

BREAST CANCER DIAGNOSIS USING MACHINE LEARNING ALGORITHMS –A SURVEY

B.M.Gayathri.¹, C.P.Sumathi² and T.Santhanam³

¹Department of Computer Science, SDNB Vaishnav College for Women,
Chennai, India

gayathri_bm2003@yahoo.co.in

²Department of Computer Science, SDNB Vaishnav College for Women,
Chennai, India

drpcsumathi@gmail.com

³Department of Computer Application, D.G.Vaishnav College for Men, Arumbakkam,
Chennai, India

santhanam_dgvc@yahoo.com

ABSTRACT

Breast cancer has become a common factor now-a-days. Despite the fact, not all general hospitals have the facilities to diagnose breast cancer through mammograms. Waiting for diagnosing a breast cancer for a long time may increase the possibility of the cancer spreading. Therefore a computerized breast cancer diagnosis has been developed to reduce the time taken to diagnose the breast cancer and reduce the death rate. This paper summarizes the survey on breast cancer diagnosis using various machine learning algorithms and methods, which are used to improve the accuracy of predicting cancer. This survey can also help us to know about number of papers that are implemented to diagnose the breast cancer.

KEYWORDS

Neural networks, SVM, RVM, ELM

1. INTRODUCTION

Breast cancer has become one of the most common disease among women that leads to death. Breast cancer can be diagnosed by classifying tumors. There are two different types of tumors such as malignant and benign tumors. Physicians need a reliable diagnosis procedure to distinguish between these tumors. But generally it is very difficult to distinguish tumors even by the experts. Hence automation of diagnostic system is needed for diagnosing tumors. Many researchers have attempted to apply machine learning algorithms for detecting survivability of cancers in human beings and it is also been proved by the researchers that these algorithms work better in detecting cancer diagnosis. This paper summarizes the application of machine learning algorithms in detecting cancer in human. In this survey section 2 gives the information of neural network, its learning rules. Section 3 specifies about literature review based on Artificial Neural Network (ANN). Section 4 specifies other related works on breast cancer using neural networks. Section 5 implies with other machine learning algorithms and its types, with related work on those algorithms.

1.1 Machine learning Algorithms

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases.

1.2 Types of Machine Learning Algorithms

- Supervised learning.
- Unsupervised learning.
- Semi-supervised learning.
- Reinforcement learning.
- Transduction .
- Learning to learn.

2. NEURAL NETWORKS AND ITS LEARNING RULES:

A neural network is a model that is designed by the way human nervous systems such as brain, that process the information. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Many neural network models, even biological neural networks assume many simplifications over actual biological neural networks.

Such simplifications are necessary to understand the intended properties and to attempt any mathematical analysis. Even if all the properties of the neurons are known, simplification is still needed for analytical purpose. Neural networks are adaptive statistical devices. This means that they can change (synaptic weights) as a function of their performance. In ANNs, all the neurons are operating at the same time, which makes ANN to perform tasks at much faster rate.

Table 1. Performance of neural networks algorithms for detecting breast cancer

S.no	Year	Author	Algorithms/Techniques used	Result obtained
1.	2010	F.Paulin ,Dr.A.Santhakumaran	Back propagation algorithm used for training Multilayer Perceptron(MLP)	99.28%
2.	2010	Dr.K .Usha rani	Feed forward, Back propagation.	92%
3.	2011	F.Paulin, Dr.A.Santhakumaran	Back propagation, Quasi newton, levennberg Marquardt algorithm.	99.28%
4.	2011	Yao ying huang,wang sen li ,Xiao jiao ye	Genetic algorithm, feature selection	99.1%

3. RELATED WORKS FOR BREAST CANCER DIAGNOSIS USING NEURAL NETWORKS

Tuba kiyan [2] et al. 2004 has discussed that statistical neural networks can be used to perform breast cancer diagnosis effectively. The scholar has compared statistical neural network with Multi Layer Perceptron on WBCD database. Radial basis function(RBF), General Regression Neural Network(GRNN), Probabilistic Neural Network(PNN) were used for classification and their overall performance were 96.18% for Radial Basis Function (RBF), 97% PNN, 98.8% for GRNN and 95.74% for MLP. Hence it is proved that these statistical neural network structures can be applied to diagnose breast cancer.

Xin Yao [24] et al. 1999 has attempted to implement neural network for breast cancer diagnosis. Negative correlation training algorithm was used to decompose a problem automatically and solve them. In this article the author has discussed two approaches such as evolutionary approach and ensemble approach, in which evolutionary approach can be used to design compact neural network automatically. The ensemble approach was aimed to tackle large problems but it was in progress.

Dr.S.Santhosh baboo and S.Sasikala [27] have done a survey on data mining techniques for gene selection classification. This article dealt with most used data mining techniques for gene selection and cancer classification, particularly they have focused on four main emerging fields. They are neural network based algorithms, machine learning algorithms, genetic algorithm and cluster based algorithms and they have specified future improvement in this field.

Afzan Adam[28] et al. have developed a computerized breast cancer diagnosis by combining genetic algorithm and Back propagation neural network which was developed as faster classifier model to reduce the diagnose time as well as increasing the accuracy in classifying mass in breast to either benign or malignant. In these two different cleaning processes was carried out on the dataset. In Set A, it only eliminated records with missing values, while set B was trained with normal statistical cleaning process to identify any noisy or missing values. At last Set A gave 100% of highest accuracy percentage and set B gave 83.36% of accuracy. Hence the author has concluded that medical data are best kept in its original value as it gives high accuracy percentage as compared to altered data.

David B.fogel [26] et al. has discussed the evolving neural networks for detecting breast cancer and the related works used for breast cancer diagnosis using back propagation method with multilayer perceptron. In contrast to back propagation David B.fogel et al. found that evolution computational method and algorithms were used often, outperform more classic optimization techniques.

The author has applied 699 data, which has missing values and removed, leaving 683 data. Using these values two experimental designs were conducted. The first experiment consisted of five trials with 9-2-1 Multi Layer Perceptron (i.e., 9 input, 2 hidden nodes, and 1 output node) and second experiment consisted of 9-9-1 Multi Layer Perceptron. The result of the first experiment after 400 generations in each five trials had accuracy of 97.5%. In second experiment, in comparison with previous experiment, best performance reported with an accuracy rate of 98.2% for lesser hidden nodes.

A.Punitha [15] et al. 2007 have discussed the genetic algorithm and adaptive resonance theory neural network for breast cancer diagnosis using Wisconsin Breast Cancer Data (WBCD). They trained 699 samples which was taken from Fine Needle Aspirates (FNA) with 16 missing

data, and 683 samples with breast tumors are used in this work of which 65% was proved to be benign and 35% malignant. The author has also compared the result of Adaptive Resonance Theory (ART) with Radial Basis Function (RBF), Probabilistic Neural Network (PNN), Multi Layer Perceptron (MLP), in which the performance of these combined approach has not only improved the accuracy but also reduced the time taken to train the network.

Val'erie Bourd'es [5] et al., 2010 have submitted the article by comparing artificial neural network with logistic regression. The author has compared multilayer perceptron Neural Networks (NNs) with Standard Logistic Regression (SLR) to identify key covariates impacting on mortality from cancer causes, Disease-Free Survival (DFS), and Disease Recurrence using Area Under Receiver-Operating Characteristics (AUROC) in breast cancer patients.

From 1996 to 2004, 2,535 patients diagnosed with primary breast cancer entered into the study at a single French centre, where they received standard treatment. For specific mortality as well as DFS analysis, the Receiver-Operating Characteristics (ROC) curves were greater with the NN models compared to LR model with better sensitivity and specificity. Four predictive factors were retained by both approaches for mortality: clinical size stage, Scarff Bloom Richardson grade, number of invaded nodes, and progesterone receptor. The results enhanced the relevance of the use of NN models in predictive analysis in oncology, which appeared to be more accurate in prediction in this French breast cancer cohort.

Chih-Lin Chi [4] et al., 2007 have presented an article on survival analysis of breast cancer on two breast cancer datasets. This article applies an Artificial Neural Networks (ANNs) to the survival analysis problem. Because ANNs can easily consider variable interactions and create a non-linear prediction model, they offer more flexible prediction of survival time than traditional methods. This study compares ANN results on two different breast cancer datasets, both of which use nuclear morphometric features. The results show that ANNs can successfully predict recurrence probability and separate patients with good and bad prognosis.

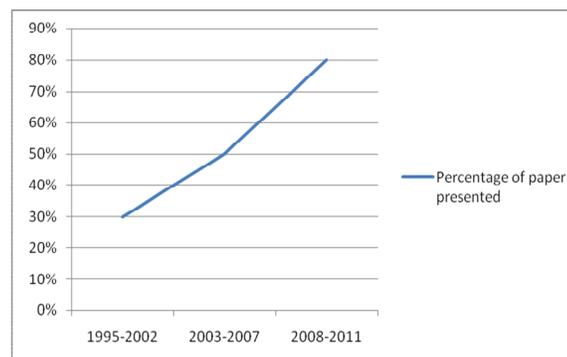


Figure 1. percentage of articles presented for diagnosing breast cancer using neural networks

From this survey, it can be found that there are many articles based on breast cancer diagnosis, which have been presented by many researchers and they are still undergoing research on developing more algorithms to get more accuracy for detecting breast cancer.

Now-a-days machine learning algorithms are mostly used for detecting the cancer disease. In the following section let us see, about other machine learning algorithms and some of the papers presented based on it.

4. OTHER MACHINE LEARNING ALGORITHMS

Some of the other machine learning algorithms are Support Vector Machine, Relevance Vector Machine. The following section gives the brief information about these learning algorithms.

4.1 SUPPORT VECTOR MACHINE

A support vector machine (SVM) is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making the SVM a non-probabilistic binary linear classifier.

Table 2. List Performance of breast cancer diagnosis using SVM

S.no	Year	Author	Algorithms	Result obtained
1.	2012	JR Marsilin	SVM	78%
2.	2011	Li Rong,Sunyuan	SVM-KNN classifier	98.06%
3.	2011	F Eddaoudi	SVM	95%
4.	2011	S. Aruna, Dr S.P. Rajagopalan	SVM	98.24%

Table 2 shows that SVM technique is used for detecting breast cancer. But Relevance vector machine (RVM) gives more accurate results than support vector machines. This has been proved by applying RVM in other cancer diagnosis such as ovarian cancer, optical cancer and general cancer classifications. Hence Relevance vector machine can also be applied to attain best result for diagnosing breast cancer.

4.1.1 RELATED WORKS ON DIAGNOSING BREAST CANCER USING SVM:

Ilias Maglogiannis [9] et al. 2009 have presented an article on An intelligent system for automated breast cancer diagnosis & prognosis using SVM based classifiers with Bayesian classifiers and ANN for prognosis & diagnosis of breast cancer disease. Wisconsin diagnostic breast cancer datasets were used to implement SVM model to provide distinction between the malignant & benign breast masses. These datasets involve measurement taken according to Fine Needle Aspirates (FNA). The article provides the implementation details along with the corresponding results for all the assessed classifiers. Several comparative studies have been carried out concerning both the prognosis and diagnosis problem demonstrating the superiority of the proposed SVM algorithm in terms of sensitivity, specificity and accuracy.

Y.Iraneus Anna Rejani and Dr.S.Thamarai selvi [12] 2009 have presented an article on Early detection of breast cancer using SVM classifier technique. In this article the authors have discussed, how to detect tumor from mammograms. In this article the authors have specified an algorithm for tumor detection and have proposed the method that includes the mammograms image, which were filtered with Gaussian filter based on standard deviation and matrix dimensions such as rows and columns.

Then the filtered image was used for contrast stretching. The background image is eliminated using Top hat operation. The top hat output is decomposed and reconstructed using Discrete Wavelet Transform (DWT). The reconstructed image is used for segmentation. Thresholding method was used for segmentation and then the features were extracted from the tumor area. This method can be summarized as the initial step based on gray level information of image enhancement. For each tumor region extract, morphological features were extracted to categorize the breast tumor and finally SVM classifiers were used for classification.

Z.Qinli [29] et al. has presented an article on, a approach to SVM and its application to breast cancer diagnosis. In this article, the authors have proposed a method for improving the performance of SVM classifier by modifying kernel functions. This is based on the differential approximation of metric. The method is to enlarge margin around separating hyper plane by modifying the kernel functions using a

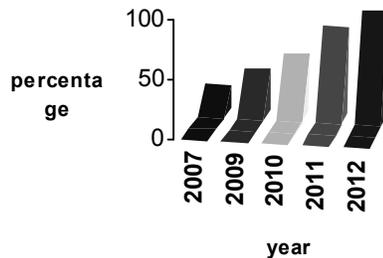


Figure 2. Percentage of articles presented for breast cancer diagnosis using SVM

positive scalar functions so that the seperability is increased. In this article, the author have specified specifically for modifying Gaussian Radial Basis function kernel. The result for both artificial and real data, show remarkable improvement of generalization error and computational cost.

4.2 RELEVANCE VECTOR MACHINE

Relevance vector machine (RVM) is a machine learning technique that uses Bayesian inference to obtain parsimonious solutions for regression and classification. The RVM has an identical functional form to the support vector machine, but provides probabilistic classification. It is actually equivalent to a Gaussian process model with covariance function:

$$k(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^N \frac{1}{\alpha_j} \varphi(\mathbf{x}, \mathbf{x}_j) \varphi(\mathbf{x}', \mathbf{x}_j)$$

where φ is the kernel function (usually Gaussian), and $\mathbf{x}_1, \dots, \mathbf{x}_N$ are the input vectors of the training set.

4.2.1 CASE STUDY ON SOME OF THE PAPERS PRESENTED FOR DETECTING CANCERS USING RVM:

Compared to that of support vector machines (SVM), the Bayesian formulation of the RVM avoids the set of free parameters of the SVM (that usually require cross-validation-based post-optimizations). However RVMs use expectation maximization (EM)-like learning method and are therefore at risk of local minima. This is unlike the standard Sequential Minimal Optimization (SMO)-based algorithms employed by SVMs, which are guaranteed to find a global optimum. RVM is used by many authors for detecting cancer in human beings. For detecting cancer such as ovarian cancers, optical cancers etc., relevance vector machine is used, which has been proved to give more accurate results than support vector machine.

S.no	Year	Author	Algorithms/Techniques used
1.	2002	Balaji Krishnapuram, Lawrence Carin, Alexander J. Hartemink	RVM(linear kernel)
2.	2003	Balaji Krishnapuram, Lawrence Carin, Alexander J. Hartemink	RVM
3.	2004	Balaji Krishnapuram, Lawrence Carin, Alexander J. Hartemink	RVM
4.	2005	Shovan K. Majumder	RVM
5.	2005	L Wei, Y Yang, RM Nishikawa	RVM
6.	2007	Wen Zhang Liu, J	RVM
7.	2009	S Ozer, MA Haider, DL Langer	RVM

Table 3. Literature review on RVM

From the above tabular column it can be found that RVM is applied for detecting cancers such as prostate cancer, optical cancer etc.,. Since RVM gives the result more accurate than SVM it can be applied for detecting breast cancer also.

5. CONCLUSION

In this survey, the performance of different machine learning algorithms such as Support Vector Machine (SVM) and Relevance Vector Machine (RVM) are assessed. Many researchers have applied the algorithm of neural networks for predicting cancers, especially the breast cancer. By going through various articles, RVM is applied for detecting optical cancer, ovarian cancer etc. Overall, if studies on RVM continues, then it is likely that the use of RVM will become much more useful in diagnosing breast cancer.

REFERENCES

- [1] M.A.H Akhand & M.Murase "Neural network ensemble construction fusing multiple popular methods.". International journal of computer science 2004.
- [2] Tuba kiyani, Tulay Yildirim "Breast cancer diagnosis using statistical neural networks", Journal of Electrical and Electronic Engineering.
- [3] P.K.Chenniappan & N.Anusheela "Early detection of cancer cells to lung cancer survival data by neural network testing" International journal of neural network and application.
- [4] Chih-Lin Chi, W. Nick Street, and William H. Wolberg " Application of Artificial Neural Network-Based Survival Analysis on Two Breast Cancer Datasets"- AMIA Annu Symp Proc. 2007; 2007: 130-134.
- [5] Val'erie Bourd'es, St'ephane Bonnevey, Paolo Lisboa, R'emy Defrance, David P'erol, Sylvie Chabaud, Thomas Bachelot, Th'er'ese Gargi,6 and Sylvie N'egrier "Comparison of Artificial Neural

- Network with Logistic regression as Classification Models for Variable Selection for Prediction of Breast Cancer Patient outcomes”
- [6] I Guyon, J Weston, S Barnhill “Gene selection for cancer classification using support vector machines” ... - Machine learning, 2002 – Springer
 - [7] MF Akay “Support vector machines combined with feature selection for breast cancer diagnosis”- Expert systems with applications, 2009 – Elsevier.
 - [8] CL Huang, HC Liao “Prediction model building and feature selection with support vector machines in breast cancer diagnosis “Expert Systems with Applications, 2008 – Elsevier.
 - [9] Ilias Maglogiannis, E Zafropoulos “An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers” Applied Intelligence, 2009 – Springer.
 - [10] AM Bagirov, B Ferguson, S Ivkovic “New algorithms for multi-class cancer diagnosis using tumor gene expression signatures”, 2003 - Oxford Univ Press.
 - [11] RF Chang, WJ Wu, WK Moon, YH Chou “Support vector machines for diagnosis of breast tumors on US images” - Academic radiology, 2003 – Elsevier
 - [12] Y Rejani- “Early detection of breast cancer using SVM”. 2009 -arxiv
 - [13] L Rong ,”Diagnosis of Breast Tumor Using SVM-KNN Classifier” - 2010 - Intelligent Systems (GCIS).
 - [14] T Mu “Detection of breast cancer using v-SVM and RBF network.” - 2005 – C Medical Applications of Signal.
 - [15] Dr.C.P.Sumathi, Dr.T.Santhanam, A.Punitha “Combination of genetic algorithm and ART neural network for breast cancer diagnosis”, 2007 -- Asian Journal of information technology, Medwell journals.
 - [16] F.Paulin, A.Santhakumaran “Classification of Breast cancer by comparing Back propagation training algorithms”, 2011-International Journal on Computer Science and Engineering (IJCSSE)
 - [17] Dr. K. Usha Rani” Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique” International Journal of Computer Applications, Volume 10– No.3, November 2010.
 - [18] G. Sophia Reena P. Rajeswari -2011 “A Survey of Human Cancer Classification using Micro Array Data”- International Journal of Computer Tech. Appl.,
 - [19] Balaji Krishnapuram, Alexander Hartemink, Lawrence Carin “Applying logistic regression and RVM to achieve accurate probabilistic cancer diagnosis from gene expression profiles” - citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.5909
 - [20] A. Bharathi and A.M. Natarajan “Cancer Classification using Support Vector Machines and Relevance Vector Machine based on Analysis of Variance Features”- Journal of Computer Science 7 (9): 1393-1399, 2011.
 - [21] Fatima Eddaoudi , Fakhita Regragui , Abdelhak Mahmoudi , Najib Lamouri “Masses Detection Using SVM Classifier Based on Textures Analysis”- Applied Mathematical Sciences, Vol. 5, 2011, no. 8, 367 – 379.
 - [22] Guoqiang Peter Zhang “Neural Networks for Classification: A Survey” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 30, NO. 4, NOVEMBER 2000.
 - [23] S. Aruna, Dr .S.P. Rajagopalan “ A Novel SVM based CSSFFS Feature Selection Algorithm for Detecting Breast Cancer” International Journal of Computer Applications (0975 – 8887) Volume 31– No.8, October 2011.
 - [24] Xin Yao, Yong Liu “Neural Networks for Breast Cancer Diagnosis” 01999 IEEE
 - [25] Dimitrios Siganos ”Neural Networks in medicine”, www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/ds12/article2.html.
 - [26] David B.Fogel, Eugene C, Wasson, Edward M.Boughton “Evolving neural networks for detecting breast cancer”. 1995 Elsevier Science Ireland Ltd.
 - [27] Dr.Santhosh baboo, S.Sasikala “A Survey on data mining techniques in gene selection and cancer classification”-April 2010 International journal of Computer science and information technology.
 - [28] Afzan Adam1 Khairuddin Omar2 “Computerized Breast Cancer Diagnosis with Genetic Algorithms and Neural Network”- fitt.mmu.edu.my/caiic/papers/afzaniCAIET.pdf.
 - [29] Zhang Qinli; Wang Shitong; Guo Qi; “A Novel SVM and Its Application to Breast Cancer Diagnosis” http://ieeexplore.ieee.org/xpl/freabs_all.jsp?arnumber=4272649.