

An Application of Classification Techniques on Breast Cancer Prognosis

Sandeep Chaurasia
Assistant Professor
Department of Computer
Science
Sir Padampat Singhania
University
India

Prasun Chakrabarti
Head, Associate Professor
Department of Computer
Science
Sir Padampat Singhania
University
India

Neha Chourasia
MTech Student,
Mewar University,
India

ABSTRACT

Breast cancer is the most frequently diagnosed cancer in USA; furthermore breast cancer is the second most frequent cause of death for women in the United States as well as in Asia. In USA 40,600 deaths from breast cancer in 2009, 400 were men.[1] Several well established tools are currently used to screen for breast cancer including clinical breast exams, mammograms, and ultrasound. Supervised training is a technique in which a set of representative input output pairs is presented to the network. Through an iterative algorithm, the interval network weights are adjusted to decrease the difference between the network prediction and the true result for the training cases. The test has been performed on the breast cancer dataset using three classification techniques: Bayes learner, Decision Tree and Neural Net. The experiment concludes that Neural Net performance is better than the Decision Tree classification and Naïve Bayes classification for early detection of breast cancer with better accuracy and precision.

Keywords

Breast cancer, Mammography, Supervised Learning, Neural Network, Naïve Bayes, Decision Tree

1. INTRODUCTION

In India between 2001-03 about 8% of the female breast cancer cases in Delhi were in the age group 20-34. Breast cancers have also been reported in younger age group of 23-24 years from many parts in India. But a challenge among younger women is detection & diagnosis of malignant lumps that aren't easy due to high density in the breast. Efforts have been made to improve the accuracy of breast cancer diagnosis using different image mortality. [2] Mammography is the gold standard for finding the breast cancer, though as a screening tool but the sensitivity and specificity of it is bit low. Alternative such as Ultrasound and CBE (clinical breast examination) are supporting each other in the breast screening process, particularly for women with mammographically dense breasts. Another standard thermography as screening tool has been approved for use by the FDA in 1982 and was initially well accepted. No single screening tool provides excellent predictability but a combination of tools that also includes thermography has provided some acceleration to both sensitivity and specificity.

2. BACKGROUND MATERIAL

Mammography: Since 1960s, the gold standard for early detection of breast cancer has been, and still is mammography. It uses low-dose amplitude-X-rays to examine

the human breast. Cancerous masses and calcium deposits appear brighter on the mammogram. Mammography has helped to decrease the mortality rate by 25%-30% in screened women when compared with a control group after 5 to 7 years [3]. Randomized trials of mammographic screening have provided strong evidence that early diagnosis and treatment of breast cancer reduces breast cancer mortality [4]. The sensitivity of mammography in the general population is believed to reside between 75% to 90% with a positive predictive value of only 25%.

Limitation of mammography: Mammography is not well suited for women with dense breasts, implants, fibrocystic breasts, or on hormone replacement therapy. As the density of a woman's breast tissue increased the mammography's ability to detect abnormalities was reduced. With mammography there is a risk of rupture of the encapsulation of a cancerous tumor, as the process of taking a mammogram involves the compression of the breast tissue. Twenty-two pounds of pressure is sufficient to rupture the encapsulation around a cancerous tumor. Today's mammogram equipment uses 42 pounds of pressure. Depending on the location of the tumor, this would be sufficient force to rupture the encapsulation and potentially release malignant cells into the bloodstream. Mammography also confers a slightly increased risk of causing radiation induced breast cancer. Younger women's breast tissue is more susceptible to the effects of radiation versus older women because undifferentiated cells are more vulnerable to the effects of ionizing radiation.

Using mammographic screening it is very difficult to detect cancer in the early stage. In abnormal breast patients the findings were screened using mammography, sonography and magnetic resonance (MR) mammography [5]. Carcinoma in situ was diagnosed in 78.9% and 68.4% of patients using mammography and MR mammography, respectively. A study by Kelly et al [6] showed that 87% of cancer detections added by AWBU were found in the 68% of studies in women with very dense breasts. Kopans[7] has suggested that sonography should always be used with mammography or other imaging techniques for providing better result

Ultrasound: Ultrasound is an adjunctive tool used in conjunction with mammography and clinical breast exam in screening for breast cancer. Breast ultrasound has been considered a useful tool in mammographically dense breasts and in characterizing an abnormality detected in mammograms. It was found that sensitivity of mammography declines with decreasing tumor size and increasing breast density, while ultrasound remained effective regardless of tumor size. The overall accuracy of ultrasound has been found

to depend on three factors: quality of the tools, expertise of the physician in conducting the procedure and in interpreting the image, and the use of a multidisciplinary approach for breast cancer detection

Clinical Breast Exam: CBE and SBE are manual exams that are performed by the clinician or the patients themselves. Well performed SBE and CBE have been found to detect at least 50% of asymptomatic cancers. The sensitivity of CBE alone (21%) is not comparable to the sensitivity of mammography (78%), the two combined tend to improve the sensitivity of breast cancer detection (82%).

Thermography: Breast thermograms are widely used for the accurate detection of breast cancer [8]. Thermography is a promising screening tool because it has the ability to diagnose breast cancer at least ten years in advance. Thermography uses a digital infrared thermal imaging to detect and record the infrared heat radiating from the surface of the body. Cluster of abnormal cells that can develop into a cancer often have an increased blood supply that leads to an elevation in temperature of the skin over the area. Breast thermography has the ability to warn women years before any other procedure that inflammation is present that could later become cancer. The accuracy of thermograms is depended on many factors such as symmetry of the breast temperature and temperature stability

3. METHODOLOGY

3.1 Neural Network

The term neural network is applied to a parallel structure of simple computation units arranged in layer mimicking the physiologic structure of the brain. There are multiple connections between units within and between layers. These connections have strengths or weights that are “learned” by the network. The training paradigm is either supervised where sample in-output pairs are presented or unsupervised when the network organized itself. The network was trained using a back propagation supervised training algorithm. Supervised training is a technique in which a set of representative input output pairs is presented to the network. Through an iterative algorithm, the interval network weights are adjusted to decrease the difference between the network prediction and the true result for the training cases. The ANN display significant better diagnostic performance than the radiologist when the network output is compared to the radiologist categorical assessment. [9]

3.2 Supervised Learning

The majority of the networks require training in a supervised and unsupervised learning node [10]. We broadly classify two different types of learning supervised & un-supervised learning.

In supervised learning we assumed we assured at every instance of applying inputs, the desired response d of the system is provided by the teacher, the distance between actual and desired response serve as a error message and is used to connect network parameters exactly. In learning classification of input patterns or situation with known response, the error can be used to modify weights so that the error decreases. The general algorithms for learning different neural network have their own learning rule. Though all methods of learning have the same general algorithm, this algorithm mainly change the networks parameter according to its learning rule to accommodate the network characteristic to desired pattern. In general for the neuron L and its input J the weight vector $w_i =$

$[w_1, w_2, w_3, \dots, w_n]^T$ increases in proportion to the product of input x and learning signal r .

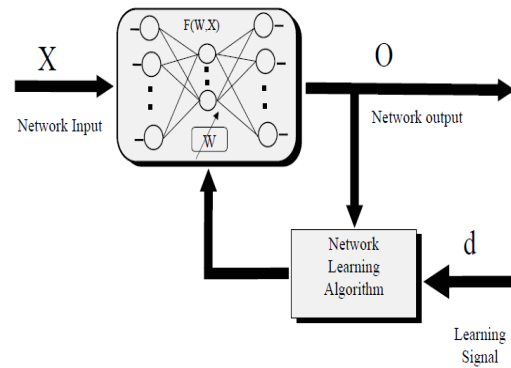


Figure 1 Supervised Learning in Neural Network

The learning signal R is in general a function of W_i , X_i and sometimes of the teacher’s signal D_i .

$$r = f(w_i, x, d)$$

The increment of the weight vector W_i produced by the learning step at the time t according to the general rule is

$$\Delta w_i(t) = cr[w_i(t), x(t), d_i(t)]x(t)$$

Where C is a positive number called the learning constant that determines the rate of learning. The weight vector adapted at time t becomes at the next instant, or learning step,

$$w_i(t+1) = w_i(t) + cr[w_i(t), x(t), d_i(t)]x(t)$$

The superscript convection will be used in this text to index the discrete-time training steps as in previous equation. For the K th step in above equation using this convection

$$w_i^{k+1} = w_i^k + cr(w_i^k, x^k, d^k)x^k$$

From the above two equation the learning can be expressed as

$$dw_i(t) / dt = crx(t)$$

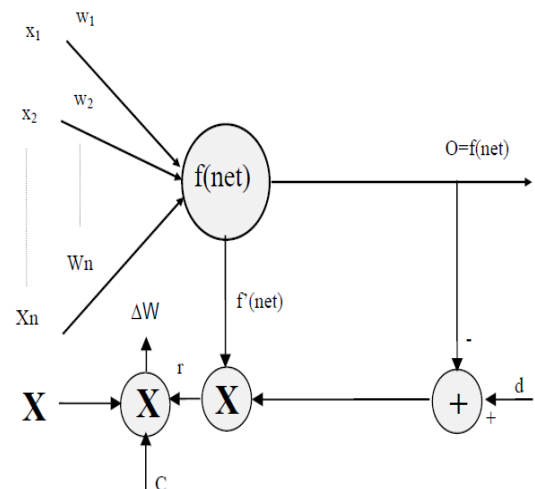


Figure 2 Neural Network learning algorithm

Table 1: Attributes of the sample dataset

Attribute	Domain
1. Sample code number	id number
2. Clump Thickness	1 – 10
3. Uniformity of Cell Size	1 – 10
4. Uniformity of Cell Shape	1 – 10
5. Marginal Adhesion	1 – 10
6. Single Epithelial Cell Size	1 – 10
7. Bare Nuclei	1 – 10
8. Bland Chromatin	1 – 10
9. Normal Nucleoli	1 – 10
10. Mitoses	1 – 10
11. Class:	2 - benign, 4- malignant

Attributes 2 through 10 have been used to represent instances.

3.3 Experiment

3.3.1 The Breast Cancer Dataset

The Breast Cancer Wisconsin Dataset available by anonymous ftp from ice.uci.edu [11]. The brief detail of the it as follows: the dataset contains 699 patterns with 9 attributes as Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses. The class output includes 2 classes, benign and malignant. (Class 2 i.e. Benign 444 and Class 4 i.e. Malignant 239 with 16 missing attribute) The original dataset was obtained by Wolberg and Mangasarian. [12]

3.3.2 The experimental setup

The original data is present in the form of analogue values with values ranging from 0-10. The data are converted to their equivalent digital form. Scaling is required to map the dataset into desired range of variable ranging between minimum and maximum range of network input. Based on total number of attribute (assume N). N-1 will be numeric feature and 1 is class category. The numerical attributes are ranging in between 0 and 1.

First from the given range of inputs, the minimum and the maximum value is picked up and this scaling is done by the following formula.

$$\text{New value (after scaling)} = (\text{current value} - \text{min value}) / (\text{max value} - \text{min value}).$$

The new values obtained are converted into binary form by the following scaling, grouping is done on the basis of range $[0, x) = '0'$ and $[x, 10] = '1'$. These attributes are fed into the various operators for training and testing to obtain the result 10 cross fold validation is applied in the validation model to calculate the performance of the three operators. We have used a open source platform named *RapidMiner 5.2* that supports Meta Learning by embedding one or several basic learners as children into a parent meta learning operator. The learning operator we have used here are the (i) Naïve Bayes operator (ii) Decision Tree (iii) Neural Net operator respectively on the breast cancer dataset.

3.3.3 Result

Decision tree: This operator learns decision trees from both nominal and numerical data. Decision trees are powerful classification methods which often can also easily be understood. In order to classify an example, the tree is traversed bottom-down. Every node in a decision tree is labeled with an attribute.

Table 2: Decision Tree Performance Vector

Performance Vector:
Accuracy: 94.42% +/- 2.89% (mikro: 94.42%)
Precision: 90.42% +/- 6.34% (Positive class: 4.0)

Table 3: Classification of data by Decision Tree

Output	True (2.0)	True (4.0)	Class Prediction
Benign	433	14	96.87%
Malignant	25	227	90.08%
Class Recall	94.54%	94.19%	

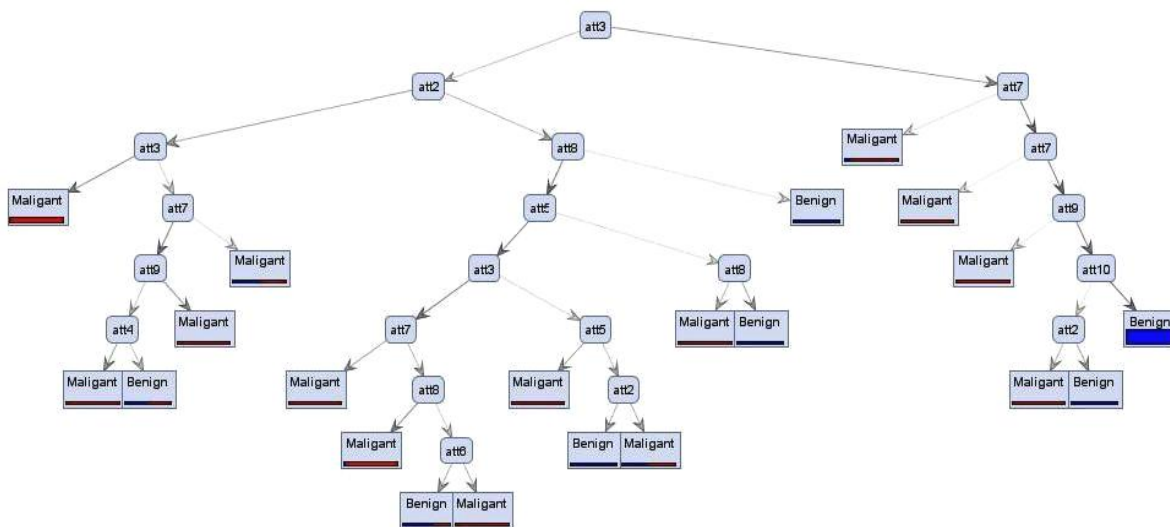


Figure 3 Decision Tree obtained with the performance matrix

Neural Network: This operator learns a model by means of a feed-forward neural network trained by a back propagation algorithm (multi-layer perception). The user can define the structure of the neural network with the parameter list "hidden layers". Each list entry describes a new hidden layer.

Table 4: Neural Net: Performance Vector
Performance Vector:
Accuracy: 94.28% +/- 3.50%
Precision: 94.64% +/- 4.65% (Positive class: 4.0)
Recall: 88.36% +/- 6.92% (Positive class: 4.0)

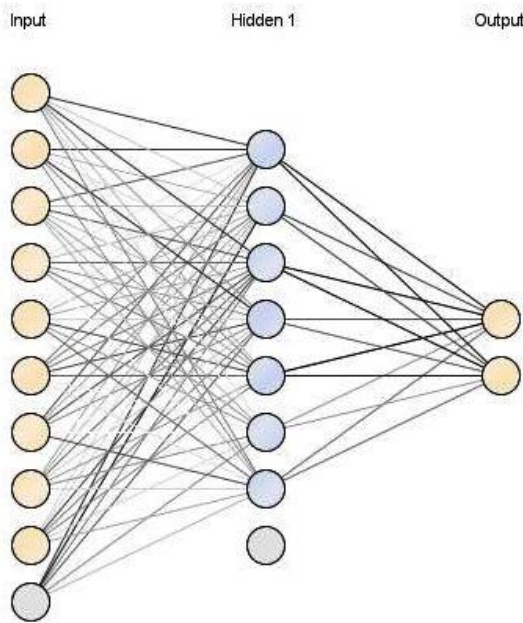


Figure 4 Neural Network with 2 output Layer Layer- (i) Benign and (ii) Malignant

Output	True 2.0	True 4.0	Class Prediction
Benign	446	28	94.09%
Malignant	12	213	94.67%
Class Recall	97.38%	88.38%	

Naive Bayes: Naive-Bayes Classification Algorithm represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems. Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data.

Table 6: Naive Bayes: Performance Matrices
Performance Vector
Accuracy: 96.14% +/- 1.92%
Precision: 91.95% +/- 4.46% (Positive class: 4.0)
Recall: 97.52% +/- 2.03% (Positive class: 4.0)

Table 7: Classification of data by Naïve Bayes

Output	True 2.0	True 4.0	Class Prediction
Benign	437	6	98.65%
Malignant	21	235	91.80%
Class Recall	95.41%	97.51%	

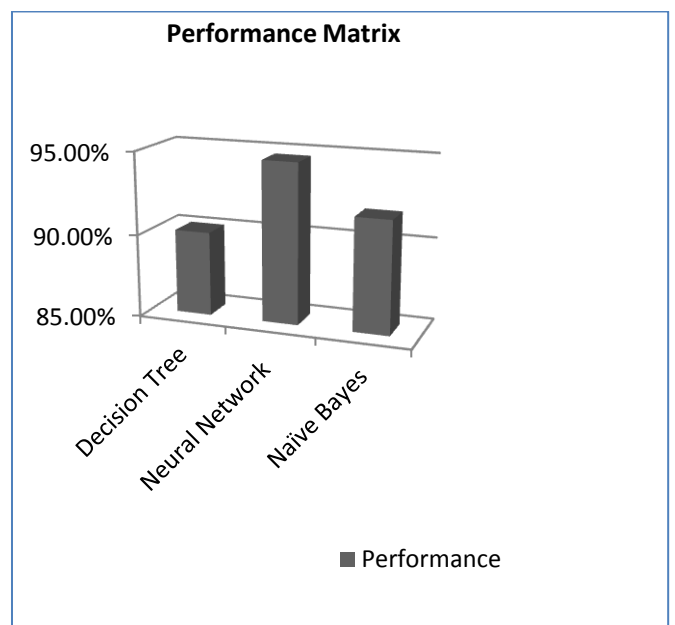
4. CONCLUSION

This paper describes the three well known techniques in relative detail. After a better understating of the strengths of each method it has been observed that the results are generated on the basis of accuracy, precision and recall. The overall prediction class for benign class is highest in Naïve Bayes classification with 98.65% with 437 true and 6 false predictions in contrast with Neural Network having 94.09% class prediction with 446 true and 28 false predictions and for Decision tree with 96.87% with 433 true and 14 false predictions.

The prediction class for malignant class is highest in Neural Network classification with 94.67% with 12 true and 213 false predictions as compare to Naïve Bayes with 91.8% class precision with 21 true and 235 false predictions and in Decision Tree with 90.08% precision (25 true and 227 false predictions).

So the efficiency of Neural Network is highest to predict the malignancy in breast cancer with +2.87% more than Bayesian classification and +4.59% more than Decision Tree.

Table 8. Performance for Data Classification



5. ACKNOWLEDGMENT

This breast cancer domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. We acknowledge to M. Zwitter and M. Soklic for providing the data. We would also like to show our appreciation to the Professor P. C Deka, Vice Chancellor, Sir Padampat Singhanian University, Udaipur for their invaluable comments and suggestion.

6. REFERENCES

- [1] Dursun Delen, Gelnn Walker, Amit Kadam “Predictive breast cancer survivability: a comparison of three data mining methods” *Artificial Intelligence in Medicine* (2004) ELSEVIER.
- [2] An article in *TIMES OF INDIA* “detecting breast cancer in the young is difficult and that’s bad news for India’s under-35 patient”.
- [3] Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. *JAMA*. 1995;273:149–154.[PubMed]
- [4] Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet*. 2002; 359:909–919.[PubMed]
- [5] Malur S, Wurdinger S, Moritz A, Michels W, Schneider A. Comparison of written reports of mammography, sonography and magnetic resonance mammography for preoperative evaluation of breast lesions, with special emphasis on magnetic resonance mammography. *Breast Cancer Res*. 2001;3:55–60. [PMC free article][PubMed]
- [6] Kelly KM, Dean J, Comulada WS, Lee SJ. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *Eur Radiol*. 2010;20:734–742. [PMC free article][PubMed]
- [7] Kopans DB. Sonography should not be used for breast cancer screening until its efficacy has been proven scientifically. *AJR Am J Roentgenol*. 2004;182:489–491.[PubMed]
- [8] J.R. Keyserlingk, P.D. Ahlgren, E. Yu, N. Belliveau, Infrared imaging of breast: Initial reappraisal using high-resolution digital technology in 100 successive cases of stage I and II breast cancer, *The Breast Journal* 4 (4) (1998) 241–251.
- [9] Online Article “Computerized detection of breast cancer with artificial intelligence and thermograms”. www.ncbi.nlm.nih.gov/pubmed/12396330
- [10] Online Article “The International Congress for global Science and Technology”. www.icgst.com
- [11] C.L Blake and C.J Merz. UCI repository of machine learning databases. <http://www.ics.uci.edu/~mllearn/> University of California, Irvine, Department of Information and Computer Science.
- [12] W.H. Wolberg and O.L. Mangasarian. “Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology”. *Proceedings of the National Academy of Sciences, U.S.A.*, Vol. 87, 1990, pages 9193-9196.