a stimulus and then desired output is produced. Broadly, there are three types of learning:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised learning is performed in the presence of a teacher. In this type of training, a supervisor or teacher is required for error minimization. It is assumed that the correct "target" output values are known for each input pattern. Some supervised learning networks are:

- Perceptron networks
- Adaptive linear neuron (Adaline)
- Multiple Adaptive linear neurons(Madaline)
- Back-Propagation network
- Radial basis Function Network
- Time delay neural network
- Functional link network
- Tree Neural Networks
- Wavelet Neural Networks

Unsupervised leaning is performed without the help of instructor. In this situation, the network itself find features, categories, patterns or regularities from the input data and relations for the input data over the output. Some unsupervised learning networks are:

- Kohonen Self-Organizing Feature Maps
- Learning vector Quantization(LVQ)
- Counter propagation networks(CPN)
- Adaptive resonance Theory network.

Advantages of ANN are as follows:

- Adaptive learning: An ANN has the ability to learn how to do tasks based on the data given for training   or initial experience.
- Self-organization: After receiving the information in learning time an ANN can create its own organization..
- Real-time operation: Many neural network computations can be carried out parallel. Specific hardware devices are being designed to take benefit of this ability of neural networks.
- Fault-tolerance via redundant information coding: Partial damage of a neural network structures leads to the degradation of performance. Though, some network abilities may be recollected even after major network damage.

Disadvantages of ANN are as follows:

- Size and Complexity: Neural networks size and complexity is very high.

- Training time is very large.

3. Literature Survey

Hu et al., [3] proposed a technique to classify cancer using supervised and unsupervised learning methods .In supervised learning, a single hidden layer feed forward NN with back-propagation training is applied for error minimization. Various activation functions such as sigmoid, sinusoid and Gaussian are tested to establish different network configurations. In unsupervised learning, fuzzy and nonfuzzy and c-means clustering methods were used. Several characteristics such as cell size, average intensity, texture, shape factor and pgDNA are used as the input for the network. Results are evaluated using 467 images from six different types of tumour classes. 96.9% classification rate is obtained using neural network techniques, while fuzzy c-means results only 76.50%.

Conventional methods for prediction of cancer in medical practice can be often imperfect or confusing while molecular level diagnostics with gene expression microarray technique is very precise, objective and efficient for cancer classification. Won et al., [4] presented a technique using ensemble of neural network classifiers studied from negatively correlated characteristics to accurately categorize cancer and it estimates the functioning of the proposed technique with using three benchmark datasets. Experimental performance proves the ensemble classifier with negatively correlated characteristics provides best recognition rate on these benchmark datasets.

Xu et al., [5] presented A hybrid method of probabilistic neural network (PNN) and discrete binary version of Particle swarm optimization. PSO is basically used for optimal selection of genes and dimensionality reduction. Feed forward neural network is used to implement neural network configurations. This method is efficiently experimented on large B-cell lymphoma dataset with 80% classification accuracy.

Bevilacqua et al., [6] described a new approach to classify cancer using ANN .ANN topology is optimized through multi objective genetic algorithm. Wisconsin breast cancer database (WBCD) is used in the classification problem. This database contains 699 cases. WBCD basically consist of two types of tumour classes, benign and malignant tumors. ANN is used for categorizing cases while multiobjective genetic algorithm is used for refining the search space and for discovering an optimal topology for ANN configuration.

Ziaei et al., [7] introduced a new method for prediction of cancer with the help of perceptron network. This network is tested on Diffuse large B-cell lymphoma (DLBCL) database. They observed 4026 genes based on their ranking, which is calculated according to their signal to noise ratios. A threshold value is obtained and those genes are removed whose ratios were less than the threshold value. Perceptron network is applied as a classifier. Thus, patients were classified with accuracy of 93%.

Hiro Takahashi et al., [8] proposed an innovative hybrid method of projective Adaptive Resonance Theory (ART) and boosted fuzzy classifier with SWEEP operator method for detection of cancer classes. They employed this method to microarray data of acute leukemia and brain tumor. Tumor is accurately classified over the classification rate of 90%. They combined wrapper and filter approaches for applying these methods to gene expression microarray data of leukemia and central nervous system tumor.

Cho et al., [9] proposed an approach using ensemble of neural networks trained from multiple significant gene subsets to identify cancer classes effectively. This method is experimented on three types of cancer data Leukaemia, Colon and Lymphoma data sets. They were used 3-layered multi-layer perceptron with 30 hidden nodes, 2 output nodes. Neural network is trained using error back propagation algorithm with the selected gene subsets.

A novel hybrid technique is adopted by Yuchun et al., [10]. This technique is a combination of granular computing, fuzzy clustering and statistical learning. Recursive feature elimination algorithm is also used in this approach. This algorithm eliminates unnecessary and noisy genes and selects only informative genes among thousands of genes. FG-SVM approach performs 100% accurately with three different open database sets.

A brain tumor identification and detection system is recommended by Joshi et al., [11]. Here, MRI images of several cancer patients are analyzed by using neural network. Various image processing techniques like image segmentation, morphological functions, histogram equalization, image enrichment and feature extraction are applied on those MRI images. Neuro fuzzy classifier is used for detection of brain tumor cells. This experiment provides good classification accuracy with Neuro fuzzy classifier.

P. Rajeswari G. and Sophia Reena[12] presented a technique for identification of tumor cells. Support Vector Machine (SVM) and FNN (Fuzzy Neural Network) are used in the classification problem. Liver cancer data set is used for testing the proposed technique .The experimental results indicate that the proposed technique has the ability to classify the cancer cells appreciably when it is compared with the conventional methods of cancer classification. Various ranking techniques are described in this paper for discriminatory selection of genes. Fuzzy Neural network provides the performance rate between 92-96%.Cancer microarray data is very high dimensional. Therefore feature selection process plays an important role.

Barnalisahu et al., [13] proposed a novel feature selection approach for the classification of high dimensional cancer microarray data, gene ranking is carried out using filtering technique such as signal-to-noise ratio (SNR) score and optimization technique as Particle swarm Optimization (PSO) is used for dimensionality reduction. Support vector machine (SVM), k-nearest neighbor (k-NN) and Probabilistic Neural Network (PNN) are used as classifiers. They used 4 microarray datasets. Leukemia, Colon, DLBCL and Breast Cancer data .Probabilistic neural network results in 96% accuracy.

Swati et al., [14] discussed ART1 network for detection of breast cancer. This method is processed in three steps such as:

recognition, comparison and search phases. Here, Back propagation algorithm is used for error minimization and training of network design. Simulation results show the efficiency of ART1 network. For unsupervised learning pattern, they got good results for breast cancer datasets with 92% accuracy.

Several Neural network technologies have been used till date for cancer classification. Dev et al.,[15] presented three different techniques for detection of tumor cells .Here, Back propagation network(BPN),Functional Link Artificial neural Network(FLANN) and PSO-FLANN are used for breast cancer classification .In this paper results of these three classifiers are analyzed and compared. Reported result for Classification rate of BPN is 56.12% and FLANN shows 63.34% whereas PSO-FLANN provides best classification rate with 92.36%.

Sharma and Kaur [16] proposed a novel method in which CT scan images of tumors are classified using Particle Swarm Optimization (PSO) and Seeker Optimization algorithm (SOA) algorithms. There are various Image segmentation approaches like Region based, Threshold based, Level set, and Clustering based and Edge detection. Here region based approach is used for segmentation. The prime objective of this research is to identify liver tumor and analyze the results of PSO and SOA in terms of detection and classification accuracy and elapsed time .Performance rate of PSO is 93.30% while SOA scores only 60% as classification accuracy. Tabel1 below summarizes the research work discussed in this paper.

Table 1: Summary of Research Work Discussed

TItle	Authors	Methods	Accuracy	Year
A comparison of neural network and fuzzy c-means methods in bladder cancer cell classification	Hu, Y, Ashenayi, K., Veltri, R., O'Dowd, G., Miller, G.,Hurst, R.,Bonner	Single hidden layer feed forward NN with error back-propagation	96.9%	1994
Paired neural network with negatively correlated features for cancer classification in DNA gene	Hong-Hee Won, Sung-Bae Cho	Feed forward multi-layer perceptron	97.1%	2003
Gene Expression Data for DLBCL Cancer Survival Prediction with A Combination f Machine Learning Technologies	RuiXu, XindiCai, Donald C. Wunsch II	Probabilistic neural network (PNN)	80%	2005
A Novel Multi-Objective Genetic Algorithm Approach to Artificial Neural Network Topology Optimization: The Breast Cancer Classification Problem	Bevilacqua, V, Mastronardi, G,Menolascina F,Pannarale, P, Pedone.	IDEST(An intelligent system) and GA(genetic algorithm)	98.60%	2006
Application of Artificial Neural Networks in Cancer Classification and Diagnosis Prediction of a Subtype of Lymphoma Based on Gene Expression Profile	L. Ziaei , A. R. Mehri , M. Salehi	Perceptron network	93%	2007
New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index"	Hiro Takahashi , Yasuyuki Murase , Takeshi Kobayashi, Hiroyuki Honda	Adaptive resonance theory and boosted fuzzy classifier	90%	2007
Cancer classification using ensemble of neural networks with multiple significant gene subsets	Sung-Bae Cho,Hong-HeeWon	Multi-layer perceptron with error back propagation algorithms	87-93%	2007
Recursive Fuzzy Granulation for Gene Subsets Extraction and Cancer Classification	Yuchun Tang; Yan-Qing Zhang, Zhen Huang,Xiaohua Hu,Yichuan Zhao	(FG-SVM-RFE)	100%	2008
Classification of Brain Cancer using Artificial Neural Network	Joshi D.M., Rana N.K, Misra, V.M	Neuro Fuzzy network	--	2010
Human Liver Cancer Classification using Microarray Gene Expression Data	P. Rajeswari, G. Sophia Reena	FNN(Fuzzy Neural Network).	92-96%	2011
A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data	Barnali Sahu, Debahuti Mishra	Probabilistic Neural Network (PNN)	96%	2012
Performance of ART1 Network in the Detection of Breast Cancer	S. Swati et al	ART1 network	92%	2012
A Classification Technique for Microarray Gene Expression Data using PSO-FLANN	Jayshree Dev et al	Functional Link FLANN and PSO	92.36%	2012
Optimized Liver Tumor Detection and Segmentation Using Neural Network	Akanksha Sharma, Parminder Kaur	Seeker Optimization algorithm and PSO	60%	2013

4. Conclusion

Cancer is one of the dreadful diseases. Diagnosis of cancer is very important in initial stage for its proper treatment. Cancer data is a collection of thousands of genes. DNA microarray is used to determine the expression level of genes. Analysis of microarray gene expression data is very difficult due to its sparse and excessive characteristics. Selection of informative genes among thousands of genes is very challenging task. By analysing these gene expression data, heterogeneous cancer can be classified into their proper subgroups. Nowadays, various kinds of machine learning and statistical approaches are used to classify tumour cells accurate such as support vector machines, k-nearest neighbour, decision trees and neural network techniques. Recently, many researchers are showing their interest in neural network techniques to classify cancer cells.

This survey clearly demonstrates the effectiveness of neural network technologies in the detection of cancer. Most of the neural network show tremendous result to classify tumour cells accurately. Especially MLP (Multi Layer Perceptron) gives 97.1% accuracy and PNN(Probabilistic Neural Network) which provides 96% accuracy, Perceptron with 93 % and ART1 shows 92% accuracy as well. After removing missing values from dataset experimental results get improved. Results of neural network structures can be enhanced by proper settings of neural network parameters. Although neural network techniques provide good classification rate, but their training time is very high.

Several researchers thus hybridize neural network techniques with optimization algorithms like PSO for further enhancement of accuracy. These optimization algorithms are used for dimensionality reduction, they suppresses search space and therefore, reduces the training time of neural network. FLANN alone shows 63.4% accuracy whereas PSO-FLANN provides good classification rate with 92.36%.

In future study, accuracy of neural network can be enhanced by increasing the number of neurons in the hidden layer .Different training and learning rules can be applied for training ANN in order to improve the performance of classifier.
.

References

1. Y Wang, F. S. Makedon , J. C. Ford, J. Pearlman. Hyk Gene: A hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data; Bioinformatics; 15; 2005; p. 1530-1537.

2. Sung-Bae Cho , Hong-HeeWon .Cancer classification using ensemble of neural networks with multiple significant gene subsets; ApplIntell; 26; 2007; p. 243–250.

3.Hu Y., Ashenayi K., Veltri, R., O'Dowd G., Miller G., Hurst R. Bonner. A comparison of neural network and fuzzy c-means methods in bladder cancer cell classification; In Proceedings of IEEE World Congress on Computational Intelligence; 6; 1994; p. 3461 – 3466.

4.Hong-Hee Won, Sung-Bae Cho. Paired neural network with negatively correlated features for cancer classification in DNA gene expression profiles; In Proceedings of the International Joint Conference on Neural Networks; 3; 2003; p. 1708 – 1713.

5.Rui Xu, Xindi Cai, Donald C. , Wunsch II. Gene Expression Data for DLBCL Cancer Survival Prediction with A Combination of Machine Learning Technologies; In Proceedings of the IEEE International Conference on Medicine and Biology, 2005, p. 894-897.

6.Bevilacqua, V., Mastronardi, G., Menolascina, F., Pannarale P., Pedone. A Novel Multi-Objective Genetic Algorithm Approach to Artificial Neural Network Topology Optimisation; The Breast Cancer Classification Problem; In Proceedings of International Joint Conference on Neural Networks (IJCNN '06);2006; p. 1958 – 1965.

7. L. Ziaei , A. R. Mehri , M. Salehi. Application of Artificial Neural Networks in Cancer Classification and Diagnosis Prediction of a Subtype of Lymphoma Based on Gene Expression Profile; Journal of Research in Medical Sciences; 11(1), 2006; p. 13-17.

8. Hiro Takahashi, Yasuyuki Murase , Takeshi Kobayashi , Hiroyuki Honda. New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index; Biochemical Engineering Journal 33; 2007; p. 100–109.

9.Sung-Bae Cho, Hong-Hee Won. Cancer classification using ensemble of neural networks with multiple significant gene subsets" C _Springer Science and Business Media; ApplIntell; 26; 2007; p. 243–250.

10. Yuchun Tang, Yan-Qing Zhang, Zhen Huang, Xiaohua Hu, Yichuan Zhao. Recursive Fuzzy Granulation for Gene Subsets Extraction and Cancer Classification; IEEE Transactions on Information Technology in Biomedicine; 2(6); 2008; p. 723 – 730.

11. Joshi D.M, Rana N.K., Misra V.M. Classification of Brain Cancer using Artificial Neural Network; In Proceedings of International Conference on Electronic Computer Technology (ICECT); 2010;p. 112 – 116.

12. P. Rajeswari, G. Sophia Reena. Human Liver Cancer Classification using Microarray Gene Expression Data; In Proceedings of International Journal of Computer Applications; 34(6); 2011;p. 25-37.

13.Sahu Barnali, Mishra Debahuti . A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data; In Proceedings of International Conference on Modeling Optimization and Computing (ICMOC-2012); Procedia Engineering; 38; 2012 ;p. 27 – 31.

14.S. Swathi, G. Anjan Babu, R. Sendhilkumar, Sreenu , Naik Bhukya. Performance of ART1 Network in the Detection of Breast Cancer; In Proceedings of International conference on Computer design and Engineering(ICCDE 2012);49; 2012; DOI:10.7763/IPCSIT.2012.V49.19;p.100-105.

15. Dev Jayshree, Dash Sanjitkumar ,Dash Sweta ,Swain Madhusmita. A Classification Technique for Microarray Gene Expression Data using PSO-FLANN; In Proceedings of International Journal on Computer science and Engineering; 4(9); 2012;p. 1534-1535.

16.Sharma Akanksha, Kuar Parminder. Optimized Liver Tumor Detection and Segmentation Using Neural Network; In Proceedings of International Journal of Recent Technology and Engineering (IJRTE); 2(5);2013;p. 7-10.