Journal of Autonomous Intelligence (2024) Volume 7 Issue 3
doi: 10.32629/jai.v7i3.1288

ORIGINAL RESEARCH ARTICLE

A deep convolutional neural network architecture for breast mass classification using mammogram images

Sivagami G.[*], Vidya K., Geetharamani R.

Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai 600025, India

*** Corresponding author:** Sivagami G., shiv.haida@gmail.com

ABSTRACT

Breast cancer is one of the second most common cancer occurring worldwide. Early identification of the disease is a major interest that promises to propose several diagnostic procedures to prevent further surgical interventions. This research paper aims to develop a breast mass classifier system using deep learning to differentiate breast mass images from normal mammographic images. The benchmark mammographic datasets CBIS-DDSM, INbreast, and mini-MIAS are used for constructing the proposed model DELU-BM-CNN. The region of interest is identified by applying image processing techniques (median filter, binarization and dilation) and the images are enhanced and sharpened using adaptive histogram equalization and unsharp masking techniques. The pre-processed images are trained with a minimum of five deep convolutional layers activated by an Exponential Linear Unit (ELU) which is developed from scratch for feature learning and classifying the given whole mammographic images. Dropout, Data normalization, and Global average pooling are some of the regularization techniques adopted to prevent the model from over-fitting. The proposed models are able to classify CBIS-DDSM images with an accuracy of 96.60%, INbreast images with 96.20% and MIAS images with 97.40%. The experimental results are also compared with conventional Rectified Linear Unit (ReLU) and Leaky ReLU activation function that promises the proposed model as a good prognosticator than the state-of-art models for cancer diagnosis using mammogram images as input.

Keywords: Breast pathologies; classification; convolution neural network; Exponential Linear Unit (ELU); image processing, mammogram

ARTICLE INFO

Received: 26 September 2023
Accepted: 6 November 2023
Available online: 29 December 2023

1. Introduction

Breast cancer is the most common cancer occurring worldwide. In 2020, globally 16% and approximately 685,000 women died from this cancer and a total of 2.3 million new cases are reported from both sexes[1,2]. Breast cancer screening is an agent to check for the disease's early symptoms, thereby reducing the mortality rate which accounted for 17% for this cancer[3]. Some of the breast imaging modalities include X-ray based like Mammography, Using ultrasound like Doppler, Sonography, Magnetic field-based like Magnetic Resonance Imaging (MRI), and Gamma radiation based like Positron Emission Tomography (PET)[4]. Irrespective of various breast screening methodologies, there occurs a serious concern about having fewer breast radiologists to analyze and interpret these results accurately. And also a double-read requirement of the screened images is a difficult strategy to run the screening programs effectively and efficiently[5].

The digitization of screening reports led to an era of Computer-

Aided Detection and Diagnosis (CAD). It also provides an exemplary way in assisting health professionals thereby overcoming the resource shortage problem. The disease identification lies in the quality of the digitized reports. Report quality can be improved with image processing algorithms and pattern recognition helps in defining the focal point for the radiologist to understand and analyze the regions affected more precisely[6]. The revolution of Deep Learning (DL) emerged exponentially with the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) intended to classify images of 1000 classes and it was won by AlexNet[7]. The implementation of DL methods for computer vision problems has paved the way for its application in medical image analysis[7–10]. DL methods exceed the traditional methods in terms of automatic feature learning, and end-to-end implementation of CAD that includes detection, segmentation, and classification of the given objective function with a unified training process[11].

With wide varies of DL models such as Convolution Neural Networks (CNN), Encoders and Decoders, Generative Adversarial Networks (GAN), and many more, the applications are not limited to specific domains and spread across Natural Language Processing (NLP), Speech recognition, Medical field, Computer Vision, Transportation, Deep learning on graphs and so on[12], this research work focuses on proposing a binary classification model using CNN that is able to classify the normal mammographic images from mass images.

The main contribution of the proposed work is

(i) A minimum of five convoluted network model is developed from scratch to classify normal and mass mammographic images.
(ii) Deep feature extraction is proposed for all the convolutional block.
(iii) An Exponential Linear Unit (ELU) activation is suggested compared to conventional ReLU and Leaky ReLU activation functions.
(iv) Image enhancement and image sharpening processes are performed for better pixel feature extraction.

This paper is organized as follows. Section 2 deals with the state-of-the-art models developed for breast pathology classification for different image inputs. Section 3 focuses and describes on the proposed methodology carried out for this research along with the dataset utilized and the image pre-processing carried out. Section 4 is detailed with experimental results and also compares with the state-of-art results while section 5 concludes the research findings.

2. Review on related studies

A diverse classification approach was proposed by researchers for the classification of breast lesion from normal images. The recent analysis approach for breast mass classification based on mammogram images and other images is discussed in this section.

An end-to-end whole mammogram image classifier was built based on constructing an efficient patch classifier thereby avoiding the Region of Interest (ROI) annotations. By using the pre-trained deep learning models an AUC in the range of 0.65 to 0.97 for the classification of breast cancer was achieved for the CBIS-DDSM and INBreast dataset[13]. A text on feature representation provided by rotation-invariant were combined with a deep CNN for classifying breast masses was tested on the CBIS-DDSM dataset with an accuracy of 94.30%[14]. Breast Mass Classification (BMC) system was developed by Malebary and Hashmi[15] with a combination of k-means, Long Short-Term Memory network (LSTM) of Recurrent Neural Network (RNN), CNN, Random Forest (RF) and boosting techniques. 96% and 95% accuracies were reported for the DDSM and MIAS dataset. Density-based CNN's were built[16] to fuse local and global features to promote a diverse contextual feature extraction mechanism to differentiate benign and malignant breast masses. An accuracy of 94.90% was achieved for the CBIS-DDSM dataset. A novel classification deep learning algorithm was developed by Sun et al.[17] by combining the features from the MLO and CC view of the mammographic mass images and thereby improving the objective function by adding a penalty term. The classifier was able to predict with 73.55% accuracy for the Digital Database for Screening Mammography (DDSM) dataset. An integrative approach of deep learning with an extreme learning machine was proposed by Sannasi Chakravarthy and Rajaguru[18] through a crow search optimizer that was able to predict with 97.19% for DDSM, 98.13% for MIAS and 98.26% for INbreast datasets. A 94% and 97% accuracy was achieved by training a customized CNN for the MIAS and a private dataset by Mahmood et al.[19] using the data augmentation and the Rectified linear unit activation function. Soulami et al.[20] proposed a capsule network deep learning architecture that was able to classify the suspicious breast mass regions from the mammograms with an accuracy of 96.03%. They also performed a multi-classification of masses into normal, benign, and malignant with an accuracy of 77.78%. A VGG architecture-based breast segmentation was developed by Sengan et al.[21] using the High-Resolution Multi-View Deep Convolutional Neural Network (HRMP-DCNN) and also utilized the weakly labeled data for the betterment of breast cancer identification. A stacked ensemble of Residual neural networks was experimented by Baccouche et al.[22] for the classification and segmentation of breast masses into benign and malignant. With XGBoost optimization, they classified with 95.13% for CBIS-DDSM and 99.20% for the INbreast dataset. A three-stage Probabilistic Anchor Assignment model-based algorithm was designed[23] for the classification of mammograms and for the detection of mass and calcification. They also concluded that multiple inputs when combined with multi views can improve the model performance rather than a single viewpoint. Baccouche et al.[24] proposed a YOLO based to classify and detect the lesions from mammograms. They did an image-to-image translation to create synthetic images and achieved an accuracy of 93% for classifying mass lesions. Digital data gets increased on a daily basis. The virtual health records get accessed by many thereby posing a threat to patient's privacy. A novel approach of reversible data concealment of the health records securely with the use of visual cryptography for medical image records was proposed by Rajesh Kumar et al.[25] and also peak-signal-to-noise ratio, mean squared error and normalized correlations are suggested for metric calculations.

A review of deep learning algorithms developed using breast MRI datasets was discussed in Adam et al.[26]. Several DL networks such as AlexNet, ResNet18, LSTM, ResNet50, DenseNet are worked out for a promising accuracy for the wide variety of MRI data. The authors concluded the limitations of dataset with respect to sample size, skewness and availability of large amount of high quality data with annotation for training the DL models. Another review paper[27] discussed on the deep learning applications for the different breast modalities like digital breast tomosynthesis, digital mammogram, ultrasound, MRI and Nuclear medicine imaging. The authors suggested for more research to happen for ultrasound and MRI modalities as there exists minimum datasets, whereas huge contributions are happening towards digital breast tomosynthesis and digital mammogram datasets.

A complete diagnostic approach was proposed by Civilibal et al.[28] for detecting normal and abnormal breast tissues from thermal images. The Mask R-CNN method with keeping ResNet-50 as backbone were able to classify with 97.10% accuracy. DenTnet method was proposed by Wakili et al.[29] and reported an accuracy of 99.28% for the histopathological BreaKHis dataset by using DenseNet as a backbone network for the transfer learning approach. They had also made a study to identify the train-test split ratio for the histopathology images. EfficientNet-B0 was applied to classify histopathological breast cancer images with the concept of multiple instance learning. Ahmed and Islam[30] were able to produce 99.52% for binary classification, 100% for Four-class (benign), 99.31% for four-class (malignant) and 99.04% for eight class. An improved version of InceptionV3 were developed by Sirjani et al.[31] that were able to attain an accuracy of 81.00% for breast lesion classification using ultrasound images. Feature fusion were obtained from the VGG19 and DenseNet201 neural network model and an HybridMultiNet framework was proposed by Rahman et al.[32] and reached an accuracy of 99.20% for BreakHis and 95.20% for the ICIAR , the two public histopathological datasets. Two new datasets had been provided by Çayır et al.[33] for the detection of the nucleus and classification of mitosis from the whole slide images of breast cancer. A collation of Cross Stage Partial

Network, DarkNet53 and Path Aggregation Network were used for the feature extraction and You Look Only Once YOLO-v4 were used for nucleus detection purpose. The pre-trained VGG16 model with decision tree classifier[34] were able to predict the classification accuracy of 95.50% from thermal images to differentiate normal from abnormal. Feature optimization using Firefly was executed.

Although there are many sophisticated models developed, this paper focus on a simple neural network architecture with deeply convoluted DELU-BM-CNN that unfolds the features of the given whole mammographic images with minimum number of layers.

3. The proposed system

This section describes on the standard dataset used for this research purpose. Image pre-processing is taken as a next step for furnishing the standard input data with better quality and thereby resulting in determinant features with which the proposed model work. This is followed by DELU-BM-CNN, proposed model architecture with description of layer configurations.

3.1. Dataset details

Three publicly available benchmark dataset for digital mammograms are considered to develop the proposed DELU-BM-CNN model. The datasets are Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM)[35], the INbreast dataset[36] and the Mammographic Image Analysis Society mini-MIAS[37].

The CBIS-DDSM dataset focus on the breast abnormalities of calcification and mass. From the dataset, the total benign images of 845 are considered as normal and malignant mass images of 739 are taken as mass images for the experimental purpose. The INbreast dataset is of the Digital Imaging and Communications in Medicine (DICOM) with the resolution of 3328 × 4084 or 2560 × 3328 pixels. The total count of normal and mass mammogram images for this dataset is 67 and 107. The mini-MIAS dataset comes in pixel resolution of 1024 × 1024 as Portable Graymap File Format (PGM) with normal images of 209 and altogether mass images (circumscribed, speculated and other masses) of 56. An overview of mammogram image dataset is provided in **Table 1**.

Table 1. Publicly available mammogram datasets.

Dataset	Number of cases	Image format	Image dimensions	Normal	Mass
CBIS-DDSM	1566	DICOM	Varying dimensions	845	739
INbreast	115	DICOM	3328 × 4084 2560 × 3328	67	107
mini-MIAS	161	PGM	1024 × 1024	209	56

3.2. Data pre-processing

Among breast cancer screening methods, Mammography is one of the prominent methods accepted and followed widely. With the advent of digitization, it provides an additional support for the radiologist to identify the lesion precisely and swiftly. The below pre-processing steps are carried out for the above datasets.

i Three datasets come with different image formats. CBIS-DDSM and INbreast dataset appears as DICOM images whereas mini-MIAS is in PGM image format. It is necessary to bring all the different image formats into a common format and hence converted into Portable Network Graphics (PNG) image format.

ii For the CBIS-DDSM dataset, there is an enclosure of white spaces around the breast region which makes it difficult to extract the region of interest for further processing. And hence, the following steps are taken exclusively for this dataset to pre-prepare the images for further pre-processing.

  a. The images are resized to 256 × 256.

b. The images are cropped in the ratio of 10:10:236:236 where the first two values represent the *x* and *y* coordinates of the image and later two values are the width and height of the image to be cropped. [The mask images are cropped in the same ratio to ensure the lesions are not missed].

iii Median filter is a non-linear statistical filtering mechanism. It works by replacing the working pixel value with the median value of the surrounding pixels.

iv The median filtered image is binarized after a careful selection of the threshold value of "0.1".

v Dilation of the binarized breast image object is performed by filling the holes which result in the better visibility of the image.

vi The breast region occupies the maximum pixels of the image than the artefacts and labels. The breast region which is the region of interest can be chosen using the bounding box by keeping the influential region and discarding the unwanted portions from the image.

vii The bounding box selected region is resized to 512 × 512 dimension that is provided as an input dimension to the CNN architecture.

viii The contrast of the image is amplified using adaptive histogram equalization (CLAHE) by keeping the tile and the enhancement limit as 8 × 8 and 0.01.

ix In addition, the images are sharpened using the Unsharp masking by keeping the sharpening strength value as 0.7.

x The images are converted to an RGB and fed as an input with size of 512 × 512 × 3 to the DELU-BM-CNN.

The proposed network architecture is detailed in **Figure 1** as an end-to-end implementation of the breast mass classification system.

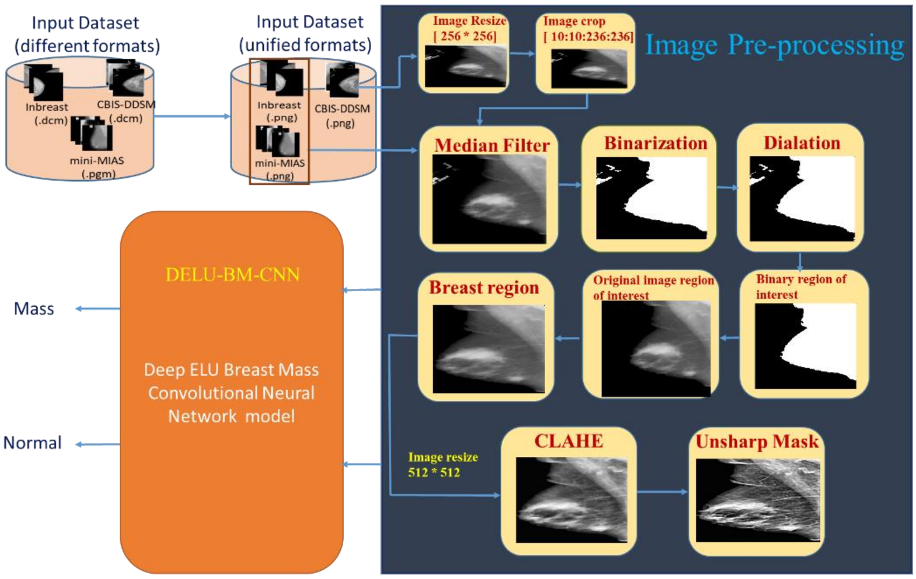


Figure 1. Proposed end-to-end implementation of the breast mass classification system.

3.3. Proposed DELU-BM-CNN

The Deep Exponential Linear Unit for constructing the Breast Mass Convolutional Neural Network (DELU-BM-CNN) is a stacked 5 layer convoluted model with ELU activation function. The input data is batch normalized during training process and drop out strategy is carried out to avoid over-fitting problem. The main layers in this model are CNN, ELU activation function, Batch Normalization, Max Pooling, Dropout and Global Average Pooling.

Convolutional Neural Network (CNN) ranks top in image classification by its ability to learn the higher order features via convolutions[38]. The Convolution operation requires a kernel filter to convolve with the input data. The kernel filter is otherwise termed as feature detector that performs the scalar product of the input data that is operated with the kernel filter[39]. A convolutional matrix is thus formed which forms the main building block of any CNN architecture. The output of the convolutional matrix[40] can be given as

$$\mathrm{Conv_{out}} = \frac{(\mathrm{I_{dim}} - \mathrm{KF_{dim}} + 2 * \mathrm{P})}{\mathrm{S}} + 1 \tag{1}$$

where $Conv_{out}$ is the number of neurons at each convoluted row output. I_{dim} is the data input dimension, KF_{dim} is the kernel filter dimension, P is the padding value and S being the stride used in that row. The kernel filters are the weights that convolve along the width and height of the input data in a sliding manner to get the convolutional output. If input data is of n-dimensional, the convolve operation is executed for n dimensions. Several kernel filters are applied to the set of neurons for feature learning which lead to an effective image recognition.

Activation functions are non-linear transformation functions that captures the abstract features as a learning process. Exponential Learning units (ELU) tries to keep the mean activation closer to zero that accelerates to a higher classification accuracy than Rectified Linear unit with a faster learning rate.

The ELU is a function f(x) satisfying with 0 < α

$$\mathrm{f(x)} = \begin{cases} \mathrm{x}, & \mathrm{x} > 0 \\ \alpha(\exp(\mathrm{x}) - 1, & \mathrm{x} \leq 0 \end{cases} \tag{2}$$

And its derivative is given as

$$\mathrm{f}'(\mathrm{x}) = \begin{cases} 1, & \mathrm{x} > 0 \\ \mathrm{f(x)} + \alpha, & \mathrm{x} \leq 0 \end{cases} \tag{3}$$

The α parameter is the ELU hyper parameter that controls the saturation point for negative net inputs[41].

Batch Normalization is performed for mini-batch by keeping the mean output close to 0 and standard deviation close to 1.0[42]. It produces a smoothness with respect to parameter space and gradients. This layer is added to each of the convolutional layer thereby normalizing the inputs by reducing the internal covariance.

The preferred pooling is Max-Pooling in the CNN architectures. Within the given set of feature map, the Max-Pooling operation picks out the maximal value and get replaced in the output of the convolution operation. The Max-Pooling function can be defined as[43].

$$\mathrm{a_j} = \underbrace{\max}_{\mathrm{N\,X\,N}} \left(\mathrm{a_i^{nX\,n}} \mathrm{u(n,n)}\right) \tag{4}$$

where the window function of u(n,n) is applied to the input data of ai. The maximum value of the operated window size is projected as the resultant matrix which is of reduced size than input data.

“Dropout” is dropping off either hidden or visible units of the CNN model randomly thereby exponentially combining the different neural network architectures and also it prevents the model from overfitting. It had proved to improve the generalization performance for various datasets[44]. Dropout value of 0.5 is used before fourth convolutional block and GAP. With the dropout value of 0.5, the neurons in the given layer are set to zero thereby forcing the network to learn the robust features and providing the generalizability.

An extension of the Stochastic Gradient Descent (SGD), Adam optimization algorithm is used as an optimizer for the proposed model. Adam is based on the adaptive moments relying on the first and second order moments. Adam outperform from SGD and RMSProp.

Global Average Pooling (GAP) performs linear transformation of feature maps, thus reducing the spatial dimension from h × w × d to 1 × 1 × d. This reduction is done by taking average of all ‘hw’ values without sacrificing the performance[45].

The structure of the proposed DELU-BM-CNN architecture is given in **Figure 2**.

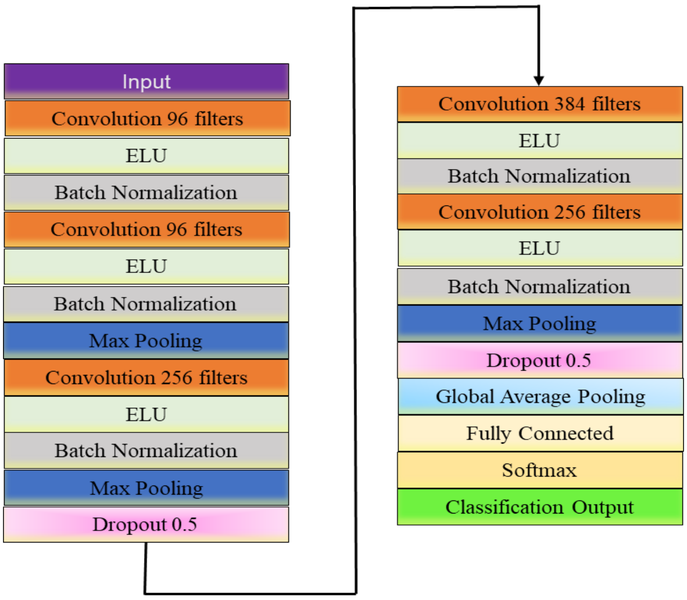


Figure 2. Layer details of DELU-BM-CNN architecture.

DELU-BM-CNN architecture consists of RGB input images of size 512 × 512. This network consists of 25 layers, out of which 5 belongs to the convolutional layers. Each of the convolutional layer is followed by the ELU and Batch Normalization layers. For easier notation, let these three layers can be combine as a single convolutional block. The Max pooling operation with a stride rate of 2 × 2 and with zero padding is implemented after second, third and fifth convolutional block. In two places, the drop out strategy is executed to ensure reduction in the network over-fitting. Every convolutional layer of the architecture extract deeper features of the input image by using a 3 × 3 filter size. The extracted deep features are taken an average using GAP and fed into next layer as input to the fully connected layer. This layer aims to solve the binary classification problem by deep features extracted from each of the convolutional blocks. A categorical distribution is provided by the softmax activation function, a normalized exponential function. The layer-level configuration details of the proposed architecture are provided in **Table 2** with the activations and the learnable parameters for each layer.

Table 2. DELU-BM-CNN configuration details.

Layer details	**Activations**	**Padding**	**Stride**	**Learnable properties**		**Total learnables**
Image input	512 × 512 × 3	-	-	-	-	0
2D Convolution—1	128 × 128 × 96	1	4	Weights 3 × 3 × 3 × 96	Bias 1 × 1 × 96	2688
ELU—1	128 × 128 × 96	-	-	-	-	0
Batch Normalization—1	128 × 128 × 96	-	-	Offset 1 × 1 × 96	Scale 1 × 1 × 96	192
2D Convolution—2	128 × 128 × 96	1	1	Weights 3 × 3 × 96 × 96	Bias 1 × 1 × 96	83,040
ELU—2	128 × 128 × 96	-	-	-	-	-
Batch Normalization—2	128 × 128 × 96	-	-	Offset 1 × 1 × 96	Scale 1 × 1 × 96	192
Max Pool—1	63 × 63 × 96	0	2	-	-	-
2D Convolution—3	32 × 32 × 256	1	2	Weights 3 × 3 × 96 × 256	Bias 1 × 1 × 256	221,440

Table 2. (*Continued*).

Layer details	Activations	Padding	Stride	Learnable properties		Total learnables
ELU—3	32 × 32 × 256	-	-	-	-	0
Batch Normalization—3	32 × 32 × 256	-	-	Offset 1 × 1 × 256	Scale 1 × 1 × 256	512
Max Pool—2	15 × 15 × 256	0	2	-	-	-
Dropout—1	15 × 15 × 256	-	-	-	-	-
2D Convolution—4	8 × 8 × 384	1	2	Weights 3 × 3 × 256 × 384	Bias 1 × 1 × 384	885,120
ELU—4	8 × 8 × 384	-	-	-	-	0
Batch Normalization—4	8 × 8 × 384	-	-	Offset 1 × 1 × 384	Scale 1 × 1 × 384	768
2D Convolution—5	8 × 8 × 256	1	1	Weights 3 × 3 × 384 × 256	Bias 1 × 1 × 256	884,992
ELU—5	8 × 8 × 256	-	-	-	-	0
Batch Normalization—5	8 × 8 × 256	-	-	Offset 1 × 1 × 256	Scale 1 × 1 × 256	512
Max Pool—3	8 × 8 × 256	0	2	-	-	0
Dropout—2	8 × 8 × 256	-	-	-	-	0
Global Average Pooling	8 × 8 × 256	-	-	-	-	0
Fully Connected	1 × 1 × 2	-	-	Weights 2 × 256	Bias 2 × 1	514
Softmax	1 × 1 × 2	-	-	-	-	0
Classification	1 × 1 × 2	-	-	-	-	0

4. Experimental results

This section discusses on the experimental analysis and the performance results of the breast mass classification based on mammographic images. Curated DDSM, INbreast and the MIAS are the gold standard mammographic dataset that are used to train the novel CNN to build a predictor model system. The proposed system classifies the given three datasets into normal and mass images.

An evaluation metric of train-validation-test split is used to estimate the performance of the proposed predictor model. The given dataset is randomly split in the ratio of 75:15:15 as train, validation and test set respectively. The hyper parameters are selected after various trials of experimental analysis and are listed in **Table 3**.

Table 3. Hyper-parameter settings.

Parameter	Value
Optimizer	Adam
Mini-batch size	64
Epoch	5 for CBIS-DDSM and 200 for MIAS and InBreast
Learning Rate	0.0001
Shuffling	After every epoch

Metrics are tied with a model objective and it is a quantitative assessment to compare with a given context. They also act as a decision-making tool. Among different metrics, for the classification models, Sensitivity, Specificity, Accuracy, and F1-score values are considered. Sensitivity is to showcase the model's performance towards the positive instances (masses). It indicates how many positive instances were able to identify by the model. A perfect model should be able to predict positive results for the targeted but high sensitivity alone doesn't satisfy the model requirement. It is also equally important to identify all the negatives (normal) i.e., persons without disease. This is given by the specificity values. Accuracy measure is the ability to differentiate the mass from the normal images. Model accuracy values goes well for balanced dataset and F1-score values for unbalanced dataset. The parameters TP, TN, FP, FN are True Positives, True Negatives, False Positives and False Negatives which will be used for the metric calculation purposes.

TP are the number of samples *correctly* predicted as *positive,*

FP are the number of samples *wrongly* predicted as *positive,*

TN are the number of samples *correctly* predicted as *negative* and

FN are the number of samples *wrongly* predicted as *negative*.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{5}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \tag{6}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{7}$$

The F1-score is based on precision and recall values. It is a harmonic mean of precision and recall values. The value ranges from 0 to 100 % and a higher F1-score better the model quality.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

F1-score can be calculated as

$$\text{F1} - \text{score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

$$\text{F1} - \text{score} = \frac{2\ \text{TP}}{(2\ \text{TP} + \text{FP} + \text{FN})} \tag{11}$$

4.1. Classification accuracy results

The performance of the proposed model for the classification of mass and normal mammographic images for the three different datasets are accessed with the results obtained from the evaluators of accuracy, sensitivity and specificity. The results are acquired for the dataset split of 70% as training, 15% as validation and 15% as testing. The top three different activation function results are also compared to determine the apt activation function to build the proposed classifier model.

The public dataset MIAS consists of different mass which includes circumscribed, speculated and other masses. All these categories are included under a single mass category and the model DELU-BM-CNN is trained accordingly. The model is able to predict the unseen data with an accuracy of 97.40% for the ELU activation function which is slightly greater than the RELU and leaky RELU functions with an average increase of 2.15% accuracy difference. The Sensitivity reported is 100 % with specificity and F1-score as 96.77% and

94.11%. The graphical representation of the model accuracy for this dataset is depicted in **Figure 3**.

The INbreast dataset is a full-field digital mammographic database that is fully focused on detection of mammary lesions. The proposed CNN model was able to detect with an accuracy of 96.2% with sensitivity of 93.75%, specificity of 100 % and F1-score of 96.77%. The RELU was able to classify with an accuracy of 84.60% while leaky RELU with 88.50%. This can be visualized in **Figure 4**.

A quiet a huge number of images are available for analysis from the benchmark dataset, CBIS-DDSM for the mass classification from the normal breast images. This dataset mainly focuses on the two breast pathologies like masses and calcifications. With around 735 mass images, the DELU-BM-CNN model were able to predict with an accuracy for the ELU activation of 96.60%. Whereas RELU was able to predict with 95.00% and leaky RELU with a 94.50% of accuracy. The model was able to produce sensitivity of 98.13%, specificity of 95.41% and F1-score of 96.33% respectively. The DELU-BM-CNN model comparison for the different activation functions for the CBIS-DDSM dataset is compared and represented in **Figure 5**.

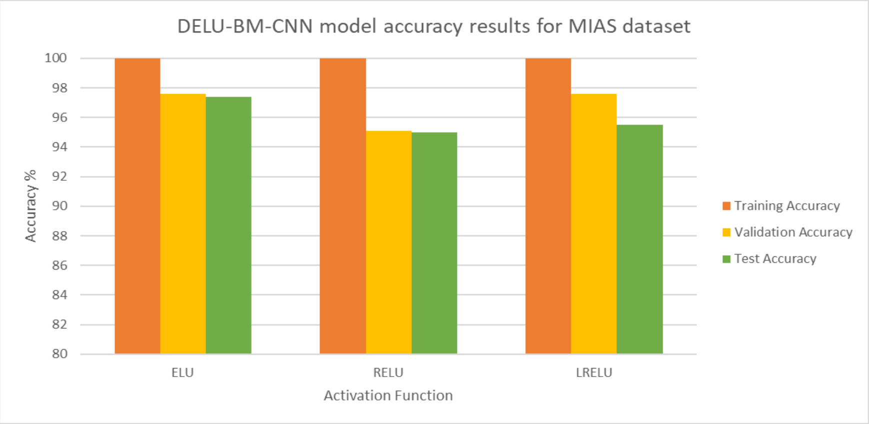


Figure 3. Classification model accuracy results for MIAS dataset.

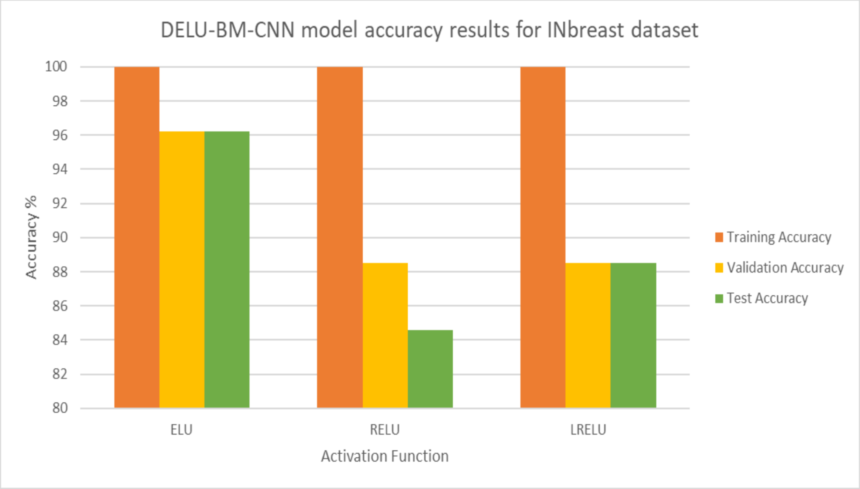


Figure 4. Classification model accuracy results for INbreast dataset.

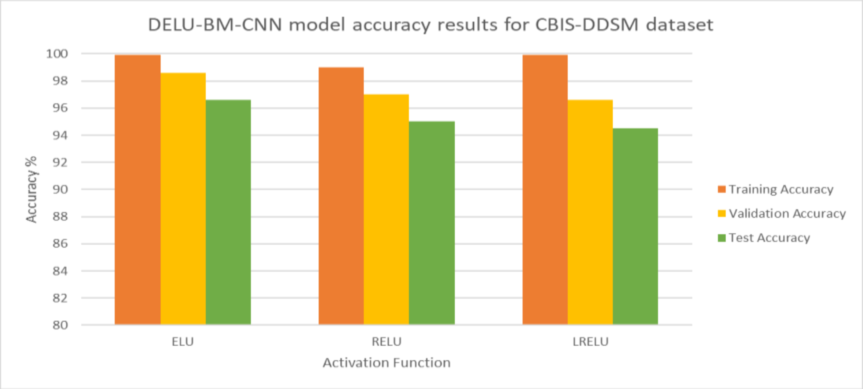
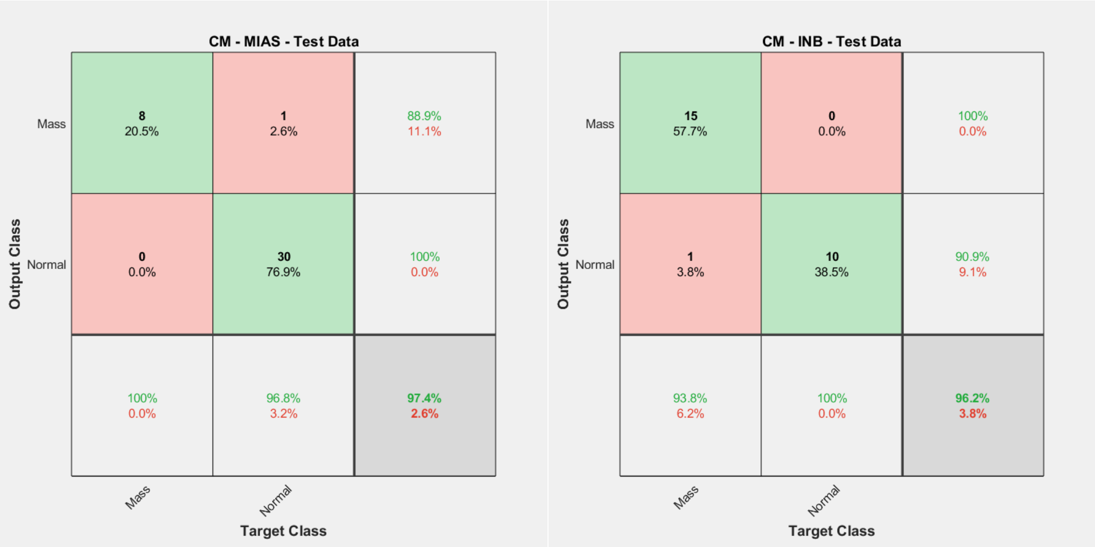


Figure 5. Classification model accuracy results for CBIS-DDSM dataset.

The confusion matrix is a table representation that summarizes the model performance. The number of rows and columns depends on the number of class labels. Since the proposed model is a binary classifier, two rows represent true positives, and false negatives while two columns represent false positives and true negatives. The four values for the test data are tabulated below for each of the datasets. The mini-MIAS dataset consists of total 265 images in which 185 images are trained by keeping 41 images for validation and the remaining 39 images for testing purpose. From the confusion matrix given in **Figure 6**, “CM-MIAS-Test Data”, out of 39 images, 8 images belong to mass and 31 images belong to normal category. All mass images were able to identify correctly whereas one image from the normal data was wrongly classified under mass category and hence the sensitivity is 100% and specificity is 96.77%. The F1-score of 94.11% is calculated from the precision value of 88.88% and recall of 100%. Similarly for the INbreast dataset, 122 images are trained and the 26 images improved the model by inputting as a validation data. The test data comes to 26, 16 belongs to mass and 10 belongs to normal images. The models were able to predict all the normal images correctly thus pushing the specificity to 100%. Here one image affected from mass was wrongly considered as a normal image leading to FN = 1 and sensitivity reporting to 93.75%. With precision being 100% and recall being 93.75%, the F1-score calculated to get 96.77%. The CBIS-DDSM dataset were partitioned into 1109 images for training, 237 images for validation and 238 images for testing with a total count of 1584 mass and normal images. In the test data, 6 images are wrongly classified into mass category and 2 images are misclassified into normal category thus giving FP = 6 and FN = 2. This produces and sensitivity of 98.13% and specificity of 95.42%. The accuracy reported is 96.64%. F1-score value for the third dataset is expressed as 96.33% with precision of 94.59% and recall of 98.13%.



CM - MIAS - Test Data

Output Class \ Target Class	Mass	Normal	
Mass	8 20.5%	1 2.6%	88.9% 11.1%
Normal	0 0.0%	30 76.9%	100% 0.0%
	100% 0.0%	96.8% 3.2%	97.4% 2.6%

CM - INB - Test Data

Output Class \ Target Class	Mass	Normal	
Mass	15 57.7%	0 0.0%	100% 0.0%
Normal	1 3.8%	10 38.5%	90.9% 9.1%
	93.8% 6.2%	100% 0.0%	96.2% 3.8%

Figure 6. (*Continued*).

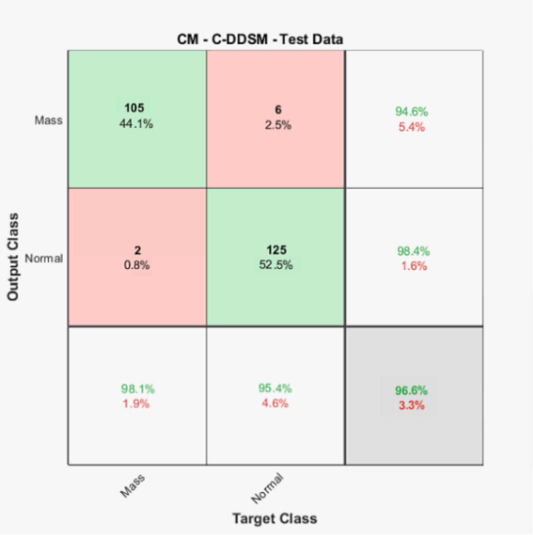


Figure 6. Confusion Matrix representation of models for three datasets.

The model evaluation results discussed above for all the three datasets are tabulated in **Table 4** for classifying the masses from the normal mammographic images.

Table 4. DELU-BM-CNN—Evaluation results.

Dataset	Sensitivity %	Specificity %	F1-score %
MIAS	100	96.77	94.11
INbreast	93.75	100	96.77
CBIS-DDSM	98.13	95.42	96.33

4.2. Comparison with existing works

The results of the proposed method are compared with the state-of-art methodologies and the results are comparison in **Table 5**. The current technologies use complex methodologies for feature extraction and pathology classification whereas the proposed methodology aimed for a simplified CNN architecture that classifies mammograms lesion from the normal mammogram images with minimum hardware and time consumption.

Table 5. Performance comparison of proposed method with existing methods

Contributors	Year	Methodology	Dataset	Accuracy
Baccouche et al.[22]	2022	Stacked residual networks are used (ResNet50V2, ResNet101V2, and ResNet152V2)	CBIS-DDSM INB	95.13% 99.20%
Mahmood et al.[19]	2022	Local features are extracted as small patches and fed to ConvNet	MIAS	94.00%
Belhaj Soulami et al.[20]	2022	Capsule CNN	DDSM, CBIS-DDSM, Inbreast	96.30%
Busaleh et al.[16]	2021	An ensemble classifier based on the contextual info from the mass regions acts as a strong indicator for discriminating benign and malignant masses from MG images	INB CBIS-DDSM	100% 94.90%
Zhang et al.[14]	2020	Texton and deep CNN are fused to train the network	CBIS-DDSM INbreast + MIAS	94.30% 93.59%
Proposed model		A five layered ELU based CNN architecture	CBIS-DDSM Inbreast MIAS	96.60% 96.20% 97.40%

5. Conclusion

This research paper focus on constructing a simple and effective deep learning based breast mass classification system for distinguishing the lesions from the normal mammographic images. Three standard and referred mammographic datasets, CBIS-DDSM, INbreast and mini-MIAS are used for devising and verifying the proposed model. Train-test model validation was used with sensitivity, specificity and accuracy as model metrics. Whole mammographic images are accounted after a careful identification of image pre-processing methods that based to propose the DELU-BM-CNN model. The 3 × 3 deep convoluted with ELU based activation function were able to produce results that are promising to assist and support the radiologists and physicians in breast cancer diagnosis. It also acted as an early identification of the lesion in the breast region, thereby preventing the disease longevity which finally leads to mastectomy.

Author contributions

Conceptualization, GR; methodology, SG; software, SG; validation, SG; formal analysis, VK; investigation, SG; resources, VK; data curation, SG; writing—original draft preparation, SG; writing—review and editing, SG, VK and GR; visualization, SG and GR; supervision, VK and GR; project administration, GR. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. The Breast. 2022, 66: 15-23. doi: 10.1016/j.breast.2022.08.010
2. Tariq M, Iqbal S, Ayesha H, et al. Medical image based breast cancer diagnosis: State of the art and future directions. Expert Systems with Applications. 2021, 167: 114095. doi: 10.1016/j.eswa.2020.114095
3. Tsochatzidis L, Costaridou L, Pratikakis I. Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. Journal of Imaging. 2019, 5(3): 37. doi: 10.3390/jimaging5030037
4. Iranmakani S, Mortezazadeh T, Sajadian F, et al. A review of various modalities in breast imaging: technical aspects and clinical outcomes. Egyptian Journal of Radiology and Nuclear Medicine. 2020, 51(1). doi: 10.1186/s43055-020-00175-5
5. Harvey H, Karpati E, Khara G, et al. The Role of Deep Learning in Breast Screening. Current Breast Cancer Reports. 2019, 11(1): 17-22. doi: 10.1007/s12609-019-0301-7
6. Abdelhafiz D, Yang C, Ammar R, et al. Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinformatics. 2019, 20(S11). doi: 10.1186/s12859-019-2823-4
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017, 60(6): 84-90. doi: 10.1145/3065386
8. LeCun Y, Bengio Y. The Handbook of Brain Theory and Neural Networks. MIT Press; 1998.
9. Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. doi: 10.1109/cvpr.2014.81
10. Zhang Y, Sohn K, Villegas R, et al. Improving object detection with deep convolutional networks via Bayesian optimization and structured prediction. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. doi: 10.1109/cvpr.2015.7298621
11. Lu L, Zheng Y, Carneiro G, et al. Deep Learning and Convolutional Neural Networks for Medical Image Computing. Springer International Publishing. 2017. doi: 10.1007/978-3-319-42999-1
12. Dong S, Wang P, Abbas K. A survey on deep learning and its applications. Computer Science Review. 2021, 40: 100379. doi: 10.1016/j.cosrev.2021.100379
13. Shen L, Margolies LR, Rothstein JH, et al. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Scientific Reports. 2019, 9(1). doi: 10.1038/s41598-019-48995-4
14. Zhang Q, Li Y, Zhao G, et al. A Novel Algorithm for Breast Mass Classification in Digital Mammography Based on Feature Fusion. Journal of Healthcare Engineering. 2020, 2020: 1-11. doi: 10.1155/2020/8860011

15. Malebary SJ, Hashmi A. Automated Breast Mass Classification System Using Deep Learning and Ensemble Learning in Digital Mammogram. IEEE Access. 2021, 9: 55312-55328. doi: 10.1109/access.2021.3071297
16. Busaleh M, Hussain M, Aboalsamh HA, et al. Breast Mass Classification Using Diverse Contextual Information and Convolutional Neural Network. Biosensors. 2021, 11(11): 419. doi: 10.3390/bios11110419
17. Sun L, Wen J, Wang J, et al. Breast mass classification based on supervised contrastive learning and multi-view consistency penalty on mammography. IET Biometrics. 2022, 11(6): 588-600. doi: 10.1049/bme2.12076
18. Sannasi Chakravarthy SR, Rajaguru H. Automatic Detection and Classification of Mammograms Using Improved Extreme Learning Machine with Deep Learning. IRBM. 2022, 43(1): 49-61. doi: 10.1016/j.irbm.2020.12.004
19. Mahmood T, Li J, Pei Y, et al. Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. PLOS ONE. 2022, 17(1): e0263126. doi: 10.1371/journal.pone.0263126
20. Belhaj Soulami K, Kaabouch N, Nabil Saidi M. Breast cancer: Classification of suspicious regions in digital mammograms based on capsule network. Biomedical Signal Processing and Control. 2022, 76: 103696. doi: 10.1016/j.bspc.2022.103696
21. Sengan S, Priya V, Syed Musthafa A, et al. A fuzzy based high-resolution multi-view deep CNN for breast cancer diagnosis through SVM classifier on visual analysis. Journal of Intelligent & Fuzzy Systems. 2020, 39(6): 8573-8586. doi: 10.3233/jifs-189174
22. Baccouche A, Garcia-Zapirain B, Elmaghraby AS. An integrated framework for breast mass classification and diagnosis using stacked ensemble of residual neural networks. Scientific Reports. 2022, 12(1). doi: 10.1038/s41598-022-15632-6
23. Jiang J, Peng J, Hu C, et al. Breast cancer detection and classification in mammogram using a three-stage deep learning framework based on PAA algorithm. Artificial Intelligence in Medicine. 2022, 134: 102419. doi: 10.1016/j.artmed.2022.102419
24. Baccouche A, Garcia-Zapirain B, Zheng Y, et al. Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques. Computer Methods and Programs in Biomedicine. 2022, 221: 106884. doi: 10.1016/j.cmpb.2022.106884
25. Rajesh Kumar N, Bala Krishnan R, Manikandan G, et al. Reversible data hiding scheme using deep learning and visual cryptography for medical image communication. Journal of Electronic Imaging. 2022, 31(06). doi: 10.1117/1.jei.31.6.063028
26. Adam R, Dell'Aquila K, Hodges L, et al. Deep learning applications to breast cancer detection by magnetic resonance imaging: a literature review. Breast Cancer Research. 2023, 25(1). doi: 10.1186/s13058-023-01687-4
27. Balkenende L, Teuwen J, Mann RM. Application of Deep Learning in Breast Cancer Imaging. Seminars in Nuclear Medicine. 2022, 52(5): 584-596. doi: 10.1053/j.semnuclmed.2022.02.003
28. Civilibal S, Cevik KK, Bozkurt A. A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. Expert Systems with Applications. 2023, 212: 118774. doi: 10.1016/j.eswa.2022.118774
29. Wakili MA, Shehu HA, Sharif MdH, et al. Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning. Computational Intelligence and Neuroscience. 2022, 2022: 1-31. doi: 10.1155/2022/8904768
30. Ahmed M, Islam MdR. A combined feature-vector based multiple instance learning convolutional neural network in breast cancer classification from histopathological images. Biomedical Signal Processing and Control. 2023, 84: 104775. doi: 10.1016/j.bspc.2023.104775
31. Sirjani N, Ghelich Oghli M, Kazem Tarzamni M, et al. A novel deep learning model for breast lesion classification using ultrasound Images: A multicenter data evaluation. Physica Medica. 2023, 107: 102560. doi: 10.1016/j.ejmp.2023.102560
32. Rahman MM, Khan MSI, Babu HMdH. BreastMultiNet: A multi-scale feature fusion method using deep neural network to detect breast cancer. Array. 2022, 16: 100256. doi: 10.1016/j.array.2022.100256
33. Çayır S, Solmaz G, Kusetogullari H, et al. MITNET: a novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue. Neural Computing and Applications. 2022, 34(20): 17837-17851. doi: 10.1007/s00521-022-07441-9
34. Kadry S, Crespo RG, Herrera-Viedma E, et al. Classification of Breast Thermal Images into Healthy/Cancer Group Using Pre-Trained Deep Learning Schemes. Procedia Computer Science. 2023, 218: 24-34. doi: 10.1016/j.procs.2022.12.398
35. Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific Data. 2017, 4(1). doi: 10.1038/sdata.2017.177
36. Moreira IC, Amaral I, Domingues I. INbreast. Academic Radiology. 2012, 19(2): 236-248. doi: 10.1016/j.acra.2011.09.014
37. Suckling J, Parker J, Dance DR, et al. Mammographic Image Analysis Society (MIAS). 1994.. 375–378.
38. Patterson J, Gibson A. Deep Learning. O'Reilly Media, Inc; 2017.
39. Zaccone G, Karim MR. Deep Learning with TensorFlow: Explore Neural Networks and Build Intelligent Systems with Python. Packt Publishing Limited; 2018.
40. Graupe D. Deep Learning Neural Networks: Design and Case Studies. World Scientific; 2020.

41. Wang SH, Satapathy SC, Xie MX, et al. ELUCNN for explainable COVID-19 diagnosis. Soft Computing. 2023. doi: 10.1007/s00500-023-07813-w
42. Sergey I, Christian S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Available online: https://proceedings.mlr.press/v37/ioffe15.html (accessed on 20 July 2023).
43. Scherer D, Müller A, Behnke S. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. Lecture Notes in Computer Science. 2010, 92-101. doi: 10.1007/978-3-642-15825-4_10
44. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. 2014; 15: 1929–1958.
45. Lin M, Chen Q, Yan S. Network in Network. Presented at the International Conference on Learning Representations. 2014.