Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 3, pp. 455–465
DOI : 10.47974/JIOS-1365

Enhancing the accuracy of breast cancer detection and determination of risk factor by using the backpropagation network theory and SVM: Machine learning

N. Bindu Madhavi *
KL Business School
Koneru Lakshmaiah Education Foundation (Deemed to be University)
Vaddeswaram
Andhra Pradesh
India

Sushil Dohare †
Department of Epidemiology
College of Public Health and Tropical Medicine
Jazan University
Saudi Arabia

G. N. R. Prasad §
Department of MCA
Chaitanya Bharathi Institute of Technology (A)
Gandipet
Hyderabad 500075
India

D. Vijendra Babu ‡
Vellore Institute of Technology
School of Electronics Engineering
Vellore 632014
Tamil Nadu
India

* *E-mail:* dr.bindumadhavi@kluniversity.in (Corresponding Author)
† *E-mail:* sdohare@jazanu.edu.sa
§ *E-mail:* gnrp@cbit.ac.in
‡ *E-mail:* drdvijendrababu@gmail.com

Abdul Rahman Mohammed Al-Ansari [@]
Department of Surgery
Salmanyia Hospital
Bahrian

Abstract

According to the world health organization, every year, more than 8% of women suffer due to breast cancer, and 40% of women die in low-poverty regions. This entire work focuses on the algorithm to detect breast cancer. This algorithm improves the accuracy of the detection and the risk factor determination by using the backpropagation network (BPN) theory and the Support vector method (SVM). By the end of the entire work, the improved accuracy is up to 95% compared to other forms; this proposed method is proper when evaluating the patient report in the image format, like a scanning report.

Subject Classification: *Primary 93A30, Secondary 49K15.*

Keywords: *Breast cancer, Backpropagation network, Support vector method.*

1. Introduction

Cancer of the breast is a kind of cancer that typically develops in females. Both breasts and a single breast are possible sites of onset. Unchecked cell growth is the first stage in the development of cancer. Breast cancer is more common in women, but males are not immune to the disease. Remember that most breast lumps are entirely benign and not cancerous (malignant). The abnormal growths known as non-cancerous breast tumours do not pose any danger to the body beyond the breast. Although gentle breast lumps seldom cause death, they increase women's risk of breast cancer. If you notice a lump or change in your breast, a doctor must check it out to determine whether or not it's cancerous. Preventing mortality from breast cancer requires a combination of measures, the most essential of which are early detection and access to cutting-edge cancer therapy. Early detection is critical to properly treat breast cancer, especially when the tumour is still relatively small and has not spread. The best approach to spot breast cancer early is to get screening done routinely.

Breast mammograms use low-dose x-rays. Regular mammograms are more effective in treating breast cancer if it is detected at an earlier stage. Mammography often detects abnormalities in the breast that might be cancer years before the onset of any noticeable symptoms. Several

[@] *E-mail:* **`Haman.alansari@gmail.com`**

decades' worth of studies demonstrates that women who regularly have mammograms have a higher chance of having their breast cancer detected at an earlier stage, requiring less invasive treatments like mastectomy and chemotherapy, and have a higher chance of being cured.

2. Existing work

According to the WHO, around the world, 10% of women die from cancer. Breast cancer is one of the significant issues for women's health. Breast cancer can detect by the existing clinical data, different medical examinations and ultrasonic results; ultrasonic inspection will give the best effect in the image format. The available images of breast cancer are one of our current study's sources.

Ultimately, the last work on breast cancer uses machine learning techniques. Different datasets used are coimbra datasets and radial functional datasets in previous work. The listed features are the BMI of the patient, age group, current medicine, the glucose level in the body, and insulin used. Specificity and sensitivity values for existing methods are 0.9 and 0.92, respectively.

3. Methodology

This method uses backpropagation networking and Artificial intelligence to detect the risk factor of breast cancer. The reports of the patients affected by breast cancer are used as input data; input data should be in the form of mammograms. This method involves two stages. Stage 1 uses the BPN and SVN methods to evaluate/analyze the datasets. After assessing the datasets, classify the risk factor of breast cancer. And stage 2, Two different models (Alex next and Resnet) are used to detect the mammograms and optimized by using the stochastic gradient descent with momentum.

3.1 *Proposed frame work:*

From figure 1, consider the patient data as input for the proposed work. This data needs to focus on the risk factor. By using this data classify the status of the patient. Suppose a mammogram is like a tumor. Then the doctor will suggest the treatment according to the patient's condition.

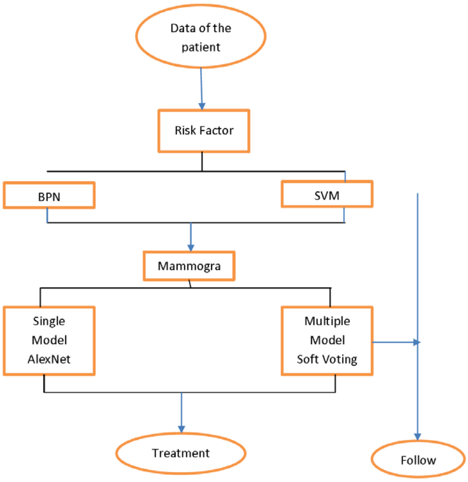

Figure 1

Proposed frame work

Classification:

Two methods are discussed below in the proposed work to classify breast cancer.

4. BPN process in step by step:

Step-1: Framework of the network system.

Input layer: Enter the data, and classify the data according to different causes of breast cancer.

Output layer: It signifies the productivity of the network. The result depends on the neural network depends on the variable problems.

Hidden layer: It links the input and the output layers. It works on several neurons that are activated.

Number of hidden layers= (Number of neurons in (Input layer X output layer)) 0.5 = (12X1)0.5 = 3.5

Number of active neurons= 0.5 (Number of neurons in (Input layer X output layer)) = 0.5(12+1) = 6.5

Activation layer = $\frac{1}{1+e^{-x}}$ It converts the output between 0 to 1.

Step-2: In this step need to control the time in each iteration.

A) Find the suboptimal value

B) Reduce the processing time like jam time/ delay time.

$$\text{Delay time} = \frac{\text{rate of initial learning}}{1+(\text{delay rate } X \text{ current iteration number})}$$

Predict the rate of accuracy by Mean square error(MSE)

$$\text{Rate of accuracy} = \frac{\text{sample classified}}{\text{total number of classified}} \times 100$$

The SVM process can perfectly optimize the decision using the data sets. The SVM process quickly deals with the non-linear model using the kernel function. Using the kernel function convert the non-linear to linear and used to perform the classifier.

5. Process of image reorganization:

For the image recognition process, use the CNN model to predict breast cancer in the patients. In the CNN process, two different types of layers are developed.

Convent layer: It is one of the primary layers. It extracts the function into the input image format (in a mammogram). In this, layer-by-layer features will develop gradually. The filter will apply to the size of the image. It reduces the size of the image by fixing the fixed value. By reducing the size, the number of parameters is also reduced as the complexity of the problem will be reduced.

Connected layer: When input starts from the first point in the convent layer, the neurons will pass through all the elements still up to future value. The future value is activated, and overall, neurons will work like a common neuron.

5.1 *Stochastic Gradient descent*

Two optimization techniques are used to train the datasets and compare the performance. Stochastic Gradient Descent with Momentum (SGDM) and Adaptive moment estimation (AME).

5.2 *Database related to breast cancer*

For the assessment of this algorithm used, nearby 2.5 lacks (2,42,928) cases with different parameters. 1. Age of the patient 2. BMI 3. Relativity 4. The density of the breast 5. Origin of breast cancer 6. Mammogram of the patient 7. Existed surgery related to this case 8. Variable response.

After applying these parameters as a filter for the total processing cases, 89822 cases are filtered as learned datasets and accordingly classified.

5.3 *Model development*

MATLAB develops the entire model. MATLAB must fix the image size and number of layers when learning datasets transfer from one stage to another.

Backpropagation Network:

The various neuron networks in hidden layers are set to 2,3,4,5,6 and 7, tested. Tabulated the tested result three times for each neuron and average the result. (Table-1).

Table 1

BPN Process Hidden layers

Hidden Layer Neuron Count	1st Result	2nd Result	3rd Result	Mean
2	75.6%	75.1%	72.1%	74.3%
3	76.2%	76.2%	76.1%	76.2%
4	72.1%	75.6%	75.2%	74.3%
5	75.3%	73.5%	74.6%	74.5%
6	73.2%	74.5%	73.8%	73.8%
7	74.9%	76.7%	74.6%	75.4%

When the risk factor of breast cancer enters the dataset to obtain the training datasets with the hidden layers, found 30221 cases of breast cancer, and the remaining 10500 are not having cancer. Among these, 13392 are correctly classified without cancer 34108 are classified with cancer.

Table 2

Accuracy and sensitivity of breast cancer in BPN and SVN

Prediction Category	Open cancer Category				
	Tumor	Tumor-Free	Tumor + Tumor-Free	Accuracy	Sensitivity
BPN	30221	10500	40721	74.21%	25.78%
	13392	34108	47500	28.19%	71.80%
SVM	40689	32	40721	93%	0.11%
	8231	39269	47500	7.9%	82.67%

Table 2 represents the accuracy and sensitivity of breast cancer in BPN and SVN. BPN sensitivity is 25.78% with cancer and 71.80% while cancer-free. The SVM method is 93% accurate with cancer.

6. Result

6.1 *Alex Net optimization*

It optimised the SGDM and AME to classify and obtain the learned datasets.

From the optimization techniques considered AME (Adaptive Moment Estimation) and SGDM.

From AME tested the accuracy for the datasets in three levels 77.22%, 83.12% and 82.12% Accuracy is obtained respectively. Average accuracy is 80.8%.

Similarly from SGDM accuracy results are 82.22%, 86.22% and 84.5%. Average accuracy is 84.5%.

By comparing AME and SDGM optimization tools the accuracy is good in SGDM which is represented in Table 3.

Table 3

AME and SDGM – AlexNet accuracy classification

Model	1st Accuracy	2nd Accuracy	3rd Accuracy	Mean
AME	77.22%	83.12%	82.12%	80.8%
SGDM	r82.22%	86.22%	85.1%	84.5%

6.2 *ResNet101 optimization technique*

Table 4

Comparison of AME's Classification Accuracy across Layers

Fixed-Weight layer	Classification Accuracy 1	Classification Accuracy 2	Classification Accuracy 3	Mean
Prior to fixing module 5c (1-333)	79.2	76.7	80.1	78.67
Prior to fixing module 5b (1-323)	81.3	82.2	80.23	81.24
Prior to fixing module 5a (1-311)	78.2	80.1	85.26	81.19

Table 5

Comparison of SGDM Classification Accuracy across Layers

Fixed-Weight layer	Classification Accuracy 1	Classification Accuracy 2	Classification Accuracy 3	Mean
Prior to fixing module 5c (1-333)	77.61	80.12	76.24	77.99
Prior to fixing module 5b (1-323)	80.72	82.23	80.22	81.06
Prior to fixing module 5a (1-311)	80.12	76.4	77.12	77.88

From the recorded experimentation results in Table 4 & 5 analyzed AME and SGDM optimizes by before fixing each module layers.

5a indicates 1-311 weight of the layer obtained by AME levels the accuracy levels are 78.2%, 80.1% and 85.26%. Similarly, from SGDM accuracy results are 80.12%, 76.4% and 77.12%.

5b indicates 1-323 weight of the layer obtained by AME levels the accuracy levels are 81.3%, 82.2% and 80.23%. Similarly, from SGDM accuracy results are 80.72%, 82.23% and 80.22%.

5c indicates 1-333 weight of the layer obtained by AME levels the accuracy levels are 77.61%, 80.12% and 76.24%. Similarly, from SGDM accuracy results are 77.61%, 80.12% and 76.24%.

6.3 *Soft voting process*

Table 6

Voting analysis

Sample category	Beginners or normal	Medium effected
Beginners or normal	32	3
Medium effected	2	36

From soft voting, Table 6 shows the result of soft voting and majority voting. Decreases the cases of misclassifications—classification rate is 94.11%.

7. Result

This entire work represents the improvement of accuracy in diagnosing breast cancer. The average accuracy achieved through AME

& SGDM methods is 80.8% and 84.5%, respectively. Alexnet reached the maximum accuracy while used in the ML-supervised algorithm process. Removing any conventional layer in the neural network leads to degrading performance. Analysing SVN, ANN, and SVN processes will avoid overfitting the data. The SVM process reaches an accuracy of up to 93%. Soft voting is the best technique to prevent misclassification while using the data. In real-time, we faced a slight proportion variation while using limited sources. Finally the end of the research, work is fine-tuned the network.

8. Conclusion

This research develops the step-by-step process to identify or predict breast cancer and improve the chances of identifying breast cancer. In the initial stage, find the breast cancer risk factor by using the existing dataset. Next, develop a model to determine the monogram using Alexnet, ResNet 101. These models studied the AME and SGDM optimizers. Finally, the developed process identifies breast cancer with more accuracy. From the entire work, it was observed that this model could classify the tumour stage, i.e. whether at the beginning or normal. Finally, soft voting classifies cases by using the developed process.

References

[1] Lin, R.-H.; Kujabi, B.K.; Chuang, C.-L.; Lin, C.-S.; Chiu, C.-J. Application of Deep Learning to Construct Breast Cancer Diagnosis Model. *Appl. Sci.* 2022, 12, 1957. https://doi.org/10.3390/app12041957.

[2] Luuk Balkenende, Jonas Teuwen, Ritse M. Mann, Application of Deep Learning in Breast Cancer Imaging, Seminars in Nuclear Medicine, Volume 52, Issue 5, Pages 584-596 (2022), ISSN 0001-2998, https://doi.org/10.1053/j.semnuclmed.2022.02.003.

[3] Shen, L., Margolies, L.R., Rothstein, J.H. *et. al.* Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Sci Rep* 9, 12495 (2019). https://doi.org/10.1038/s41598-019-48995-4.

[4] Mridha MF, Hamid MA, Monowar MM, Keya AJ, Ohi AQ, Islam MR, Kim JM. A Comprehensive Survey on Deep-Learning-Based Breast Cancer Diagnosis. Cancers (Basel). 13(23):6116 (2021 Dec 4). doi: 10.3390/cancers13236116. PMID: 34885225; PMCID: PMC8656730.

[5] Arooj Sahar, Atta-ur-Rahman , Zubair Muhammad, Khan Muhammad Farhan, Alissa Khalid, Khan Muhammad Adnan, Mosavi Amir, Breast Cancer Detection and Classification Empowered With Transfer Learning, Frontiers in Public Health, Volume 10, (2022) ISSN- 2296-2565, 10.3389/fpubh.2022.924432, URL=https://www.frontiersin.org/articles/10.3389/fpubh.2022.924432.

[6] BramahHazela, J. Hymavathi, T. Rajasanthosh Kumar, S. Kavitha, D. Deepa, Sachin Lalar, Prabakaran Karunakaran, "Machine Learning: Supervised Algorithms to Determine the Defect in High-Precision Foundry Operation", *Journal of Nanomaterials*, vol. 2022, Article ID 1732441, 9 pages (2022). https://doi.org/10.1155/2022/1732441.

[7] Dewangan, K.K., Dewangan, D.K., Sahu, S.P. *et. al.* Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique. *Multimed Tools Appl* 81, 13935–13960 (2022). https://doi.org/10.1007/s11042-022-12385-2.

[8] Saad Awadh Alanazi, M. M. Kamruzzaman, Md Nazirul Islam Sarker, Madallah Alruwaili, Yousef Alhwaiti, Nasser Alshammari, Muhammad Hameed Siddiqi, "Boosting Breast Cancer Detection Using Convolutional Neural Network", *Journal of Healthcare Engineering*, vol. 2021, Article ID 5528622, 11 pages (2021). https://doi.org/10.1155/2021/5528622.

[9] Yue, Wenbin & Wang, Zidong & Chen, Hongwei & Payne, Annette & Liu, Xiaohui. Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis. Designs. 2. 13 (2018). 10.3390/designs2020013.

[10] Hamed, Ghada & Marey, Mohammed & Amin, Safaa & Tolba, Mohamed. Deep Learning in Breast Cancer Detection and Classification (2020). 10.1007/978-3-030-44289-7_30.

[11] Buda M, Saha A, Walsh R, et. al. A Data Set and Deep Learning Algorithm for the Detection of Masses and Architectural Distortions in Digital Breast Tomosynthesis Images. *JAMA Netw Open.* 4(8) : e2119100 (2021). doi:10.1001/jamanetworkopen.2021.19100.

[12] Yassine Laarichi, Mohammed Barmaki. Explicit formulas for computing matrix trigonometric functions. *Journal of Interdisciplinary Mathematics* 25:8, pages 2321-2331 (2022).

[13] Malihe Niksirat. Intuitionistic Fuzzy Hub Location Problems: Model and Solution Approach. *Fuzzy Information and Engineering* 14:1, pages 74-83 (2022).

[14] Neetika Bhandari, Payal Pahwa. Evaluating Partitioning Based Clustering Methods for Extended Non-negative Matrix Factorization (NMF). *Intelligent Automation & Soft Computing* 35 : 2, pages 2043-2055 (2023).

[15] Joseph Thomas Eghwerido, Pelumi E. Oguntunde, Friday Ikechukwu Agu. The Alpha Power Marshall-Olkin-G Distribution: Properties, and Applications. *Sankhya A* 85:1, pages 172-197 (2023).

[16] Abdalla Rabie, Eslam Hussam, Abdisalam Hassan Muse, Ramy Abdelhamid Aldallal, Amirah Saeed Alharthi, Hassan M. Aljohani, Melike Kaplan. Estimations in a Constant-Stress Partially Accelerated Life Test for Generalized Rayleigh Distribution under Type-II Hybrid Censoring Scheme. *Journal of Mathematics*, pages 1-9 (2022).