Received: August 11, 2023. Revised: September 13, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Early Detection and Classification of Breast Cancer Using Weight based AdaBoost Algorithm

Asha Kethaganahalli Hanumanthaiah[1] Manju More Eshwar Rao[2] Soumyashree Muralidhar Panchal[3]*

[1]*Department of Information Science and Engineering, Global Academy of Technology, Bengaluru, India*
[2]*Department of Computer Science and Engineering, PES University, Bengaluru, India*
[3]*Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India*
* Corresponding author's Email: soumya.shree@manipal.edu

Abstract: Breast cancer is a deadly disease; an accurate and early diagnosis of breast cancer is the most efficient method to decrease the death rate. But, in the early detection and diagnosis of breast cancer, differentiating abnormal tissues is a challenging task. In this paper, a weight-based AdaBoost algorithm is proposed for an effective detection and classification of the breast cancer. An AdaBoost algorithm effectively classifies the breast cancer classes by adding the weights to the samples in the week classifier during the training phase. A weighted vote is performed on the results of each week classifier, and the strong classifier is integrated according to the weight of the week classifier. The breast cancer image datasets named CBIS-DDSM and MIAS are utilized for effective classification. Tumor-like regions (TLRs) are diagnosed by utilizing the optimum method of Otsu thresholding to enhance training abilities. The convolutional neural network (CNN) architectures of the AlexNet and ResNet50 are utilized for the feature extraction. A weight-based AdaBoost algorithm is proposed for the classification of breast cancer mammogram images into four classes benign calcification (BC), malignant calcification (MC), benign mass (BM) as well and malignant mass (MM). The results shows that the proposed weight based AdaBoost algorithm delivers the performance metrics such as accuracy, specificity, sensitivity, precision and F1-score values about 99.56%, 99.38%, 99.40%, 98.89% and 99.18% respectively, which ensures the accurate classification results compared with the existing methods such as IMPA-ResNet50, Gray difference weight and MSER detector, MLO and CC methods.

Keywords: AlexNet, Breast cancer detection, Convolutional neural network, Deep learning, Otsu thresholding, Resnet50, Weight based AdaBoost.

1. Introduction

According to the world health organization (WHO), breast cancer is the most frequent cancer in women and the second most common cancer in the world [1]. Breast cancer can occur in both men and women, but the occurrence rate in men is 1 out of 100 [2]. An early diagnosis is important, because, the cancer spreads to affect the whole breast or other parts [3]. Breast cancer can be divided into two types according to the visualization of cells using a microscope. The two types are invasive ductal carcinoma (IDC) as well as ductal carcinoma in situ (DCIS) [4]. Screening is an efficient way of minimizing the rate of mortality from breast cancer. Various imaging techniques like mammography, elastography, and ultrasound imaging are utilized for the diagnosis of breast cancer [5]. The female population-based systematic screening with mammograms and early diagnosis of breast cancer can increase the patient's survival as well as minimize the essential treatment side effects [6].

Early detection of breast cancer is a significant method to maximize the chances of treatment and survival rate of the affected patients [7]. The breast cancer diagnosis majorly consists of pathological and imaging diagnosis. The diagnosis using image is a non-invasive diagnostic method compared to pathological diagnosis [8]. Medical images are important sources of useful information and various imaging techniques such as mammography, ultrasound (US) imaging, thermography, and elastography are being widely used for diagnosing breast cancer [9]. Thermography is a low-cost technique, safer, non-invasive as well as non-contact screening method that can detect cancer at an early stage, even in precancerous conditions [10]. The thermal images were normally interpreted by human experts. A breast X-ray image is captured by the mammography to identify and diagnose breast cancer disease [11]. Mammography is the most popular imaging technique for detecting breast cancer, but it has some limitations such as low sensitivity, over-ionizing radiations and is not suitable for dense breasts [12]. The computer-based programs that alert radiologists to variations of optimum in mammography as well as permit integrated film misleading are the potential points [13]. The proposed method's goal is early detection and classification of breast cancer using a weight based AdaBoost algorithm with the feature extraction using AlexNet and ResNet50. The advantages include enhanced feature extraction, improved classification accuracy, efficiency and flexibility. These advantages collectively contribute to the development of accurate and efficient classification. The major contributions of this proposed study are given as follows:

- The optimum Otsu thresholding method is utilized to identify the tumor-like-regions (TLRs) to enhance training abilities.
- The AlexNet and ResNet50 architectures of convolutional neural network (CNN) are utilized for the extraction of features to detect the features in the image of breast cancer detection.
- In classification, a weight based AdaBoost algorithm is used for the classification of breast cancer into four classes benign calcification (BC), malignant calcification (MC), benign mass (BM) and malignant mass (MM) by adding the weights to the samples in the week classifier during the training phase.

The rest of the manuscript is organized as follows: Section 2 illustrates the literature review. The proposed model is presented in section 3. The experimental result of the proposed model is illustrated in section 4. The summary is described in section 5.

2. Literature survey

Houssein [14] implemented a novel classification model for the diagnosis of breast cancer using marine predators' algorithm and an improved optimization algorithm called IMPA-ResNet50. This model utilized an opposition-based learning strategy to cope using the implied weakness of the original MPA for improving the model. This model utilized two datasets MIAS as well as CBIS-DDSM. This model improved the accuracy of results of breast cancer diagnosis. The IMPA-ResNet50 was only implemented to classify mammography images. Those results were limited to a specific dataset, MIAS dataset, and CBIS-DDSM dataset and may not be generalized to the other dataset.

Sanchez-Cauce [15] implemented a novel method for the early detection of breast cancer by combining various perspectives with clinical and personal data. This method utilized a multiple input classification, which utilizes the convolutional neural network (CNN) benefits for the analysis of the image. This method initially utilized the thermal images of personal and clinical data for an analysis of cancerous images. This method enhanced the classification performance as well as achieved the best accuracy by adding thermal images. The DMR database had small size (only 287 patients before data cleansing) makes it difficult to apply deep learning techniques.

Allugunti [16] implemented a method of computer-aided diagnosis (CAD) for breast cancer patient identification as well as diagnosis. This method utilized multiple classes such as cancer, no cancer and non-cancerous for the identification and diagnosis of patients under the management of a database. This method enhanced the model performance and achieved better accuracy results by pre-processing the data and by controlling the mammogram images in advance. However, in order to pre-process the images, the imaging method demands a significant amount of processing resources.

Divyashree and Kumar [17] performed background supervision and pectoral muscle removal utilizing a gray difference weight and MSER detector. A contrast-limited adaptive histogram equalization (CLAHE) and de-correlation stretch were performed to improve the detection of breast region. This model utilized two datasets MIAS as well as CBIS-DDSM. This model achieved better accuracy results in segmentation and detection of breast cancer. The CBIS-DDSM dataset was able to detect 192 images with exact locations as given in dataset, but failed to detect the eight mass images.

Alanazi [18] implemented boosting automatic

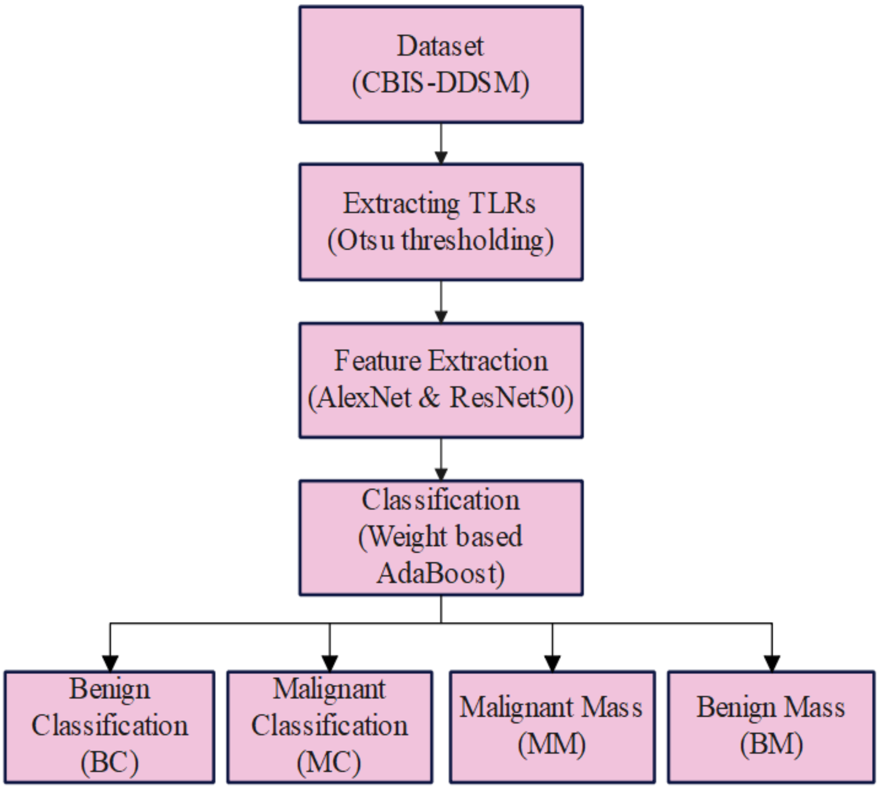

Figure. 1 Workflow of proposed methodology

breast cancer identification by determining the tissue zones of hostile ductile carcinoma in whole-side images. This implemented method was evaluated using the number of architectures of convolutional neural networks for automatic breast cancer detection by comparing the machine learning algorithms. By a combination of machine learning algorithms, this method achieved better accuracy, but this method had consumed a much time for obtaining better performance.

Salama and Aly [19] developed a new framework for breast cancer image segmentation and classification. This method utilized the cranio caudal (CC) vision and mediolateral oblique (MLO) technique for the identification and diagnosis of breast cancer. The transfer learning and data augmentation approach was utilized to solve the problem of tagged data in breast cancer. This method achieved the best result and less computational time but this method does not clearly classify all the images.

Tembhurne [20] implemented a computer-aided transfer learning with a deep learning model as a binary classifier for breast cancer detection. The sequential architecture of deep learning was followed on an individual channel for the extraction of features as well as classification. This method utilized the combining techniques of multiple channels for devising the architecture of a dual ensemble. This method achieved better accuracy results by utilizing multiple channels. But for more precise diagnosis it needs to implement the segmentation task.

The limitations found from the related work are the time complexity to train multiple classifiers, the small number of databases to apply, the proper screening images as well as utilization of the secondary Kaggle databases. These limitations caused the inappropriate results so to overcome this, a weight-based AdaBoost classification algorithm is proposed for breast cancer detection. To classify the breast cancer mammogram images into four classes such as BC, MC, BM and MM as well as earlier diagnosis of this disease.

3. Proposed methodology

The proposed weight based AdaBoost algorithm is utilized for breast cancer detection using CBIS-DDSM dataset. This section consists of four different stages which include data collection, extraction of TLRs, feature extraction and classification. Fig. 1 represents the workflow of the proposed breast cancer detection methodology.

3.1 Dataset description

The experiments of this proposed method utilize the breast cancer dataset named digital database for screening mammography (DDSM) and mammographic image analysis society (MIAS). The curated breast imaging subset (CBIS-DDSM) [21] is a standard and advanced form of the DDSM dataset that is utilized to take not only normal cases but also abnormal cases. This dataset consists of four classes such as BC, MC, BM and MM. This dataset consists 3549 region of interest as 1852 (1132 benign and 720 malignant) for calcification nodules and 1697 (913 benign and 784 malignant) for mass nodules. This type of image can be converted from the format of DICOM to the format of PNG with the size of. The MIAS [22] dataset contains 322 mammographic images with size of 1024×1024 in the format of Portable Grey Map. Further image size is reshaped into 256×256. This dataset has two classes: normal (209) and abnormal (113). The abnormal category was divided into two classes: benign and malignant of 65 and 48 images. The classifications, asymmetries and masses are the abnormalities types identified in the data. For the CBIS-DDSM and MIAS dataset, this process utilized the 5-fold cross-validation due to its increased size. This is divided into two portions such as 80% for training and 20% for validation. The provided results are evaluated as the average number of folds. The collected dataset is provided to the tumor-like regions for extraction.

3.2 Extraction of tumor-like-regions

After collecting the data, the Otsu thresholding technique is used to extract the TLRs. The extracted TLRs are utilized for low-cost feature extraction in the sense of time complexity. The TLRs are a grey-level image, and the tumor regions are extracted using the Otsu thresholding. The Otsu thresholding is utilized for identifying the TLRs to enhance the deep learning training abilities. Initially, a Gaussian smoothing is employed to traditional ROI image, IM_{ORG}, to access the Otsu method to settle on features of the universal image. The Gaussian kernel G is described in Eq. (1) as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Where, (x, y) – pixel's cartesian coordinates; σ – Gaussian kernel standard deviation. The optimum Otsu thresholding, which reduces the variance of class, is a common measure utilized in the analysis of statistical discriminant.

3.3 Feature extraction

After extracting the TLRs, the image features are extracted by using the convolutional neural network (CNN) [23]. The CNN contains two layers such as fully connected layers have to perform the classification and the convolutional layers have to extract the lower and higher order image features. The AlexNet and ResNet50 are the most commonly utilized architectures in the CNN model with the aim of significant achievement in the medical field, especially in the detection of breast cancer. Thus, these two architectures are utilized to achieve the best performance for the proposed system.

3.3.1. AlexNet architecture

The AlexNet [24] is one of the CNN architectures, which is majorly utilized to extract the image features. The AlexNet consists of multiple layers such as five convolutional layers, three pooling layers as well as three fully connected layers, and its basic architecture is depicted in Fig. 2. An AlexNet utilizes $[256 \times 256 \times 3]$ image as an input to the feature extraction.

Every neuron in convolutional layers calculates the dot product among its weights as well as local regions, which are attached to the input volume. All convolutional layers are later with the pooling layer, to execute down sampling tasks besides the spatial dimensions to minimize the number of calculations as well as enhance the robustness of the model. A fully connected layers task is to extract the convolutional layers features and then extracted feature vectors are considered as the output for classification.

3.3.2. ResNet50 architecture

The ResNet is a short form of residual network [25]. The ResNet is a CNN architecture and the convolutional layer with 50 layers is called ResNet50. The ResNet50 is designed by combining the five convolutional layers and it utilizes the input of $[256 \times 256 \times 3]$ five convolutional layers. The convolutional Resnet50 utilizes a neural network comprised of 50 convolution layers, that are trained on the known dataset called CBIS-DDSM to classify the objects into four classes. This method utilizes the pre-trained ImageNet weights. The various kernel sizes are utilized in an average pooling to minimize the variance as well as extract the low-level features from images. The basic idea of ResNet50 is to skip the blocks of convolutional layers by utilizing the shortcut connections of the convolutional layers. The ResNet50 is developed by the residual learning framework and the network layers are given to the residual mapping instead preferred secret mapping between inputs and outputs. These convolutions are a max pool layer, which is followed by 1 fully

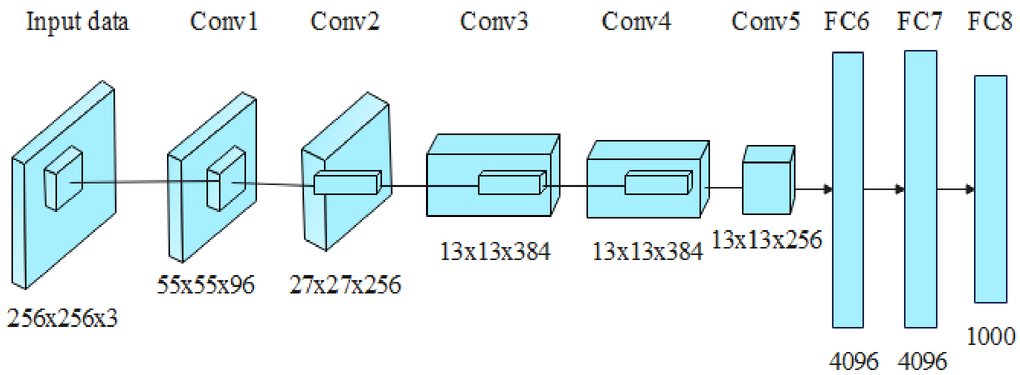

Figure. 2 AlexNet architecture

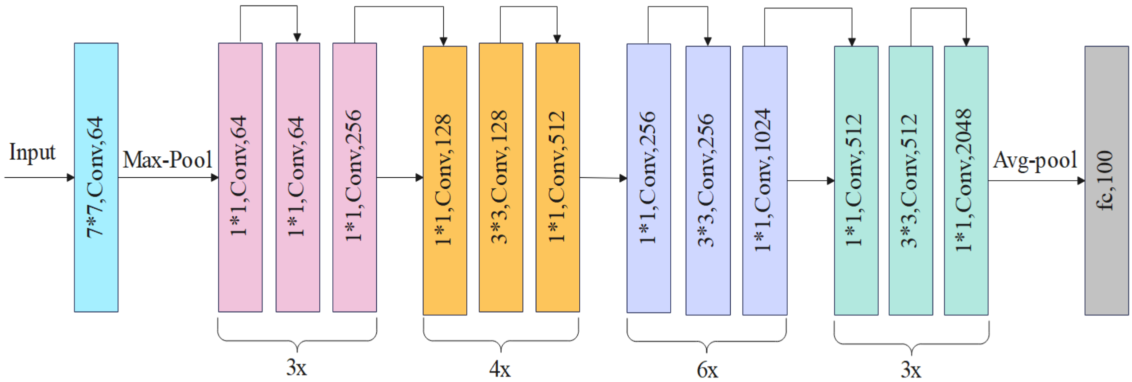

Figure. 3 ResNet50 architecture

connected layer with size $[1000]$. In ResNet50, $[fc1000]$ is utilized to extract features and then used these features as input for the classification. Fig. 3. depicts the basic architecture of ResNet50.

3.4 Classification

After the extraction of the feature, the AdaBoosting algorithm [26] is utilized to classify breast cancer detection. An extracted features output is further proceeded to the classification of images in breast cancer. The AdaBoost is an effective and repetitive ensemble algorithm, that requires less amount of training samples benefit using updating weight of training samples. The AdaBoost algorithm is a classification and iterative-based algorithm.

3.4.1. Weight based AdaBoost algorithm

In the AdaBoost algorithm, the predicted output of the weak learning machine-generated by each iteration is compared with the actual output. A recent training sample as well as a recent weak learning weight compound predictor are updated based on the variation between the predicted and actual output. After updating and iterating training samples for the next weak learning machines, analysis of prediction is performed again until the last weak learning machine stops. Thus, this algorithm majorly merges the amount of weak learning machines' output to make the strongest prediction. Initially, each sample weight is equal and can be formulated in Eq. (2) as;

$$w_k(i) = \frac{1}{n} \tag{2}$$

Where, $w_k(i)$ – sample weight i in k round. The test error is utilized to modify the weak classifier weight, hence to modify the latter classifier process of training as well as train the whole classifier one after another and acquire the strongest classifier based on the weak classifier final weight. In AdaBoost algorithm, is divided into two types such as additive and forward step-by-step models. In the additive model, the binary classification model with T weak classifier, when the training sample is n, the strongest classifier is obtained by the integration, which is formulated in Eq. (3) as;

$$H(x) = sign\big(f(x)\big) = sign\left(\sum_{t=1}^{T} \alpha_t\, h_t(x)\right) \tag{3}$$

In the step-by-step model, the classifier is produced by the new iteration is trained based on the classifier from the last iteration and it can be described in Eq. (4) as;

$$H(x)_m = H(x)_{m-1} + \alpha_t h_t(x) \tag{4}$$

Where, α_t – t^{th} weak classifier weight;

$h_t(x)$ – weak classifier result of the classification in t^{th} iteration;

$H(x)$ – weak classifiers linear combination;

$H(x)_{m-1}$- a combination of weak classifiers in the last iteration.

The calculation of the weight for every weak classifier round is expressed in Eq. (5) and training sample distribution is adjusted using α_t is expressed in Eq. (6) as;

$$\alpha_t = \frac{1}{2} ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{5}$$

$$w_{t-1} = \frac{w_t}{z_t} exp\left(-\alpha_t y h_t(x)\right) \tag{6}$$

Where, w_t – last round distribution weight training samples;

y – label of classification;

z_t – normalization factor.

In this study, an original AdaBoost algorithm is improved by adding the weights of the samples that are classified wrongly in every iteration. The samples are misclassified number of times, which causes the samples' weight to increase frequently and outcomes in a severe weight imbalance sample distribution. To solve this issue, in this study, the weights in the AdaBoost algorithm. Adjusting the weight w_t obtained from t^{th} iteration is to decrease the variance in sample weights between $(t-1)$th iteration and the t^{th} iteration is to decrease the imbalance of sample weight distribution and it can be expressed in Eq. (7) as;

$$s_t = w_{t-1} + \frac{z.e^{-|z|}}{1+e^{-|z|}} \tag{7}$$

Where, $z = w_t - s_{t-1}, w_{t-1}$ - weight acquired from the last iteration.

s_t – adjusted weight.

In the original AdaBoost algorithm, each sample's initial weight is arranged uniformly. The method utilized to arrange each sample weight can mentioned in Eq. (2). In a weight based AdaBoost, each sample's initial weight is arranged based on the time point, that produces the sample. This can be formulated in Eq. (8) as;

$$w_1(i) = \frac{(1.01)^{Y_i - min_\varsigma(Y_\varsigma)}}{\sum_{i=s}^{n}(1.01)^{Y_s - min_\varsigma(Y_\varsigma)}} \tag{8}$$

Where, Y_i – the year that sample i is generated. Thus, the sample will obtain a nearby maximum weight, if this is a newer sample. With the above function, the newest sample can obtain a nearby maximum weight in step 1. A weight-based AdaBoost algorithm achieves better classification in the newest sample.

An AdaBoost algorithm introduces a model of one-hot vector space. The fault classification index location in the mapping vector is 1, and other locations are 0, K is the mapping vector size and it can be expressed in Eq. (9) as;

$$\vec{y}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}, \ldots, \vec{y}_k = \begin{bmatrix}0\\1\\0\end{bmatrix}, \ldots \vec{y}_k = \begin{bmatrix}0\\0\\1\end{bmatrix} \tag{9}$$

Obtaining the output vector of the AdaBoost by combining the output vector of the base classifier and the predicted fault AdaBoost label can be defined in Eq. (10) as;

$$Y^p = arg\,max\,\vec{h}^p(\vec{x}^p(i)) \tag{10}$$

The output $\vec{h}^p(\vec{x}^p)$ of AdaBoost obtained by P different measurement methods is regarded as a sub classifier. The cosine similarity matrix-vector $\vec{c}_t^p$ of the t^{th} classifier is obtained, according to the input one-shot vector $\vec{y}^p$ and the output vector $\vec{y}_t^p$ generated by the t^{th} base classifier. This can be represented by Eqs. (11) and (12) as;

$$\vec{c}_t^p = \left[c_t^p(1) \ldots c_t^p(k) \ldots c_t^p(K)\right], c_t^p(K) = \frac{sim_t^p(K)}{N^p(K)} \tag{11}$$

$$\alpha_t^p(K) = \frac{exp(c_t^p(K))}{\sum_{p=1}^{p} exp(c_t^p(K))} \tag{12}$$

The improved strong classification results are obtained and it is expressed in Eq. (13) as;

$$\vec{H}(\vec{x}) = \sum_{p=1}^{p} \vec{\alpha}^p . \vec{h}_t^p(\vec{x}) = \sum_{p=1}^{p} \sum_{t=1}^{T} \vec{\alpha}_t^p . \beta_t \vec{y}_t \tag{13}$$

The final classification label of the fault diagnosis can be defined in Eq. (14) as;

$$Y(i) = arg\,max(\vec{H}(\vec{x}(i))) \tag{14}$$

In this classification, the AdaBoost algorithm is improved by adding and adjusting the weights to improve the classification performance and accuracy.

Received: August 11, 2023. Revised: September 13, 2023.

Table 1. Performance analysis of various network methods with proposed network method

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
GoogleNet	90.94	90.40	89.45	90.21	89.89
MobileNet	92.86	92.34	92.03	91.82	91.64
AlexNet	95.86	95.64	95.19	94.64	94.23
ResNet50	98.59	98.33	98.08	97.07	97.68
AlexNet-ResNet50	99.56	99.40	99.38	98.89	99.18

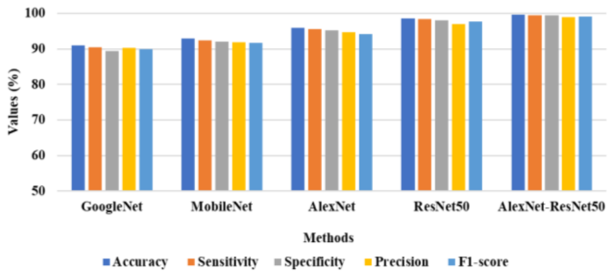

Figure. 4 Graphical representation of proposed methods various network methods

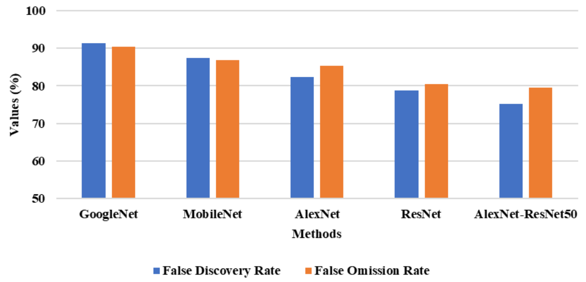

Figure. 5 Graphical representation of proposed methods various network methods

4. Results

The proposed method is implemented on MATLAB software and system specification with Windows 10 operating system, Intel core i5 processor and 4GB RAM. The proposed breast cancer detection method utilized a number of performance metrics to estimate the system performance. Statistical parameters like accuracy, specificity, precision, sensitivity, F1-score, false discovery rate (FDR), false omission rate (FOR) andarea under the curve (AUC) are used for evaluating the proposed method. The mathematical expression for each metric is described as the following Eqs. (15 to 22) as;

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{16}$$

$$Specificity = \frac{TN}{TN+FP} \tag{17}$$

$$Precision = \frac{TP}{TP+FP} \tag{18}$$

Table 2. Performance analysis of various network methods with the proposed network method

Methods	False Discovery Rate (%)	False Omission Rate (%)
GoogleNet	91.39	90.39
MobileNet	87.45	86.75
AlexNet	82.29	85.38
ResNet	78.68	80.46
AlexNet-ResNet50	75.23	79.59

$$F1-score = \frac{2TP}{2TP+FP+FN} \tag{19}$$

$$False\ Discovery\ Rate\ (FDR) = \frac{FP}{TP+FP} \tag{20}$$

$$False\ Omission\ Rate\ (FOR) = \frac{FN}{TN+FN} \tag{21}$$

$$AUC = \frac{\sum R_i(I_I) - I_I(I_I+I)/2}{I_I+I_f} \tag{22}$$

Where,
TP – True positive;
TN – True negative;
FP – False positive;
FN – False negative;
I_I – Number of positive images;
I_f – Number of negative images;
R_i – Rate of ith image.

4.1 Performance analysis

This section shows the quantitative and qualitative analysis of the proposed weight based AdaBoost algorithm with AlexNet and resNet50 for breast cancer detection and early diagnosis in terms of achievable sum rate.

Table 1 and Fig. 4 represents performance analysis of various network for the CBIS-DDSM dataset. The GoogleNet, MobileNet, AlexNet and ResNet50 are measured and compared with the AlexNet-ResNet50. The obtained result shows that the proposed AlexNet-ResNet50 model achieves better results by using performance metrics like Accuracy, Sensitivity, Specificity, Precision, and F1-

Received: August 11, 2023. Revised: September 13, 2023.

Table 3. Performance analysis of proposed methods with existing methods

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
KNN	91.87	91.34	91.18	90.27	90.86
SVM	93.72	93.69	92.39	92.88	92.69
RF	95.87	95.52	95.31	94.15	94.65
AdaBoost	97.96	97.69	97.38	96.02	96.59
Weight based AdaBoost	99.56	99.40	99.38	98.89	99.18

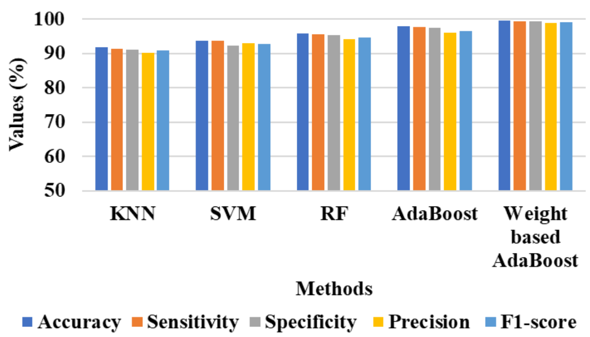

Figure. 6 Graphical analysis of the proposed method with different methods

score values of about 99.56%, 99.40%, 99.38%, 99.89% and 99.18%.

Table 2 and Fig. 5 represents performance analysis of various network for the CBIS-DDSM dataset. The GoogleNet, MobileNet, AlexNet and ResNet50 are measured and compared with the proposed AlexNet-ResNet50. By combining AlexNet and ResNet50, which enables much faster training at each layer, achieves better accuracy. The obtained result shows that the AlexNet-ResNet50 model achieves a better result by using performance metrics like false discovery rate (FDR) and false omission rate (FOR) values of about 75.23% and 79.59%.

Table 3 and Fig. 6 represent the performance analysis of different classification methods for the CBIS-DDSM dataset. The K-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and AdaBoost are measured and compared with the proposed weight based AdaBoost. The obtained result shows that the proposed weight based AdaBoost model achieves better results by using the performance metrics like accuracy, sensitivity, specificity, precision, and F1-score values of about 99.56%, 99.40%, 99.38%, 99.89% and 99.18%.

Table 4 and Fig. 7 represent the performance analysis of different classification methods for the CBIS-DDSM dataset. The K-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and AdaBoost are measured and compared with the proposed weight based AdaBoost. The weight-based AdaBoost increased the accuracy by converting the number of weak learning classifiers

Table 4. Performance analysis of proposed methods with existing methods

Methods	False Discovery Rate (%)	False Omission Rate (%)
KNN	89.93	91.34
SVM	85.29	86.68
RF	81.89	85.27
AdaBoost	79.52	80.34
Weight based AdaBoost	75.23	79.59

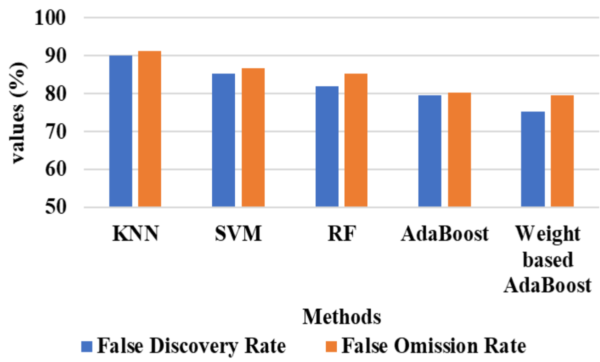

Figure. 7 Performance analysis of proposed methods with existing methods in terms of FDR and FOR

into a single strong classifier. The obtained result shows that the proposed weight based AdaBoost model achieves a better result by using performance metrics like false discovery rate (FDR) and false omission rate (FOR) values about 75.23% and 79.59%.

4.2 Comparative analysis

This section shows the comparative analysis of the proposed weight AdaBoost algorithm in terms of method, datasets, accuracy, sensitivity, specificity, precision, F1-score and AUC of the CBIS-DDSM dataset are shown in Table 5 and the proposed method with MIAS dataset is shown in Table 6. The accuracy, sensitivity, specificity, precision, F1-score and AUC of the proposed weight-based AdaBoost are more efficient than the existing methods. Table 5 shows the comparative analysis of the proposed method with recent research methods using CBIS-DDSM dataset. Table 6 shows the comparative

Table 5. Comparative analysis of proposed method with existing methods of the CBIS-DDSM dataset

Method	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)	AUC (%)	Computational time (sec)
IMPA-ResNet50 [14]	CBIS-DDSM	98.32	96.61	98.56	98.68	97.65	97.88	N/A
Gray difference weight and MSER detector [17]		90.38	84.47	75.71	75.71	N/A	N/A	N/A
MLO and CC [19]		98.87	98.98	N/A	98.79	97.99	98.88	1.2134
Proposed Weight based AdaBoost Algorithm		99.56	99.40	99.38	98.89	99.18	99.29	1.0148

Table 6. Comparative analysis of proposed method with existing methods of the MIAS dataset

Method	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)	AUC (%)
IMPA-ResNet50 [14]	MIAS	98.88	97.61	98.40	98.30	97.10	99.24
Gray difference weight and MSER detector [17]		94.12	90.68	77.70	91.96	N/A	N/A
Proposed Weight based AdaBoost Algorithm		99.29	99.06	99.18	98.97	99.21	99.04

analysis of the proposed method with recent research methods using MIAS dataset.

4.3 Discussion

This section illustrates the limitations of existing methods and explains how the proposed method overcomes such limitations. The existing model such as the computer-aided diagnosis (CAD) [16] method has some limitations which consume more time while training multiple classifiers. Gray difference weight and MSER detector [17] has few risks in some images which has no significant contrast pectoral muscle. The proposed weight-based AdaBoost algorithm achieved better performance by combining the AlexNet and ResNet50 architectures in the feature extraction and increasing the weights in the AdaBoost algorithm. The proposed weight based AdaBoost algorithm on CBIS-DDSM achieved an accuracy of 99.56%, sensitivity of 99.40%, specificity of 99.38%, precision of 99.89, F1-score of 99.18%, AUC of 99.29 and computational time of 1.0148 sec respectively. While on MIAS dataset, proposed weight based AdaBoost algorithm achieved an accuracy of 99.29%, sensitivity of 99.06%, specificity of 99.18%, precision of 98.97, F1-score of 99.21% and AUC of 99.04%.

5. Conclusion

In this research, the breast cancer detection and early diagnosis weight based AdaBoost algorithm is proposed. The proposed method is used to evaluate the breast cancer image datasets named CBIS-DDSM and MIAS. The Otsu thresholding method is utilized to identify the tumor-like-regions (TLRs) to enhance the training abilities. The network architectures of the AlexNet and ResNet50 are utilized for the feature extraction. An AdaBoost algorithm is a classification algorithm, mostly utilized for medical applications, especially in the detection of breast cancer. A weight-based AdaBoost algorithm is used for the classification of the mammogram images into four classes benign calcification (BC), malignant calcification (MC), benign mass (BM) as well and malignant mass (MM). The proposed method achieves better accuracy results in the classification and achieved the accuracy result of 99.56% by comparison of the existing research models and the evaluation results using parameter metrics. In the future, this proposed method will extend to utilize the huge number of datasets to improve the best classification results of breast cancer.

Notation

Variables	Descriptions
(x, y)	Pixel's cartesian coordinates
σ	Gaussian kernel standard deviation
$w_k(i)$	Sample weight i in k round
$H(x)$	Weak classifiers linear combination

$H(x)_{m-1}$	Combination of weak classifiers in the last iteration
$h_t(x)$	Weak classifier result of the classification in t^{th} iteration
α_t	t^{th} weak classifier weight
w_t	last round distribution weight training samples
y	label of classification
z_t	normalization factor
$z = w_t - s_{t-1}, w_{t-1}$	weight acquired from the last iteration
s_t	adjusted weight
Y_i	Year that sample i is generated
K	Mapping vector size
$\vec{h}^p(\vec{x}^p)$	Output of AdaBoost obtained by P different measurement methods
$\vec{c}_t^p$	Cosine similarity matrix-vector of the t^{th} classifier
$\vec{y}^p$	input one-shot vector
$\vec{y}_t^p$	output vector generated by the t^{th} base classifier
$Y(i)$	Final classification label of fault diagnosis

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st and 2nd author. The supervision and project administration, have been done by 3rd author.

References

[1] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, and A. Liotta, “Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning”, *Computer Methods and Programs in Biomedicine*, Vol. 223, p. 106951, 2022.

[2] M. A. Naji, S. E. Filali, K. Aarika, E. L. H. Benlahmar, R. A. Abdelouhahid, and O. Debauche, “Machine learning algorithms for breast cancer prediction and diagnosis”, *Procedia Computer Science*, Vol. 191, pp. 487-492, 2021.

[3] H. Liu, G. Cui, Y. Luo, Y. Guo, L. Zhao, Y. Wang, A. Subasi, S. Dogan, and T. Tuncer, “Artificial intelligence-based breast cancer diagnosis using ultrasound images and grid-based deep feature generator”, *International Journal of General Medicine*, Vol. 2022, No. 15, pp. 2271-2282, 2022.

[4] Y. K. Qawqzeh, A. Alourani, and S. Ghwanmeh, “An improved breast cancer classification method using an enhanced AdaBoost classifier”, *International Journal of Advanced Computer Science and Applications*, Vol. 14, No. 1, 2023.

[5] A. Ahila, M. Poongodi, S. Bourouis, S. S. Band, A. Mosavi, S. Agrawal, and M. Hamdi, “Meta-heuristic algorithm-tuned neural network for breast cancer diagnosis using ultrasound images”, *Frontiers in Oncology*, Vol. 12, p. 834028, 2022.

[6] X. Cai, X. Li, N. Razmjooy, and N. Ghadimi, “Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm”, *Computational and Mathematical Methods in Medicine*, Vol. 2021, p. 5595180, 2021.

[7] K. B. Soulami, N. Kaabouch, M. N. Saidi, and A. Tamtaoui, “Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation”, *Biomedical Signal Processing and Control*, Vol. 66, p. 102481, 2021.

[8] S. Iqbal, A. Imran, and M. Adnan, “Breast Tumor Detection using Machine Learning Boosting Classifiers”, *Journal of Computing & Biomedical Informatics*, Vol. 4, No. 01, pp. 118-131, 2022.

[9] I. Sechopoulos, J. Teuwen, and R. Mann, “Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art”, *Seminars in Cancer Biology*, Vol. 72, pp. 214-225, 2021.

[10] A. Das, M. N. Mohanty, P. K. Mallick, P. Tiwari, K. Muhammad, and H. Zhu, “Breast cancer detection using an ensemble deep learning method”, *Biomedical Signal Processing and Control*, Vol. 70, p. 103009, 2021.

[11] A. Khamparia, S. Bharati, P. Podder, D. Gupta, A. Khanna, T. K. Phung, and D. N. H. Thanh, “Diagnosis of breast cancer based on modern mammography using hybrid transfer learning”, *Multidimensional Systems and Signal Processing*, Vol. 32, pp. 747-765, 2021.

[12] R. Brahimetaj, I. Willekens, A. Massart, R. Forsyth, J. Cornelis, J. D. Mey, and B. Jansen, “Improved automated early detection of breast cancer based on high resolution 3D micro-CT microcalcification images”, *BMC Cancer*, Vol. 22, p. 162, 2022.

[13] T. Kavitha, P. P. Mathai, C. Karthikeyan, M. Ashok, R. Kohar, J. Avanija, and S.

Neelakandan, "Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images", *Interdisciplinary Sciences: Computational Life Sciences*, Vol. 14, No. 1, pp. 113-129, 2022.

[14] E. H. Houssein, M. M. Emam, and A. A. Ali, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm", *Neural Computing and Applications*, Vol. 34, No. 20, pp. 18015-18033, 2022.

[15] R. S. Cauce, J. P. Martín, and M. Luque, "Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data", *Computer Methods and Programs in Biomedicine*, Vol. 204, p. 106045, 2021.

[16] V. R. Allugunti, "Breast cancer detection based on thermographic images using machine learning and deep learning algorithms", *International Journal of Engineering in Computer Science*, Vol. 4, No. 1A, pp. 49-56, 2022.

[17] B. V. Divyashree and G. H. Kumar, "Breast cancer mass detection in mammograms using gray difference weight and mser detector", *SN Computer Science*, Vol. 2, p. 63, 2021.

[18] S. A. Alanazi, M. M. Kamruzzaman, M. N. I. Sarker, M. Alruwaili, Y. Alhwaiti, N. Alshammari, and M. H. Siddiqi, "Boosting breast cancer detection using convolutional neural network", *Journal of Healthcare Engineering*, Vol. 2021, p. 5528622, 2021.

[19] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: Automated CNN approach", *Alexandria Engineering Journal*, Vol. 60, No. 5, pp. 4701-4709, 2021.

[20] J. V. Tembhurne, A. Hazarika, and T. Diwan, "BrC-MCDLM: breast Cancer detection using Multi-Channel deep learning model", *Multimedia Tools and Applications*, Vol. 80, No. 21-23, pp. 31647-31670, 2021.

[21] CBIS-DDSM dataset link: https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset.

[22] MIAS dataset link: https://www.kaggle.com/datasets/kmader/mias-mammography.

[23] M. Amini, M. M. Pedram, A. Moradi, and M. Ouchani, "Diagnosis of Alzheimer's disease severity with fMRI images using robust multitask feature extraction method and convolutional neural network (CNN)", *Computational and Mathematical Methods in Medicine*, Vol. 2021, p. 5514839, 2021.

[24] S. Mohapatra, S. Muduly, S. Mohanty, J. V. R. Ravindra, and S. N. Mohanty, "Evaluation of deep learning models for detecting breast cancer using histopathological mammograms Images", *Sustainable Operations and Computers*, Vol. 3, pp. 296-302, 2022.

[25] D. Clement, E. Agu, M. A. Suleiman, J. Obayemi, S. Adeshina, and W. Soboyejo, "Multi-class breast cancer histopathological image classification using multi-scale pooled image feature representation (mpifr) and one-versus-one support vector machines", *Applied Sciences*, Vol. 13, No. 1, p. 156, 2023.

[26] K. M. M. Uddin, N. Biswas, S. T. Rikta, and S. K. Dey, "Machine learning-based diagnosis of breast cancer utilizing feature optimization technique", *Computer Methods and Programs in Biomedicine Update*, Vol. 3, p. 100098, 2023.