



Dense Tissue Pattern Characterization Using Deep Neural Network

Indrajeet Kumar¹ · Abhishek Kumar² · V D Ambeth Kumar³ · Ramani Kannan⁴ · Vrince Vimal¹ · Kamred Udham Singh⁸ · Mufti Mahmud^{5,6,7}

Received: 31 August 2020 / Accepted: 21 November 2021
© The Author(s) 2022

Abstract

Breast tumors are from the common infections among women around the world. Classifying the various types of breast tumors contribute to treating breast tumors more efficiently. However, this classification task is often hindered by dense tissue patterns captured in mammograms. The present study has been proposed a dense tissue pattern characterization framework using deep neural network. A total of 322 mammograms belonging to the mini-MIAS dataset and 4880 mammograms from DDSM dataset have been taken, and an ROI of fixed size 224×224 pixels from each mammogram has been extracted. In this work, tedious experimentation has been executed using different combinations of training and testing sets using different activation function with *AlexNet*, *ResNet-18* model. Data augmentation has been used to create a similar type of virtual image for proper training of the DL model. After that, the testing set is applied on the trained model to validate the proposed model. During experiments, four different activation functions ‘sigmoid’, ‘tanh’, ‘ReLU’, and ‘leakyReLU’ are used, and the outcome for each function has been reported. It has been found that activation function ‘ReLU’ perform always outstanding with respect to others. For each experiment, classification accuracy and kappa coefficient have been computed. The obtained accuracy and kappa value for MIAS dataset using *ResNet-18* model is 91.3% and 0.803, respectively. For DDSM dataset, the accuracy of 92.3% and kappa coefficient value of 0.846 are achieved. After the combination of both dataset images, the achieved accuracy is 91.9%, and kappa coefficient value is 0.839 using *ResNet-18* model. Finally, it has been concluded that the *ResNet-18* model and *ReLU* activation function yield outstanding performance for the task.

Keywords Deep neural network · Dense tissue characterization · AlexNet · ResNet-18 · Kappa coefficient

Introduction

✉ Mufti Mahmud
muftimahmud@gmail.com; mufti.mahmud@ntu.ac.uk

Ramani Kannan
ramani.kannan@utp.edu.my

¹ Graphic Era Hill University, Dehradun, India

² JAIN (Deemed To Be University), Bangalore, India

³ Panimalar Engineering College, Anna University, Chennai, India

⁴ Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia

⁵ Department of Computer Science, Nottingham Trent University, Nottingham NG11 8NS, UK

⁶ Medical Technologies Innovation Facility, Nottingham Trent University, Nottingham NG11 8NS, UK

⁷ Computing and Informatics Research Centre, Nottingham Trent University, Nottingham NG11 8NS, UK

⁸ Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan

The statistical report produced by the American cancer society shows that among twelve women one is the possibility of rising breast-related lesions. Among the various breast lesions, breast tumor is one of the repeatedly occurring infections among women in every region of the world [1–5]. The major attribute to developing breast cancer is breast tissue pattern density, body weight, age, genetic history of breast lesions, type of radiation therapy, alcohol consumption, etc. [4, 5]. The relation between the development of cancer and risk factor is shown in Fig. 1.

After the study of past statistics and Fig. 1, it has been seen that the tissue density plays a crucial role in developing breast cancer. Breast tissue density is the ratio between fibro-glandular tissues to the fatty tissues. According to the Breast Imaging-Reporting and Data System (BIRADS) notation [6], breast tissue characterization is further classified into four different levels as per the tissue density availability

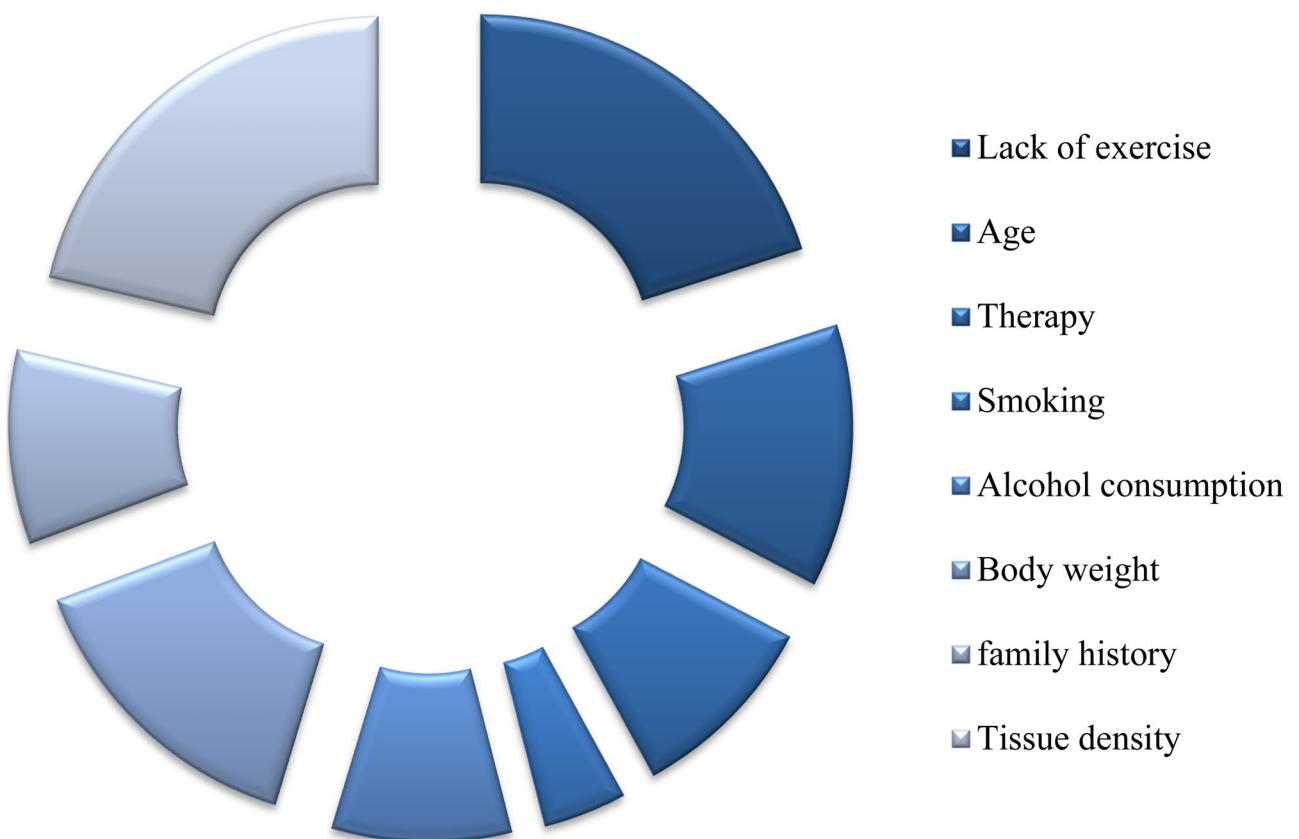


Fig. 1 Relation between the development of cancer and risk factor [3]

in mammograms. The four class of tissue characterization [7–13] is as BIRADS-I which covers completely fatty tissue, i.e., ratio is less than 25%, BIRADS-II which covers the ratio between 25 and 50%, BIRADS-III which covers the ratio between 51 and 75% and BIRADS-IV includes the ratio between 76 and 100%.

In most of the cases, BIRADS-I and BIRADS-II combined together as fatty breast tissue class and BIRADS-III and BIRADS-IV combined together as dense breast tissue class. The sample image of each ACR-BIRADS class taken from the Digital Database for Screening Mammography (DDSM) dataset [14] is shown in Fig. 2.

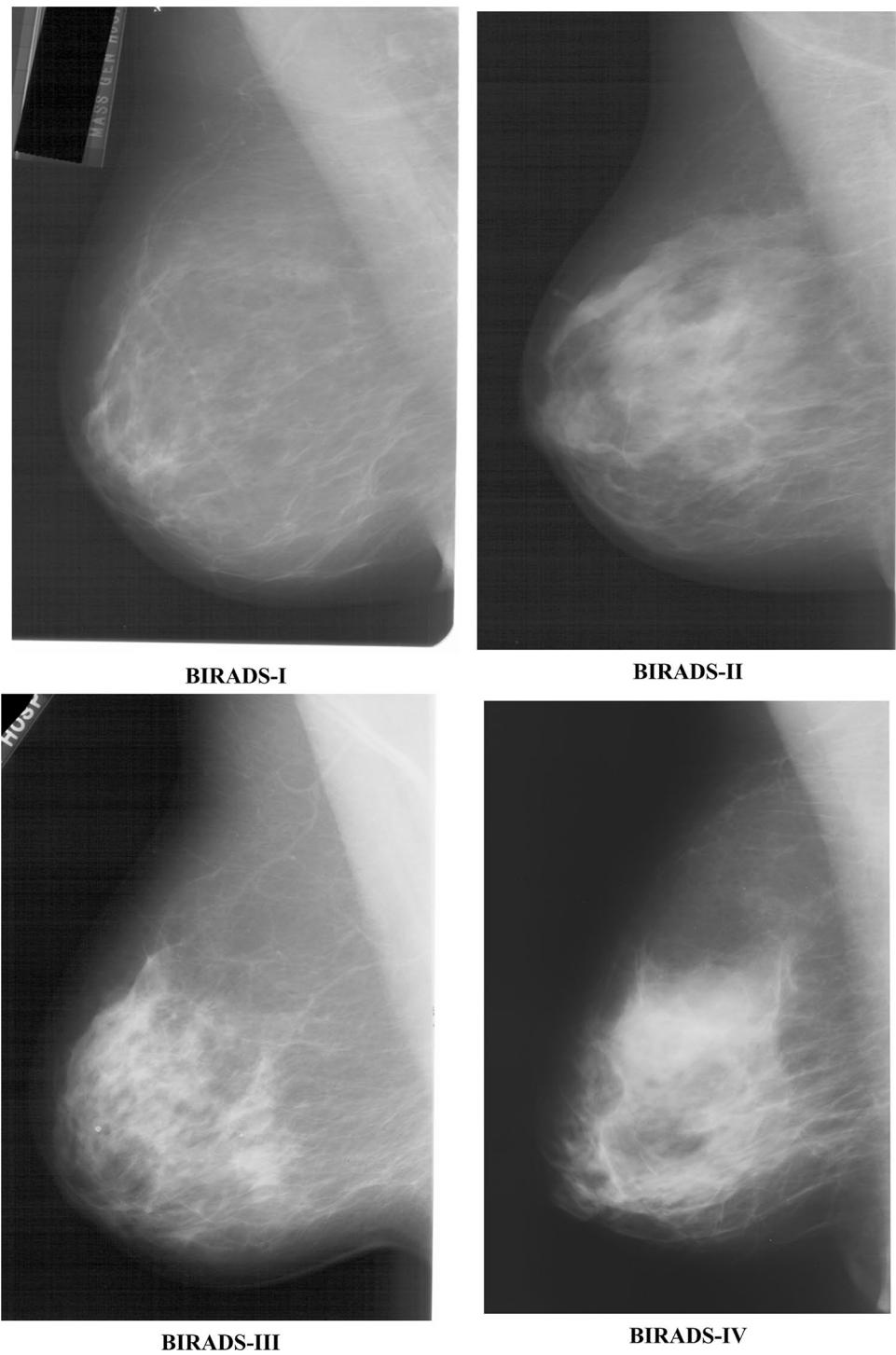
The characterization of breast tissue pattern density is important because of (i) adequate scheduling of treatment related to breast lesions, and (ii) it has been also found that lesions are masked behind the dense tissue so that the treatment is not going in proper directions, for such type of cases expert may take the opinion of secondary imaging modalities. It has been also observed that the early predictions of increased breast tissue density reduce the development of cancerous cell and improve the treatment adequacy. For characterization of breast tissue pattern, machine learning and deep learning models have played a crucial role [7, 8]. Therefore, the growth of

efficient computer-aided classification (CAC) system is a challenging task for the research community. Regarding this problem, so much work based on the machine learning concept has been already completed, but these models suffered from either complexity or less accuracy [9, 10, 15–17].

Based on the challenges obtained from the previous studies, a dense tissue pattern classification framework has been proposed using a deep neural network model. The major contributions of the proposed work as follows:

1. Deep neural network-based dense tissue pattern classification framework is proposed for prediction of breast tissue pattern. This framework is suitable for clinical practice as a secondary opinion tool for prediction of breast density.
2. The extensive experiments have been carried out on 322 mammograms taken from mini-MIAS dataset and 4880 mammograms taken from DDSM dataset.
3. The present work is different from past studies in terms of complexity because it is based on the region of interest. In which a region of interest (ROI) is cropped from the center region of the mammograms and passed to the framework, and the decision of the system is generated.

Fig. 2 Sample image of ACR-BIRADS class taken from DDSM dataset [14]



4. For attaining better result and accuracy, different activation functions like *ReLU*, *Sigmoid*, *Tanh*, and *Leaky ReLU* are investigated on the same dataset with *AlexNet* and *ResNet-18* deep learning model.
5. The results achieved by proposed the model are compared with state-of-art and results achieved in our model outperform existing works by achieving an accuracy of 92.3%, and the kappa coefficient value of 0.846 for

characterization of dense tissue using mammographic images.

The rest of the paper is ordered as a literature review section as "Related Studies" section, in which major findings and limitations of past studies have been illustrated. In next "Materials and Methods" section, material and methods are described in which data preparation protocols, ROI

extraction, data augmentation and dataset bifurcation are discussed. In "Experiments and Results" proposed model, experiments and results are discussed, and finally, the complete work has been summarized in the "Conclusion" section as a conclusion.

Related Studies

After a deep study of literature related to dense tissue pattern characterization, it has been found that the characterization of tissue can be done in two methods: (i) segmented tissue-based characterization (ST) and (ii) region of interest tissue-based characterization (ROIT). ROI-based method is simpler than segmentation-based method because segmentation-based approach needs additional preprocessing steps, but in case of ROI-based approach, a fixed size ROI has to be cropped and passed to the model without any additional preprocessing steps. Therefore, the present work is based on an ROI-based method for dense tissue pattern characterization.

In past studies, so many studies for breast tissue pattern characterization using ST and ROIT have been done [7, 8, 18, 19]. For these methodologies, machine learning (ML) and deep learning (DL) are prominently used [20, 21]. In ML, support vector machine (SVM), artificial neural network (ANN), k-nearest neighbor (kNN), probabilistic neural network (PNN), smooth SVM (SSVM), etc. [7, 8, 18] are applied with spatial domain feature extraction, transform domain-based feature extraction, and law's texture energy features [18]. The past studies show some promising results with ML models, but it suffers if the number of the input sample is large. For such types of problems, DL models yield better performance. Due to that, DL model is getting more attention in the last few years.

With time, several deep learning model-based computer aided diagnostic systems have been developed for the detection or classification of tumors. In the study [22], convolutional neural network (CNN) is used for lesion classification on 736 mammograms and attained the accuracy of 82.6%. In this model, computed features are passed to the classification module. The experiment performed by Qiu et al. in study [23] used a DL model having eight layers for classification task. The proposed eight-layer DL model can extract the features automatically, and extracted features are used for feature classification. The designing of the DL model is a hectic task; therefore, pre-trained models can be also used for the classification problem [24]. The study [25] used a pre-trained model for benign and malignant cancer classification on DDSM and InBreast dataset [26]. The obtained ROC curve shows the accuracy of 90.0%. For tissue density classification, the study [15] shows the accuracy of 83.6% for two-class classification on MIAS dataset using pre-trained *VGG16* model. The breast tissue density can be

also classified into four classes as reported in previously published work [27]. In this work, the DL model is used and validated on 200,000 mammographic samples. The achieved accuracy for the proposed model is 84.2%. The obtained result shows the promising results for four class classification tasks. The similar type of task has been performed in [16] using the *Inception V3* model and achieves the accuracy of 84.4% on 3813 number of samples.

The DL model is not only limited to the classification task. Such a type of model can be also used for the feature extraction, and the computed values are passed to the machine learning model called transfer learning [28–30]. The study [28] shows the application of transfer learning and achieved outstanding results. In the proposed work, DL model is used for feature extraction, and extracted features are passed to the ML algorithm for classification task. The proposed model is validated on self-collected mammograms. The concept of transfer learning is also used in [29] where 22000 samples are used, and the CNN model is used for feature extraction. Finally, the reported accuracy for this study is 92.6% for testing and 94.2% for training.

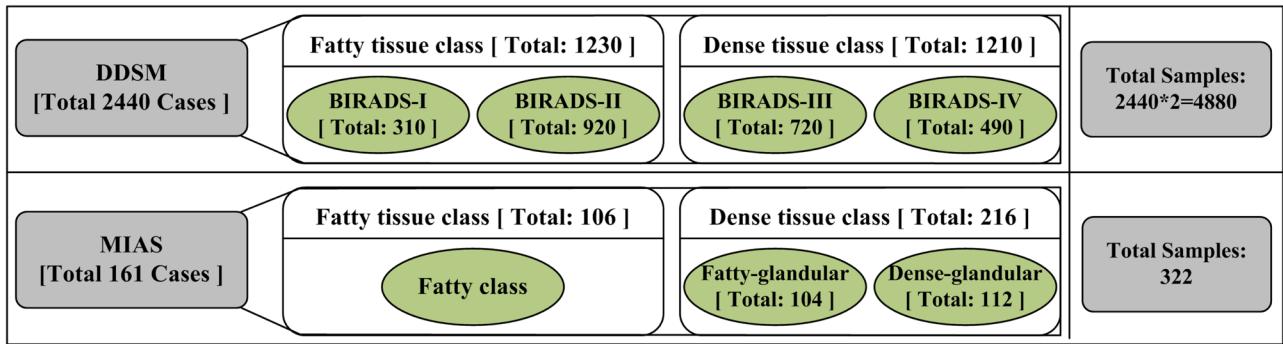
Keeping in the view of previously published results, it has been conceded that most of work has been concentrated around complete mammogram processing. In past studies, it has been also found that the density of tissue pattern is highest at behind the nipple and center location of breast. The same fact has been experimentally proved by Li [31]. It is also noticed that the same facts have been also observed by radiologist and medical experts [7, 8]. Therefore, ROIT concept is used to design a dense tissue pattern characterization. In the proposed work, a fixed size of ROI is extracted from the mammograms and then it is passed to the DL model.

In this work, two pre-trained DL models *AlexNet* and *ResNet-18* [19, 25, 32] have been used to develop the proposed system. For each model, four activation functions '*sigmoid*', '*tanh*', '*ReLU*', and '*leakyReLU*' [33] are used for activation of neurons. The training and testing image samples are taken from mini-MIAS and DDSM database. The image augmentation has been performed to increase the number of samples. After that, training and testing of DL model is performed and obtained results are evaluated in terms of accuracy, misclassification accuracy, and kappa coefficient.

Materials and Methods

Dataset Preparation

In this work, two scientific datasets (mini-MIAS, DDSM) are used. Both datasets are freely available for research purposes. The various experiments have been performed on firstly individual dataset then the combination of both

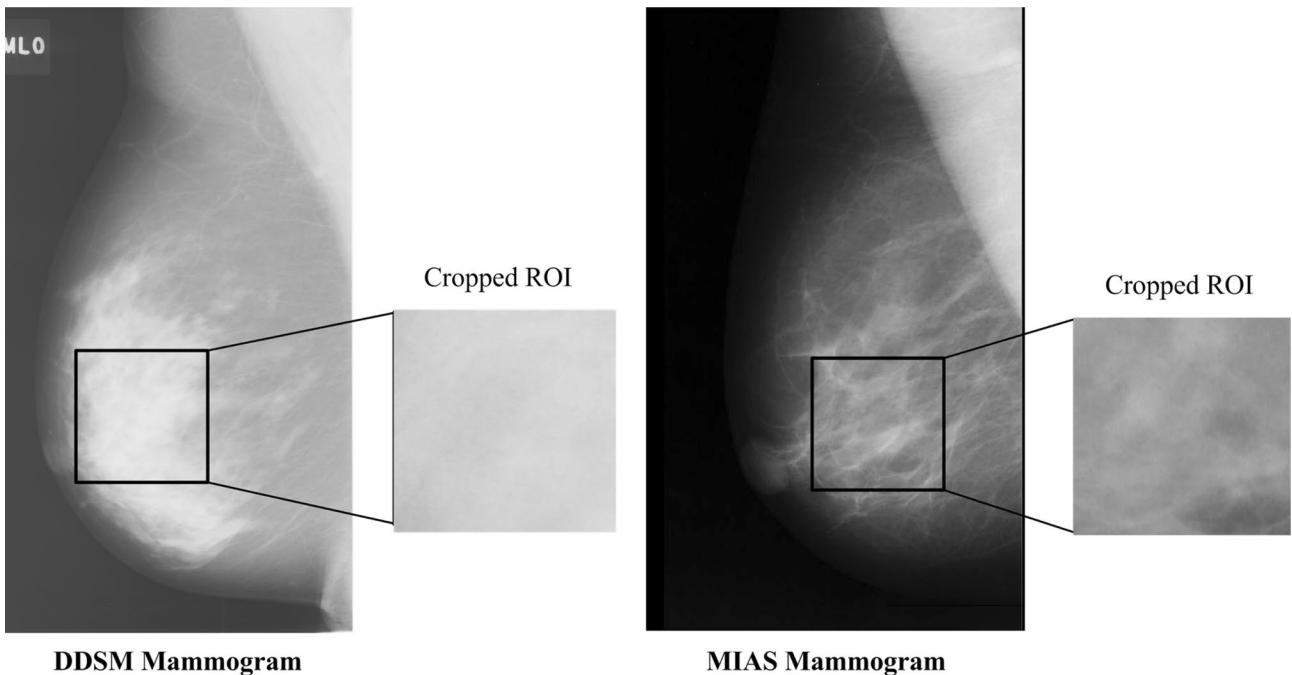
**Fig. 3** Dataset preparation

dataset and obtained findings has been reported in results sections.

The *mini-MIAS* [34] consists of 322 mammograms of 161 patients. Each mammogram of the dataset is digitized at 50-micron pixels. The density label of each mammogram is labeled by three expert radiologists as fatty tissue class, fatty-glandular tissue class, and dense-glandular class. For two class tissue pattern classification, fatty tissue class is considered as class 1, i.e., fatty whereas fatty-glandular and dense-glandular are treated as class 2, i.e., dense class. The total number of samples in the fatty class is 106 mammograms (106 samples ∈ fatty tissue classes), and the number of samples in dense class is 216 mammograms (104 samples ∈ fatty-glandular and 112 samples ∈ dense-glandular class).

In *DDSM* dataset [14], 10000 multi-view (CC view and MLO view) mammograms of 2500 patients are available

of three classes as benign, malignant, and normal. Each mammogram of the dataset is digitized at 42 to 50 microns. Each study includes two projections (MLO and CC view) of each breast, along with essential patient information like patient age, tissue pattern density rating, rating for lesions, and description of architectural distortion. The density label of each mammogram is categorized into four classes, i.e., BIRADS-I, BIRADS-II, BIRADS-III, and BIRADS-IV. To attain the objective of proposed work, a total sample of 4880 mammographic images are used as a fatty class and dense class. The fatty class comprised of 2460 mammograms as 620 samples ∈ BIRADS-I and 1840samples ∈ BIRADS-II and dense class comprised of 2420 mammograms as 1440 samples ∈ BIRADS-III and 980 samples ∈ BIRADS-IV class of MLO view. The brief detail of dataset preparation is given in Fig. 3 for MIAS and DDSM dataset.

**Fig. 4** Steps for ROI extraction

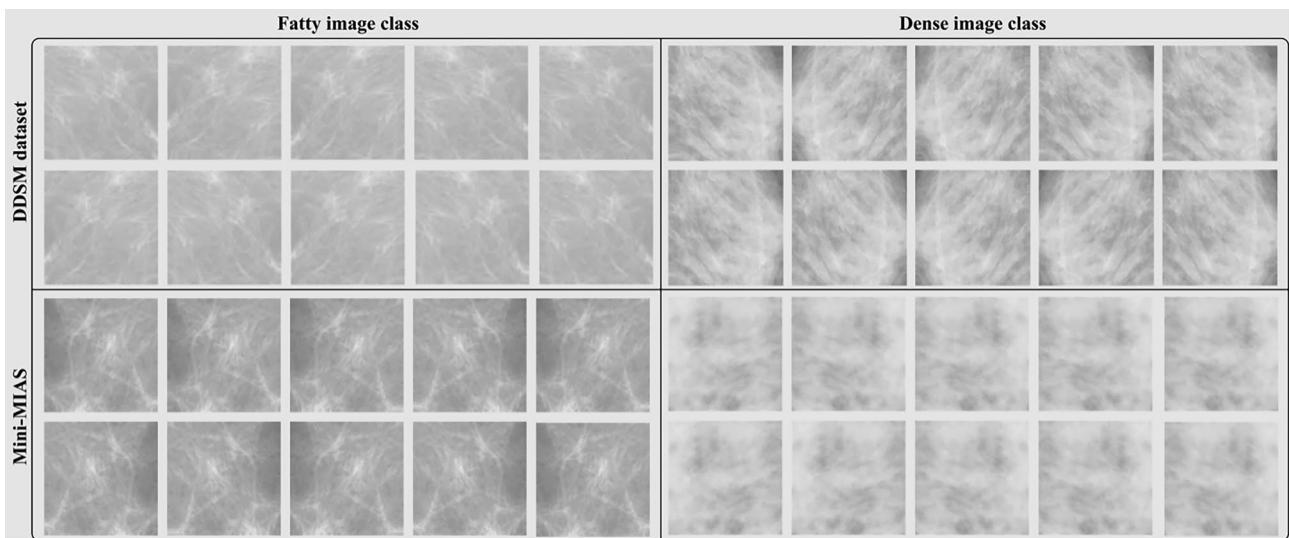


Fig. 5 Augmented image of DDSM and MIAS dataset based on fatty and dense image class

ROI Extraction

From the past studies, it has been observed that the maximum tissue pattern density found at the center of mammograms just behind the nipple [7, 8, 18, 31]. After the concern of an expert radiologist, they had also concluded a similar fact. Therefore, an ROI of size 224×224 pixels has been cropped from the center of each mammogram for the entire experiments. For mini-MIAS dataset, a total of 322 ROIs have been cropped, and from DDSM dataset, a total of 4880 ROIs have been cropped. The steps for ROI extraction are shown in Fig. 4.

Data Augmentation

It is well known that the performance of deep learning models depends on the amount of training dataset. In case of fewer amounts of data, data augmentation is used for generating a large amount of virtual samples from the available samples [35–37]. The virtual images are generated with the help of angle rotation of 5° with height, width, shear, and zoom range of value 0. In the case of DDSM mammograms, ten virtual samples are generated using one sample

with abovementioned parameters, and a set of twenty virtual images is generated using mini-MIAS mammograms by using the above mentioned parameters. After the augmentation, overall mammographic samples from the DDSM dataset are 48800 and 6440 samples of mini-MIAS mammographic images are available. The sample image of the augmented dataset is shown in Fig. 5.

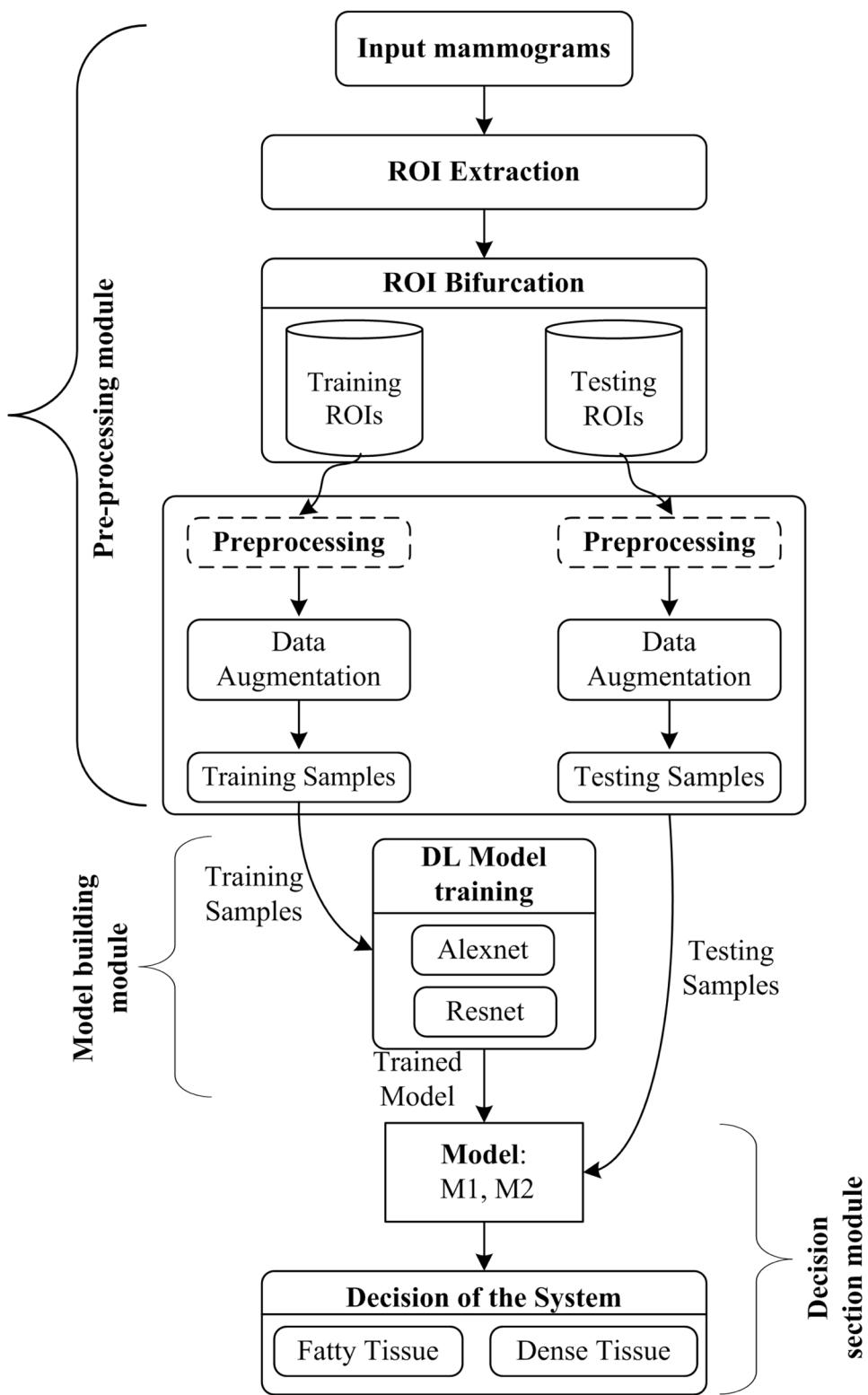
Dataset Bifurcation

The complete set of samples further bifurcated into training and testing samples in a balanced and unbalanced manner. In a balanced manner, bifurcation with 50:50 ratio is used for training and testing samples. According to this, 24400 samples of DDSM dataset and 3220 samples of the mini-MIAS dataset are used as training and testing dataset. Similarly, for unbalanced bifurcation of samples 70:30 ratio is maintained for training and testing samples. Thus 34160 samples of DDSM dataset and 4508 samples of the mini-MIAS dataset are used as a training set and 14640 samples of DDSM dataset 1932 samples of the mini-MIAS dataset are used as a testing set. The brief description of dataset preparation and bifurcation is given in Table 1.

Table 1 Brief description of dataset preparation and bifurcation

		DDSM dataset	Mini-MIAS
Total cases		4480	322
Total ROIs		4880	322
Augmented sample		$4880 \times 10 = 48800$ ROIs	$322 \times 20 = 6440$ ROIs
Dataset bifurcation	Balanced bifurcation	Training set: 24400	Training set: 3220
		Testing set: 24400	Testing set: 3220
Unbalanced bifurcation	Balanced bifurcation	Training set: 34160	Training set: 4508
		Testing set: 14640	Testing set: 1932

Fig. 6 Proposed workflow diagram for dense tissue pattern characterization using deep learning model



Proposed Work

The proposed workflow chart is shown in Fig. 6. The proposed work is divided into three main levels as a pre-processing module, model building module, and decision section. In the

preprocessing section, dataset preparation, ROI extraction, data augmentation, and ROIs bifurcation are performed, and in model building section, DL-based model is trained using the training set, and trained model is used to predict the testing samples of testing dataset described in decision section.

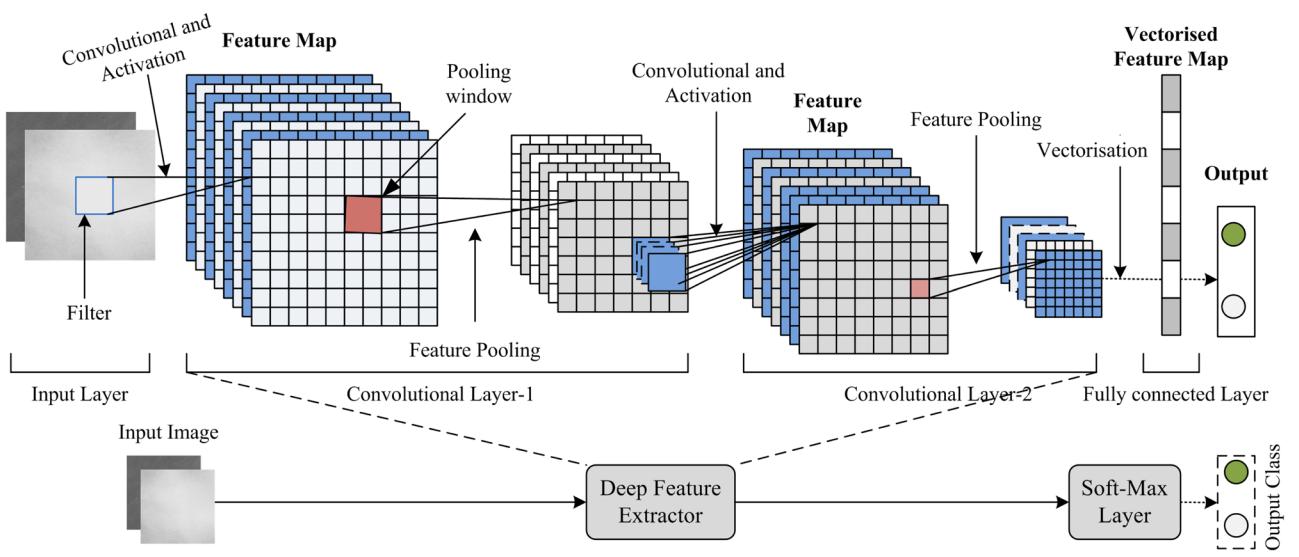


Fig. 7 Architecture of CNN model as a classifier

Deep Learning Model

In the recent trends of artificial intelligence, deep learning (DL) model plays a significant role in the development of computer-assisted framework [25] and different domains of research [38]. DL is the subset of ML which learns underlying features from data using neural a network. It is well known that the ML-based framework performance is degraded with a large amount of data, but DL models show promising results on the large amount of dataset. Another important limitation of the ML algorithm is learning from

hand-engineered features, which are time consuming, brittle, and non-scalable whereas in case of the DL model tries to learn high-level features from the data itself. The DL model is similar to the neural network having multiple hidden layers, convolution layers, pooling layers, fully connected layers, activation functions, etc. Some popular pre-trained neural network architectures like CNN and recurrent neural network (RNN) are suitable for classification and object detection type problems [19, 33]. The generalized architecture of the DL model (CNN) as a classifier is shown in Fig. 7.

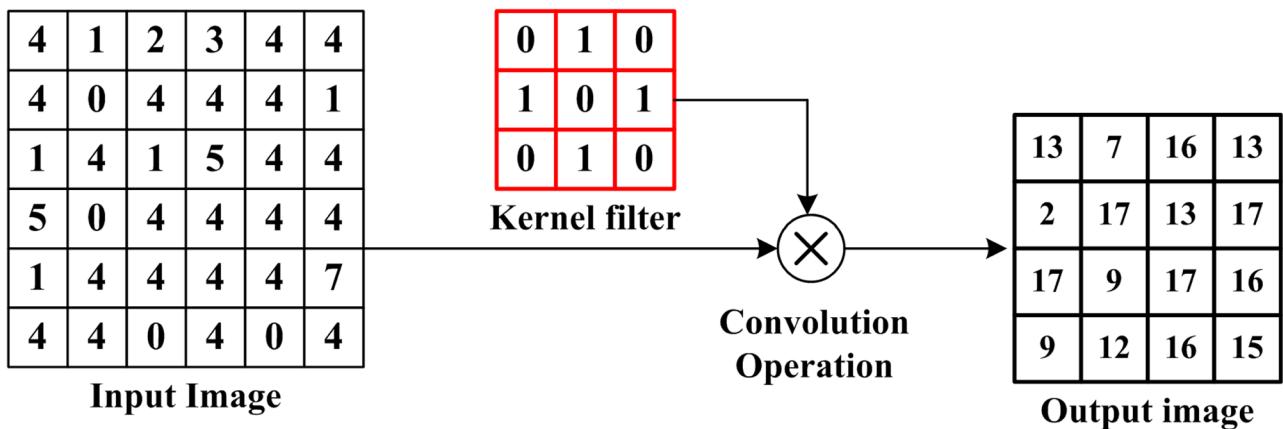
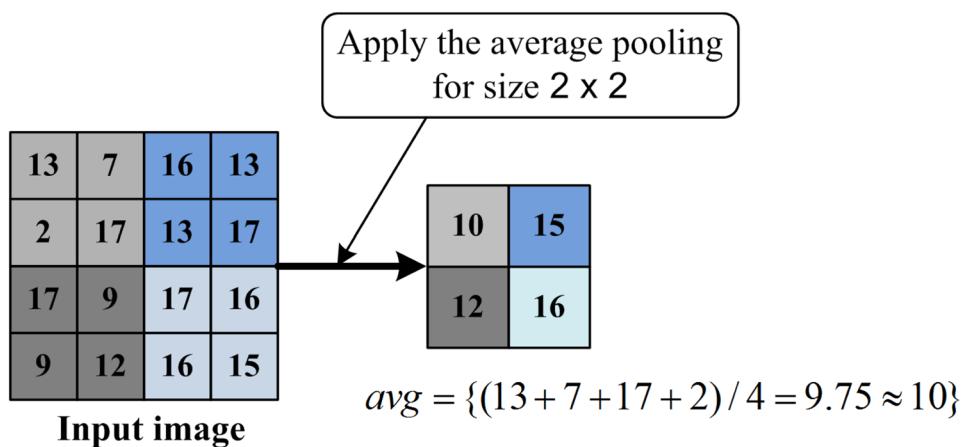


Fig. 8 Convolution operation

Fig. 9 Average pooling operation at pooling layer



Convolutional Layer

In CNN, the convolution layer is used as a feature extractor layer of the input image. To compute the feature, the input image is convolved with the weight matrix of the convolutional layer [39]. The output of every neuron is obtained using dot matrix multiplication between the weight matrix of the convolutional layer and part of the input image. The explanation of the convolution operation is shown in Fig. 8.

The size of the output image is defined as given in Eq. (1) for input image $I(W,H)$.

$$\begin{aligned} \text{outputsiz}(W) &= \{(W - F + 2P)/(S + 1)\} \\ \text{outputsiz}(H) &= \{(H - F + 2P)/(S + 1)\} \end{aligned} \quad (1)$$

where W and H are the width and height of the image. F is used for kernel filter, P is pooling function, and S is used for stride value.

Activation Layer

In this layer, a nonlinear function, known as activation function, is applied to the input matrix and performs the

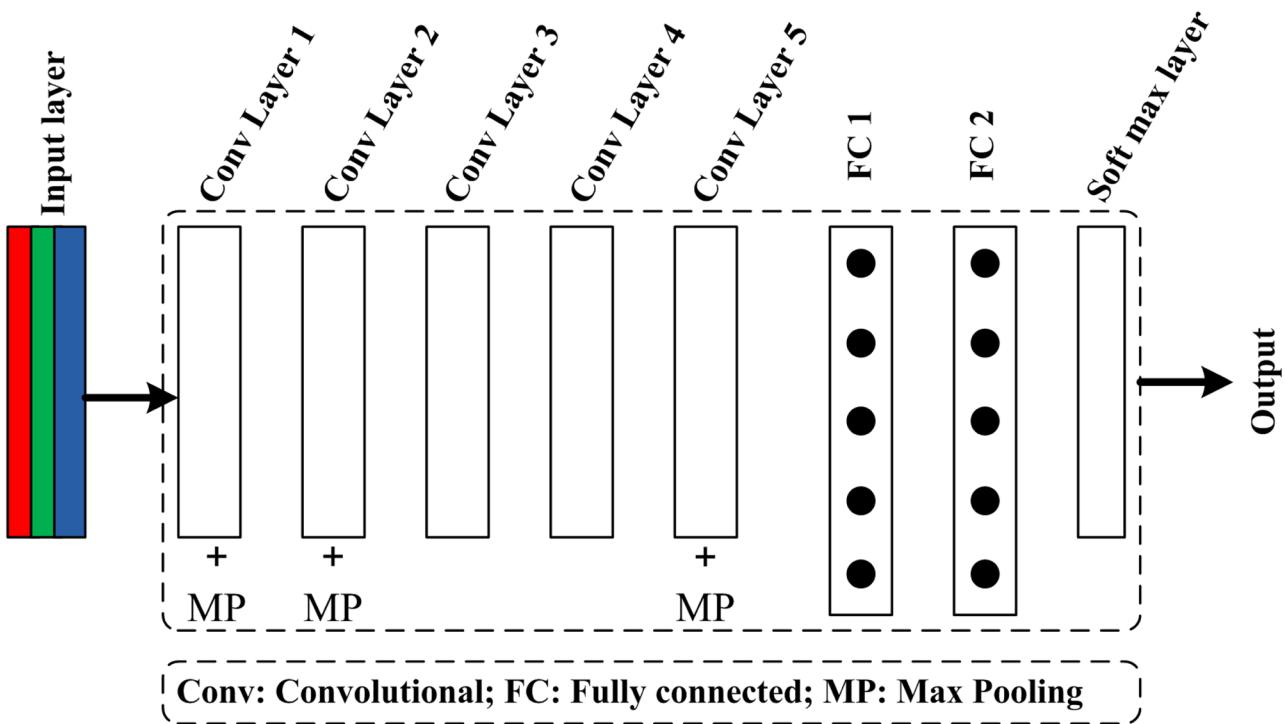


Fig. 10 Architecture of ‘AlexNet’ model

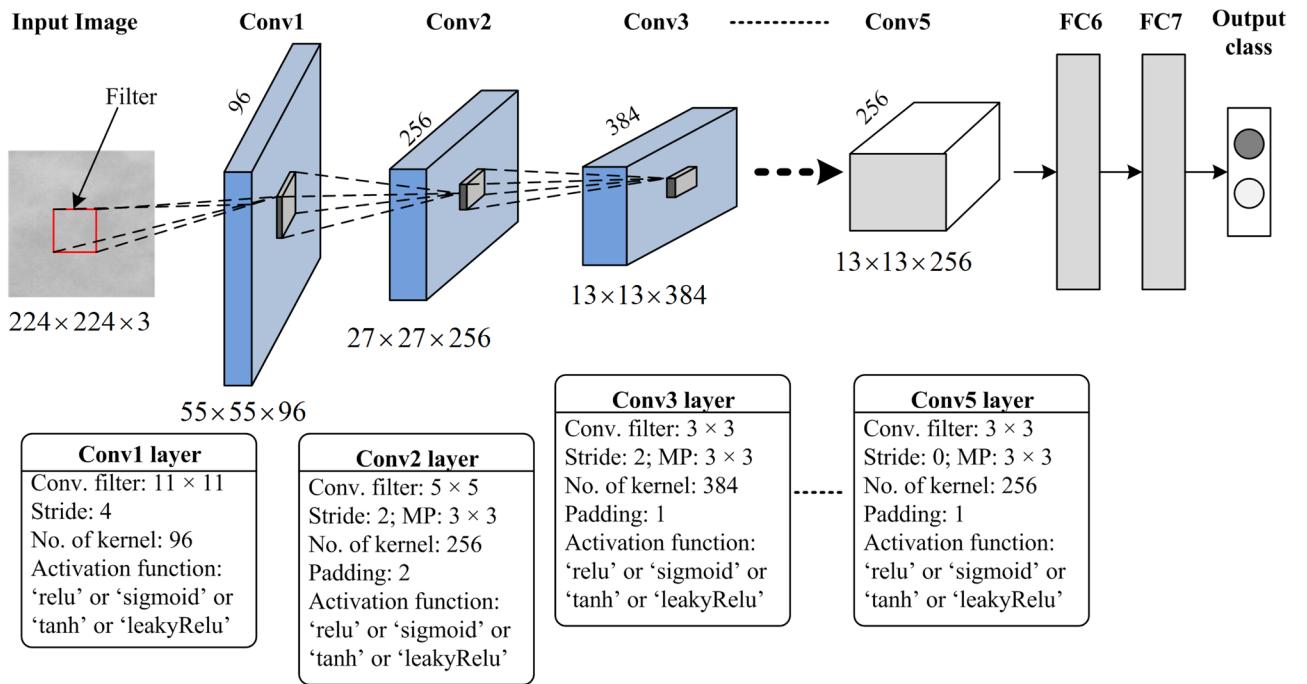


Fig. 11 Experimental structure of *AlexNet* with input sample taken from the used dataset

complex operation. The resultant value of the activation function is used to decide whether a neuron is active or not. The most frequently used activation function in DL

models is 'sigmoid', 'tanh', 'ReLU', 'leakyReLU', and so many variants of ReLU [33].

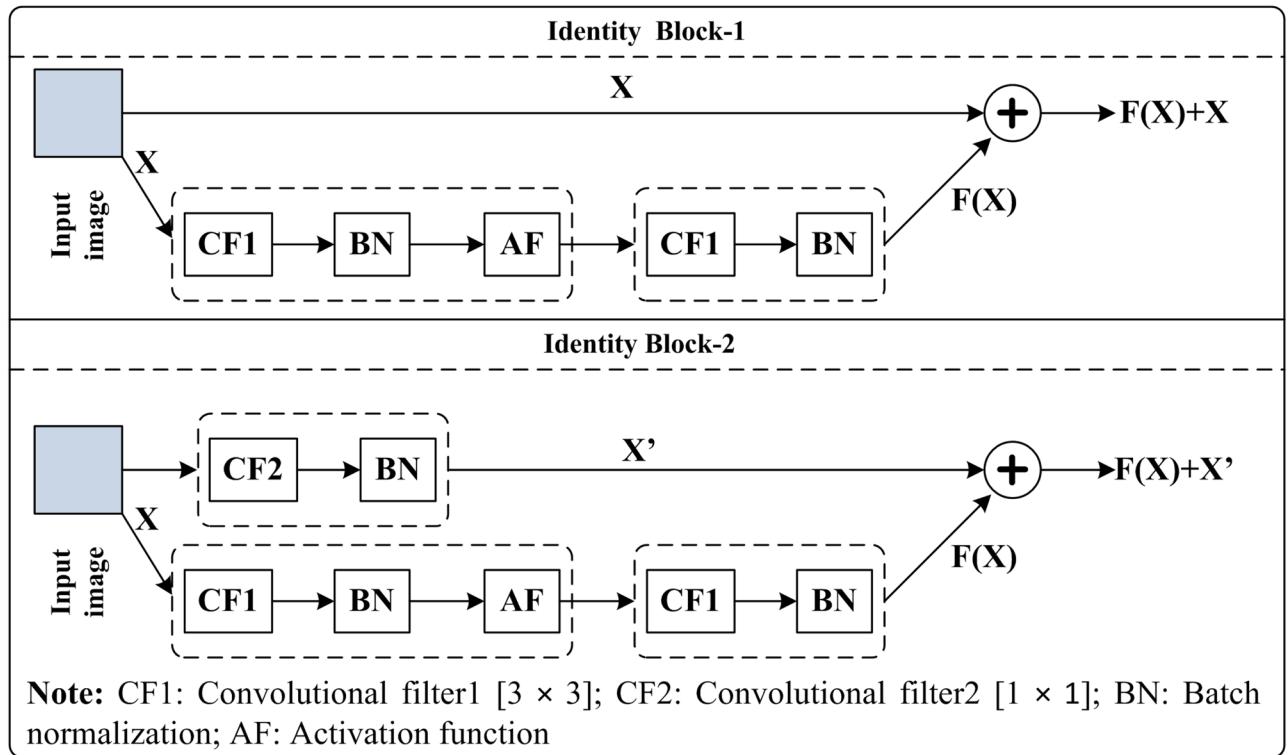
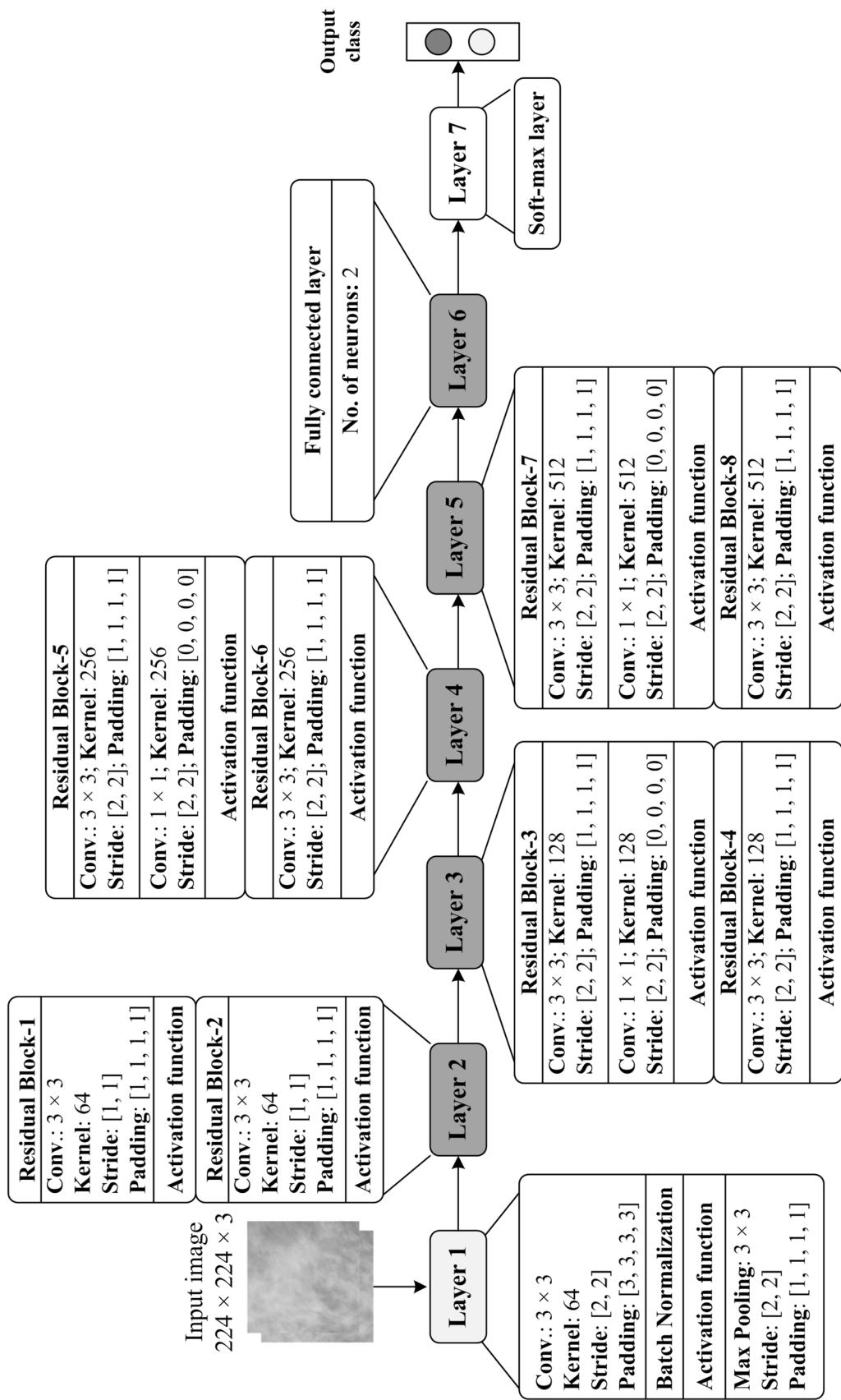


Fig. 12 Residual Block or identity block

**Fig. 13** ResNet-18 architecture

Pooling Layer

This layer is used to reduce the spatial dimension of the input image so that fewer operations have to be performed at the next layer. The most frequently used pooling layer operations are average pooling and max pooling. The example of average pooling operation is shown in Fig. 9.

In this work, two different pre-trained DL models (AlexNet, ResNet-18) are used for dense tissue pattern characterization. The brief description of each model is given below.

AlexNet

The first deep learning model ‘AlexNet’ is developed by Alex Krizhevsky [40] which makes a huge revolution in ML and AI research field. The model consists of five convolutional layers, two fully connected layers, and a softmax layer. The architecture of ‘AlexNet’ model is given in Fig. 10.

Each convolutional layer uses a convolutional filter followed by a non-linear activation function. Out of these five

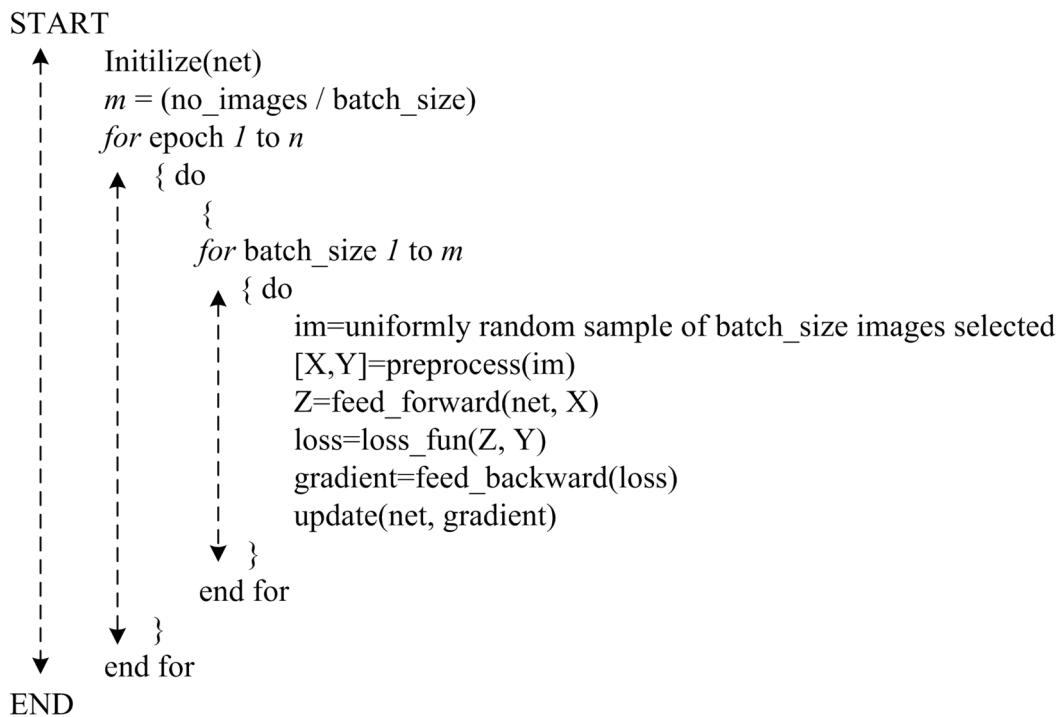
convolutional layers, three layers are followed by a pooling layer as shown in Fig. 9. The input size of the AlexNet model is $224 \times 224 \times 3$, so the ROI of every mammogram has been resized to $224 \times 224 \times 3$. A convolution filter size of 11×11 with stride size of 4 is applied on input images of size $224 \times 224 \times 3$. Therefore, the size of the conv2 layer is $55 \times 55 \times 96$ using $\{(224 - 11/4) + 1\} = 55$ and a kernel size of 96 is generated. After that, a max-pooling filter of size 3×3 with a stride rate of 2 is applied so that the size of the next layer becomes $27 \times 27 \times 256$.

The similar calculation has been done for others layers, and experimental structure of AlexNet with input sample taken from the used dataset is shown in Fig. 11.

ResNet

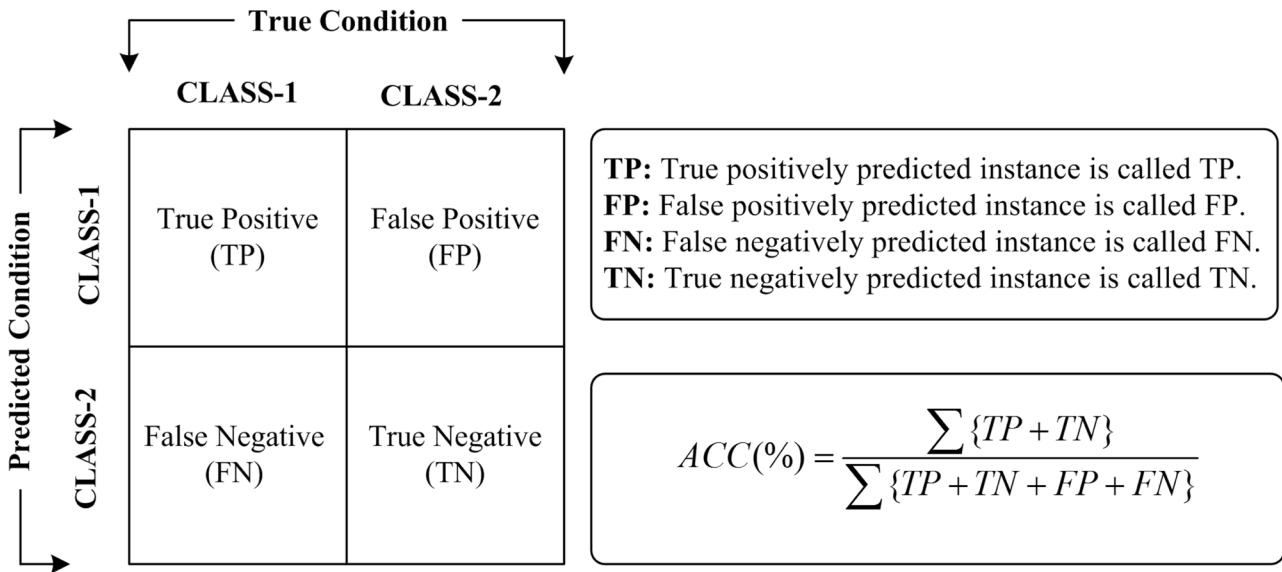
The *ResNet* model is the same as the GoogleNet model having seven numbers of layers [41, 42]. Each layer consists of an identity block. The structure of identity block-1 and identity block-2 is shown in Fig. 12. Each block consists of a convolutional filter, batch normalization, and non-linear activation function. The convolutional filter size of block-1 is 3×3 , and 1×1 convolutional filter size for

Algorithm 1: Train a deep neural network model with defined batch_size for SGD approach



NOTE: SGD stands for stochastic gradient descent method

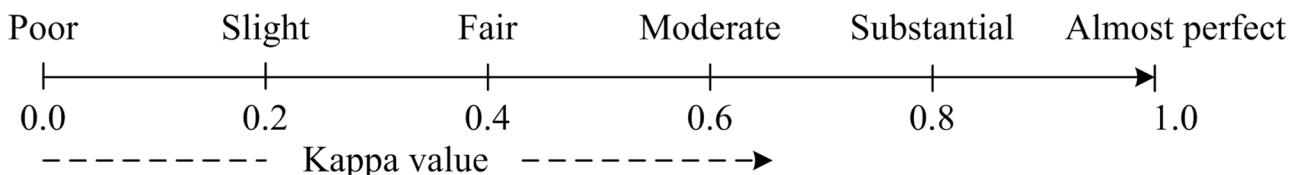
Algorithm 1 Train a deep neural network model with defined batch size for SGD approach

**Fig. 14** Confusion matrix

block-2 is used. The resultant vector of residual block-1 is directly added, i.e., element wise addition with input vector so that the extracted features are preserved and the DL model tries to learn maximum features from low to high level. In the same manner, block-2 performs the same set of operations with a minor difference as shown in Fig. 12. In this work, ResNet-18 model is used for dense tissue pattern characterization.

The complete architecture of ResNet-18 is shown in Fig. 13. The ResNet-18 [19] is composed of 8 residual

blocks. Among these, residual block-1, residual block-2, residual block-4, residual block-6, and residual block-8 are composed of identity block-1, and residual block-3, residual block-, and residual block-7 are composed of identity block-2. The size of the $224 \times 224 \times 3$ pixels input image is applied to the ResNet model. At layer one convolution filter of size 3×3 , a total number of kernels 64, stride of [2, 2] with padding [3, 3, 3, 3] is applied. This operation is followed by batch normalization and activation function, and finally, max pooling, stride, and padding are applied on the normalized



Kappa value	Significance
<0	Poor
0.01-0.20	Slightly significant
0.21-0.40	Fairly significant
0.41-0.60	Moderately significant
0.61-0.80	Substantially significant/Acceptable
0.81-1.0	Almost perfect

Fig. 15 Relation between kappa coefficient and the significance of the system

vector. The resultant vector of layer one is passed to layer two, which comprises identity block 1. The set of operations performed at layer 2 is mentioned in Fig. 11. Layer 3 is divided into two identity blocks as residual block 3 and block 4. Residual block 3 consists of identity block-1 and residual block 4 consists of identity block-2. In the same manner, the remaining layer has consisted, and details of each layer are shown in Fig. 12. To train the DL model, stochastic gradient descent algorithm is used.

Stochastic Gradient Descent

In past studies, it has been found that stochastic gradient descent (SGD) is frequently used for DL model training and

Table 2 Description of bifurcation of ROIs belonging from DDSM and MIAS datasets

	MIAS ROIs	DDSM ROIs
Balanced Bifurcation	Training set: 3220 Testing set: 3220	Training set: 24400 Testing set: 24400
Unbalanced Bifurcation	Training set: 4508 Testing set: 1932	Training set: 34160 Testing set: 14640

attains the promising results [43]. SGD is an optimization technique which is mathematically defined as a given expression in Eq. (2) for training sample $tr(x)$ with label $tr_b(y)$.

$$\theta = \theta - \eta^{\nabla} \partial j(\theta, tr(x)^{(i)}; tr_b(y^{(i)})) \quad (2)$$

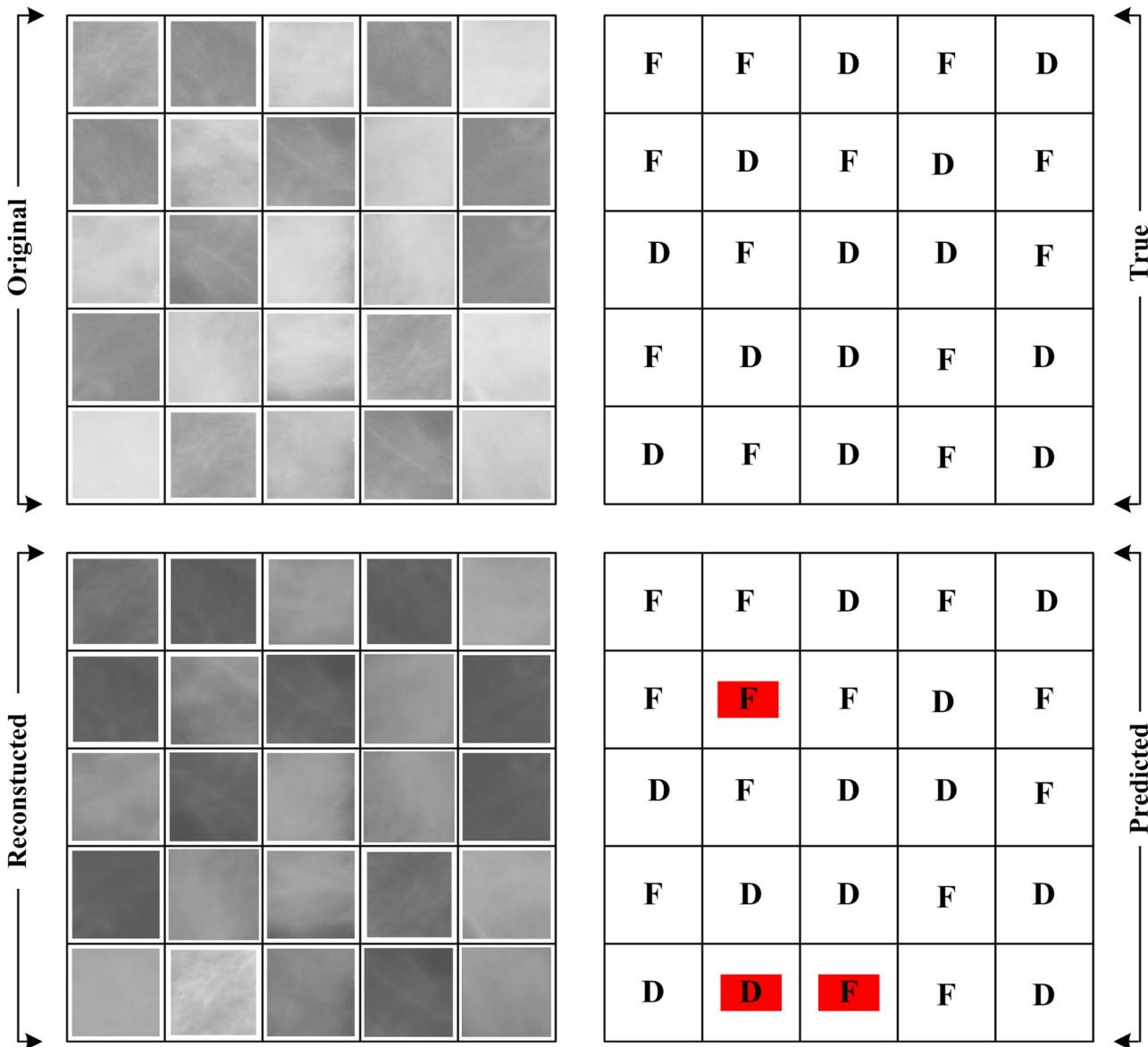


Fig. 16 Sample of the original and augmented image with true class and predicted class

The complete training algorithm is given as Algorithm 1.

Performance Evolutionary Parameters

In this work, the proposed work has been evaluated through accuracy and kappa coefficient. For accuracy calculation, a confusion matrix (CM) is used as shown in Fig. 14, and the mathematical expression of accuracy is also given in the same.

Kappa Coefficient

The kappa coefficient [44] is the statistical analysis of the proposed work, which shows the significance or reliability of the proposed work. The kappa coefficient τ is calculated with the help of given Eq. (3).

$$\tau = \frac{P_0 - P_e}{1 - P_e} \quad (3)$$

where τ is kappa value, P_0 is the probability of observed agreement, and P_e is the probability of hypothetical agreement. With the help of CM, P_0 and P_e are computed as:

$$P_0 = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

and

$$P_e = P_{class1} + P_{class2} \quad (5)$$

where P_{class1} and P_{class2} are computed as given expression.

$$P_{class1} = \frac{TP + FP}{TP + FP + TN + FN} * \frac{TP + FN}{TP + FP + TN + FN} \quad (6)$$

$$P_{class2} = \frac{FN + TN}{TP + FP + TN + FN} * \frac{FP + TN}{TP + FP + TN + FN}$$

The relation between kappa coefficient and the significance of the system is shown in Fig. 15.

Experiments and Results

Experimental Setup

The complete experimentation has been performed at HP Z4 G4 workstation. The specification of the system is given as Intel Xeon W-2014 CPU @ 3.2 GHz, 64 GB RAM, 4 GB NVIDIA Quadro P1000, 256 GB SSD, and 2 TB SATA HDD. All the images and ROIs are stored in this system, and the Python environment is used for performing the experiments.

Experiments

In this work, meticulous experimentations have been carried out for the dense tissue pattern characterization using

deep learning models. To achieve the desired outcome, AlexNet and ResNet-18 deep learning models have been used. The input mammograms were taken from mini-MIAS and DDSM databases. Due to the limited number of samples available in MIAS and DDSM dataset, virtual images were generated using data augmentation, for the training and testing purpose of the model. After the augmentation, $322 \times 20 = 6440$ ROIs of MIAS and $4880 \times 10 = 48800$ ROIs of the DDSM dataset are generated. Further, the complete set of ROIs is bifurcated into the training and testing set. The description of the bifurcation of ROIs belonging from DDSM and MIAS dataset is shown in Table 2.

The sample of the original and augmented dataset with true class and predicted class is shown in Fig. 16.

In this work, experiments have been carried out for MIAS, DDSM, and MIAS + DDSM mammograms. Initially, a model is designed for MIAS images and tested with the test samples from MIAS mammograms. In the next model, input mammogram ROIs are taken from the DDSM dataset, and the test set is also generated from the same set. In the next experiments, input ROIs are taken from DDSM and MIAS both dataset and then train the model from a training set ROIs and lastly test the model using the testing set. The bifurcation of the training and testing set is performed by balanced and unbalanced methods. The list of experiments carried out for the work is given in Table 3.

Experiment 1

In this experiment, a total number of 6440 ROIs have been used for dense tissue pattern classification using *AlexNet* and *ResNet-18* models. From 6440 ROIs, training and testing sets are created using balance bifurcation; therefore, 3220 ROIs are used as a training set, and the remaining 3220 ROIs are used as a testing set. From 3220 ROIs of the training set, 1060 ROIs belong to the fatty class, and 2160 ROIs belong to dense tissue class. In the same manner, the testing set is created. The four activation functions ‘ReLU’, ‘Sigmoid’, ‘Tanh’, and ‘Leaky ReLU’ are used for the experiment, and the obtained results are reported in Table 4. The distribution of ROIs is given as

Total ROIs of fatty class : $106 \times 20 = 2120$ (1060 training + 1060 testing)

Total ROIs of dense class : $216 \times 20 = 4320$ (2160 training + 2160 testing).

Experiment 2

In this experiment, a total number of 6440 ROIs have been used for dense tissue pattern classification using *AlexNet* and *ResNet-18* model. From 6440 ROIs, training and testing sets are created using unbalanced bifurcation; therefore, 4508 ROIs are used as a training set, and remaining 1932 ROIs are used as a testing set.

Table 3 Details of experiments carried out for the work

Dataset	Dataset bifurcation	Experiment no	Description
MIAS	Balanced bifurcation	Experiment 1	Dense tissue pattern characterization using <i>AlexNet</i> and <i>ResNet-18</i> model
	Unbalanced bifurcation	Experiment 2	Dense tissue pattern characterization using <i>AlexNet</i> and <i>ResNet-18</i> model
DDSM	Balanced bifurcation	Experiment 3	Dense tissue pattern characterization using <i>AlexNet</i> and <i>ResNet-18</i> model
	Unbalanced bifurcation	Experiment 4	Dense tissue pattern characterization using <i>AlexNet</i> and <i>ResNet-18</i> model
DDSM+MIAS	Balanced bifurcation	Experiment 5	Dense tissue pattern characterization using <i>AlexNet</i> and <i>ResNet-18</i> model
	Unbalanced bifurcation	Experiment 6	Dense tissue pattern characterization using <i>AlexNet</i> and <i>ResNet-18</i> model

The obtained results for four different activation functions are reported in Table 5. The description of ROIs distribution ROIs is given as

Total ROIs of fatty class : $106 * 20 = 2120$ {1484 training + 636 testing}

Total ROIs of dense class : $216 * 20 = 4320$ {3024 training + 1296 testing}.

are reported in Table 6. The description of ROIs distribution is given as

Total images fatty class : 24600{12300 training + 12300 testing}

Total images dense class : 24200{12100 training + 12100 testing}.

Experiment 4

In this experiment, 4880 ROIs are extracted from each mammogram taken from the DDSM dataset. To improve the learning of the DL model, the large number of input samples required; therefore, augmentation is used to create virtual 48800 ROIs for dense tissue pattern classification using *AlexNet* and *ResNet-18* model. Further, these samples are divided into training and testing sets using an unbalanced manner, i.e., 70;30 ratio. The obtained results for four different activation functions are reported in Table 7, and the description of ROIs distribution ROIs is given as

Total images fatty class : 24600{17220 training + 7380 testing}

Total images dense class : 24200{16940 training + 7260 testing}.

Experiment 3

In this experiment, 4880 cases are taken from the DDSM dataset, and an ROI from each mammogram has been extracted according to previously mentioned steps and area. After the augmentation, a total set of a total number of 48800 ROIs have been generated for dense tissue pattern classification using *AlexNet* and *ResNet-18* model. From 48800 ROIs, training and testing sets are created using balance bifurcation; therefore, 24400 ROIs are used as a training set, and remaining 24400 ROIs are used as a testing set. The obtained results for four different activation functions

Table 4 Classification accuracy for MIAS dataset

DL model	Activation function	Confusion matrix		Accuracy (%)	Kappa coefficient (τ)
		Fatty	Dense		
AlexNet	<i>ReLU</i>	Fatty	897	163	88.9
		Dense	194	1966	0.750
		Fatty	897	163	86.3
	<i>Sigmoid</i>	Dense	278	1882	0.698
		Fatty	895	165	86.1
		Dense	282	1878	0.694
	<i>Tanh</i>	Fatty	873	187	85.0
		Dense	296	1864	0.669
		Fatty	908	152	89.8
	<i>Leaky ReLU</i>	Dense	176	1984	0.770
		Fatty	864	196	87.2
		Dense	216	1944	0.711
ResNet-18	<i>ReLU</i>	Fatty	870	190	85.0
		Dense	293	1867	0.668
		Fatty	873	187	85.2
	<i>Sigmoid</i>	Dense	289	1871	0.673
		Fatty	864	196	87.2
		Dense	216	1944	0.711

Table 5 Classification accuracy for MIAS dataset

DL model	Activation function	Confusion matrix		Accuracy (%)	τ
		Fatty	Dense		
<i>AlexNet</i>	<i>ReLU</i>	Fatty	569	67	89.3
		Dense	138	1158	0.766
	<i>Sigmoid</i>	Fatty	581	55	89.1
		Dense	155	1141	0.763
	<i>Tanh</i>	Fatty	548	88	86.1
		Dense	179	1117	0.698
	<i>Leaky ReLU</i>	Fatty	541	91	88.6
		Dense	125	1171	0.749
	<i>ResNet-18</i>	<i>ReLU</i>	581	55	91.3
		Dense	113	1183	0.807
	<i>Sigmoid</i>	Fatty	570	66	89.6
		Dense	135	1158	0.770
	<i>Tanh</i>	Fatty	568	68	87.8
		Dense	168	1128	0.734
	<i>Leaky ReLU</i>	Fatty	558	78	86.5
		Dense	183	1113	0.706

Experiment 5

In this experiment, a total of 55240 ROIs (6440 belonging to MIAS and 48800 belonging to DDSM) are considered. The total number of fatty class ROI is 26720, and 28520 ROIs belong to dense tissue class. In this experiment, balance bifurcation of the dataset is used to create training and testing sets. The *AlexNet* and *ResNet-18* model is used for dense tissue pattern classification. The obtained results for four different activation functions are reported in Table 8, and the description of ROIs distribution ROIs is given as

Total images fatty class : 26,720{13,360 training + 13,360 testing}
Total images dense class : 28,520{14,260 training + 14,260 testing}.

Experiment 6

In this experiment, the same number of samples as Experiment 5 is used, but the bifurcation of the training and testing sets is made according to the unbalanced manner. The ratio of 70:30 is used for the training and testing set creation. The description of ROIs distribution ROIs is given as:

Table 6 Classification accuracy for DDSM

DL model	Activation function	Confusion matrix		Accuracy (%)	τ
		Fatty	Dense		
<i>AlexNet</i>	<i>ReLU</i>	Fatty	10849	1451	88.2
		Dense	1428	10672	0.764
	<i>Sigmoid</i>	Fatty	11123	1177	86.3
		Dense	2166	9934	0.725
	<i>Tanh</i>	Fatty	10418	1882	85.6
		Dense	1631	10469	0.712
	<i>Leaky ReLU</i>	Fatty	10568	1732	86.0
		Dense	1684	10416	0.720
	<i>ResNet-18</i>	<i>ReLU</i>	11112	1188	90.2
		Dense	1206	10894	0.803
	<i>Sigmoid</i>	Fatty	10910	1390	88.7
		Dense	1368	10732	0.773
	<i>Tanh</i>	Fatty	10676	1624	86.8
		Dense	1597	10503	0.736
	<i>Leaky ReLU</i>	Fatty	11218	1082	89.7
		Dense	1420	10680	0.794

Table 7 Classification accuracy for DDSM using unbalanced bifurcation

DL model	Activation function	Confusion matrix		Accuracy (%)	τ
		Fatty	Dense		
<i>AlexNet</i>	<i>ReLU</i>	Fatty	6389	991	87.8
		Dense	794	6466	0.756
	<i>Sigmoid</i>	Fatty	6332	1048	85.1
		Dense	1133	6127	0.702
	<i>Tanh</i>	Fatty	6103	1277	84.6
		Dense	977	6283	0.692
	<i>Leaky ReLU</i>	Fatty	6456	924	87.1
		Dense	964	6296	0.742
	<i>ResNet-18</i>	<i>ReLU</i>	6812	568	92.3
		Dense	559	6701	0.846
		<i>Sigmoid</i>	6718	662	89.0
		Dense	946	6314	0.780
		<i>Tanh</i>	6287	1093	87.3
		Dense	763	6497	0.746
		<i>Leaky ReLU</i>	6676	704	90.3
		Dense	702	6558	0.807

Total images fatty class : 26,720{18704 training + 8016 testing}

Total images dense class : 28,520{19964 training + 8556 testing}.

The obtained results for different activation functions are reported in Table 9.

Results Analysis

To achieve the efficient model for dense tissue characterization, extensive experiments have been carried out on MIAS and DDSM mammograms using *AlexNet* and *ResNet-18*. Initially, MIAS and DDSM images have been considered

separately for model designing, and later on, combined samples are used, so that maximum variability of input samples are considered. After the experimentations, following major outcomes have been induced as:

- (i) For the MIAS dataset, two experiments (Experiment 1 and Experiment 2) have been performed using the *AlexNet* and *ResNet-18* model. In these experiments, training and testing sets have been created using balance and unbalanced bifurcation approaches. The various activation functions have been checked for the convolutional layer, and results have been reported

Table 8 Classification accuracy for MIAS + DDSM images using balanced bifurcation

DL model	Activation function	Confusion matrix		Accuracy (%)	τ
		Fatty	Dense		
<i>AlexNet</i>	<i>ReLU</i>	Fatty	11583	1777	86.7
		Dense	1897	12363	0.733
	<i>Sigmoid</i>	Fatty	11336	2024	85.6
		Dense	1654	12306	0.730
	<i>Tanh</i>	Fatty	11429	1931	84.8
		Dense	2268	11992	0.695
	<i>Leaky ReLU</i>	Fatty	11649	1711	87.2
		Dense	1825	12435	0.743
	<i>ResNet-18</i>	<i>ReLU</i>	12097	1263	88.3
		Dense	1969	12291	0.766
		<i>Sigmoid</i>	11730	1630	87.8
		Dense	1740	12520	0.755
		<i>Tanh</i>	11416	1944	86.2
		Dense	1868	12392	0.723
		<i>Leaky ReLU</i>	11667	1693	88.1
		Dense	1594	12666	0.761

in Tables 4 and 5. The performance of each experiment has been evaluated using classification accuracy (Acc) and kappa coefficient (τ). The highest accuracy of 89.8% (2892/3220) has been reported for Experiment 1 using *ResNet-18* model for *ReLU* non-linear activation function and value of the kappa coefficient (τ) is 0.770. For the same testing set, the accuracy of 88.9% (2863/3220) has been achieved using the *AlexNet* model for *ReLU* activation function, and the value of τ is obtained as 0.750. Further, the accuracy of 86.3%, 86.1%, and 85.0% have been obtained and the value of τ is obtained as 0.698, 0.694, and 0.669 for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ using the *AlexNet* model. Similarly, 87.2%, 85.0%, and 85.2% of accuracy and the kappa coefficient as 0.711, 0.668, and 0.673 have been obtained for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ using the *ResNet-18* model.

For Experiment 2, the highest Acc of 89.3% (1727/1932) and 91.3% (1764/1932) has been achieved for the *AlexNet* and *ResNet-18* model, respectively using the *ReLU* activation function. The value of kappa coefficient τ is 0.776 and 0.807 for the *AlexNet* and *ResNet-18* model, respectively. In this experiment, the non-linear activation functions ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ have been also checked, and the obtained classification accuracy for each function is noted as 89.1%, 86.1%, and 88.6%, and the value of kappa coefficient is obtained as 0.763, 0.698, and 0.749 for *AlexNet* model. In the similar manner, the accuracy of 89.6%, 87.8%, 86.5%, and value of τ is 0.770, 0.734, and 0.706 for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ activation function using the *ResNet-18* model.

- (ii) For the DDSM dataset, two extensive experiments (Experiment 3 and Experiment 4) have been performed for dense tissue pattern characterization using *AlexNet* and *ResNet-18* model. In this experiment, four activation functions ‘*ReLU*’, ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ is used for both DL model and obtained classification accuracy, and kappa coefficient value is reported in Tables 6 and 7.

In Experiment 3, the bifurcation of training and testing samples is made according to the balanced bifurcation. The achieved accuracy is 90.2% (22006/24400) using the *ResNet-18* model and 88.2% (21521/24400) accuracy is achieved using *AlexNet* model for *ReLU* activation function. The value of τ is 0.803 and 0.764 obtained for the *ResNet-18* and *AlexNet*, respectively. The accuracy for other activation functions (‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’) for *AlexNet* model is obtained as 86.3%, 85.6%, and 86.0% respectively. The kappa value for the same activation function is 0.725, 0.712, and 0.720 for *AlexNet* model, respectively. Similarly, the obtained accuracy for *ResNet-18* model is as 88.7%, 86.8%, and 89.7% respectively. The kappa value for the ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ using the *ResNet-18* model is as 0.773, 0.736, and 0.794, respectively.

In Experiment 4, the same number of samples as Experiment 3 are used, but the distribution of samples in the training and testing sets is made according to the ratio of 70:30. It means 70% of the total sample is used as a training set, and the remaining 30% is used as a testing set. The highest classification accuracy of 92.3% (13513/14610) is achieved using

Table 9 Classification accuracy for MIAS + DDSM using unbalanced bifurcation

DL model	Activation function	Confusion matrix		Accuracy (%)	τ
		Fatty	Dense		
<i>AlexNet</i>	<i>ReLU</i>	Fatty	7378	638	90.8
		Dense	888	7668	0.815
		<i>Sigmoid</i>	7014	1002	87.5
			1070	7486	0.749
	<i>Tanh</i>	Fatty	6965	1051	86.9
		Dense	1130	7426	0.736
		<i>Leaky ReLU</i>	7054	962	88.0
			1027	7529	0.759
	<i>ResNet-18</i>	<i>ReLU</i>	Fatty	7289	727
			Dense	599	7957
		<i>Sigmoid</i>	Fatty	7102	914
			Dense	976	7580
		<i>Tanh</i>	Fatty	7014	1002
			Dense	1070	7486
		<i>Leaky ReLU</i>	Fatty	7078	938
			Dense	998	7558

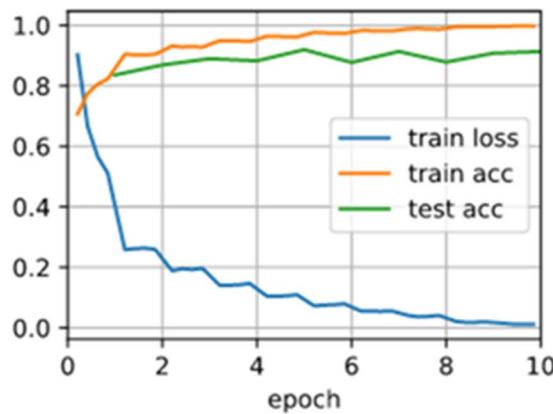
the *ResNet-18* model and 87.8% (12855/14610) of accuracy is achieved for the *AlexNet* model for the *ReLU* function. The kappa value for the same activation function for the *AlexNet* and *ResNet-18* model is obtained as 0.846 and 0.756, respectively. The accuracy for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ using *ResNet-18* is 89.0%, 87.3%, and 90.3%, respectively. The kappa values for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ activation function are 0.780, 0.746, and 0.807 respectively for the *ResNet-18* model. Similarly, the accuracy for three activation functions is obtained as 85.1%, 84.6%, and 87.1% for *AlexNet* model using ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ activation functions, respectively, and the kappa

coefficient values are as 0.702, 0.692, and 0.742 respectively.

- (iii) For MIAS + DDSM dataset, two extensive experiments (Experiment 5 and Experiment 6) have been performed for dense tissue pattern characterization using *AlexNet* and *ResNet-18* model. In this experiment, four activation functions ‘*ReLU*’, ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ are used for both DL model and obtained classification accuracy, and kappa coefficient value is reported in Tables 8 and 9, respectively.

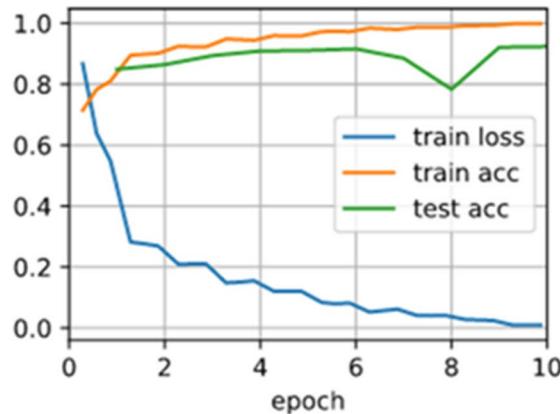
From Table 8, it has been found that the maximum classification accuracy of 88.3% (24388/27620) is attained for the *ReLU* activation function using

loss 0.012, train acc 0.997, test acc 0.913
35.3 examples/sec on cpu(0)



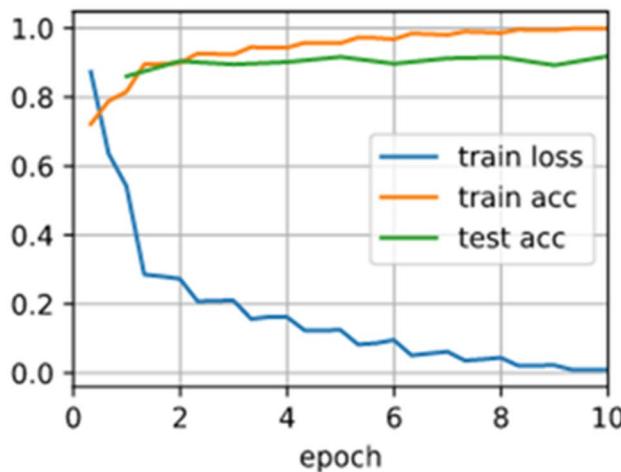
(a) MIAS dataset

loss 0.010, train acc 0.998, test acc 0.923
31.9 examples/sec on cpu(0)



(b) DDSM dataset

loss 0.009, train acc 0.999, test acc 0.919
23.5 examples/sec on cpu(0)



(c) MIAS+DDSM dataset

Fig. 17 Training, testing, and train loss curves for *ResNet-18* model

Table 10 Misclassification analysis of experiments

Experiment no	Model	Dataset	No. of samples	Miss_F	Miss_D	Mis_Acc
<i>Experiment no. 1</i>	AlexNet	MIAS	3220	152	176	10.2
<i>Experiment no. 2</i>	ResNet-18	MIAS	1932	55	113	8.7
<i>Experiment no. 3</i>	AlexNet	DDSM	24400	1188	1206	9.8
<i>Experiment no. 4</i>	ResNet-18	DDSM	14640	568	559	7.7
<i>Experiment no. 5</i>	AlexNet	MIAS + DDSM	27620	1263	1969	11.7
<i>Experiment no. 6</i>	ResNet-18	MIAS + DDSM	16572	727	599	8.1

Miss_F Misclassified samples of fatty class, *Miss_D* misclassified samples of dense class, *Mis_Acc*, misclassification accuracy

ResNet-18 model, and the kappa value of 0.766 is achieved. The accuracy for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ function is reported as 87.8%, 86.2%, and 88.1% respectively, and the kappa value is obtained as 0.755, 0.723, and 0.761 respectively. With the help of the *AlexNet* model, 87.2% (24084/27620) of accuracy is achieved using the *Leaky ReLU* activation function, and the kappa value is obtained as 0.743. The accuracy for ‘*ReLU*’, ‘*Sigmoid*’, and ‘*Tanh*’ activation function is obtained as 86.7%, 85.6%, and 84.8%, respectively. The kappa coefficient value for these activation functions is obtained as 0.733, 0.730, and 0.695 respectively.

From Table 9, it has been observed that the *AlexNet* and *ResNet-18* model is used for the development of dense tissue pattern characterization systems. The highest classification accuracy is obtained as 91.9% (15246/16572) and the kappa value as 0.839 using *ResNet-18* model for *ReLU* activation function. The accuracy for *Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ activation function is obtained as 88.6%, 87.5%, and 88.3%, respectively. The kappa coefficient values of 0.771, 0.749, and 0.766 are obtained for *Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ activation functions using the *ResNet-18* model. The highest accuracy for *AlexNet* model is obtained as 90.8% (15046/16572) using the *ReLU* activation function, and the kappa value is obtained as 0.815. The accuracy for ‘*Sigmoid*’, ‘*Tanh*’, and ‘*Leaky ReLU*’ activation functions

is obtained as 87.5%, 86.9%, and 88.0%, respectively, and the kappa coefficient value is obtained as 0.749, 0.736, and 0.759, respectively.

- (iv) In this work, six different experiments have been performed with the help of three different combinations of the dataset (i.e., MIAS, DDSM, and MIAS + DDSM). The obtained maximum accuracy for MIAS dataset is 91.3% (1764/1932), for DDSM, dataset is 92.3% (13513/14610), and for MIAS + DDSM, dataset is 91.9% (15246/16572). The training and testing accuracy for every model is shown in Fig. 17.

Misclassification Analysis

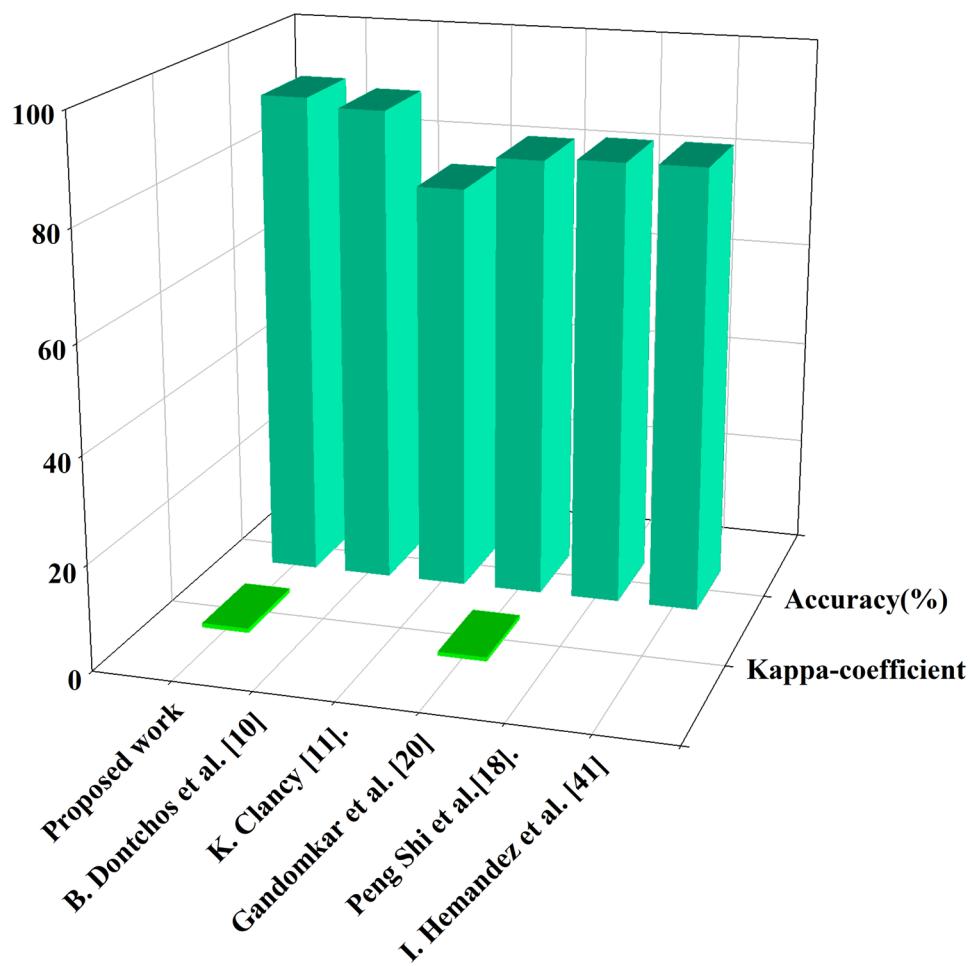
After the successful completion of experiments, so many samples are misclassified. The analysis of the misclassification of every experiment is given in Table 10.

From Table 10, it has been observed that the minimum misclassification accuracy is 7.7% (1127/14640) using the *ResNet-18* model for DDSM dataset. It shows that 1127 samples are not correctly predicted. From 1127 samples, 568 samples of the fatty class and 559 samples of the dense class have been incorrectly classified. The minimum misclassification accuracy for MIAS dataset is 8.7% using the *ResNet-18* model. In the same manner, the minimum misclassification accuracy for MIAS + DDSM dataset is 8.1% using the *ResNet-18* model.

Table 11 Comparison between proposed work and the state of art

	DL model	Number of samples	Accuracy (%)	Kappa coefficient
Proposed work	<i>ResNet-18</i>	48800	92.3	0.846
Valencia-Hernandez et al. [26]	<i>Takagi–Sugeno</i>	1010	84.2	–
Dontchos et al. [9]	<i>CNN</i>	2174	90.7	–
Clancy et al. [10]	<i>AlexNet</i>	22000	77.0	–
Gandomkar et al. [16]	<i>Inception-V3</i>	3813	83.3	0.775
Shi et al. [15]	<i>VGG-16</i>	10304	83.9	–

Fig. 18 Comparative analysis between proposed works and previously published



Comparative Analysis

The performance of proposed work has been compared with the current state of the art, and comparative analysis table is shown in Table 11. The comparative analysis has been performed on the basis used DL model, the total number of considered cases for experiment, accuracy, and kappa coefficient. Figure 18 shows the comparative analysis between proposed works and previously published.

In Table 11, previously published work [9, 10, 15, 16, 26] is compared with present work. The comparative table shows the accuracy of present work is better than the accuracy of the state of art work done in [9, 10, 15, 16]. The accuracy of the work reported in a study [9] is 90.7% that is the highest among the accuracy of study reported in [10, 15, 16, 26]. It is also worth mentioning that the proposed work accuracy is 92.3% which is higher than the previously reported accuracy in study [9].

Among previously published work [9, 10, 15, 16, 26], only study done by Gandomkar et al. [16] done kappa coefficient evaluation for their work. The obtained kappa value for proposed work is 0.846, and kappa value for the study

[16] is 0.775. The kappa coefficient value obtained for the proposed work shows that the significance of proposed work is more than the previously published work. Thus, the present study is more suitable for the clinical purpose for dense tissue pattern characterizations.

Conclusion

It is very well known that dense tissue is a major risk factor for the growth of cancerous cells in women's breasts. Therefore, the present study reports the performance of the proposed dense tissue pattern characterization model using deep neural networks. Initially, MIAS and DDSM datasets are used for input image samples. Due to less number of samples, data augmentation has been performed to generate virtual samples, so that the training and testing of models are done properly. After the augmentation, the problem of under-fitting and over-fitting of the model is reduced.

In this work, two deep learning models (*AlexNet* and *ResNet-18*) are used, and four activation functions ('*Sigmoid*', '*Tanh*', '*ReLU*', and '*LeakyReLU*') have been tested

with each model. To achieve the desired objective, extensive experiments have been performed with the different combination of training and testing samples with different activation functions using *AlexNet* and *ResNet-18* model. For every experiment, the outcome of the results has been measured in terms of accuracy and kappa coefficient. The obtained accuracy and kappa value show the activation function '*ReLU*', and deep neural network model *ResNet-18* is more suitable for dense tissue pattern characterization. Finally, it has been concluded that the designed model is more suitable for clinical purposes, and it shall be helpful for an expert person for proper and adequate scheduling of the treatment.

This work is also suffering from the manual ROIs extraction. If the ROIs extraction performed automatically then the execution time and performance will be improved. The same limitations will be also considered as a future work of the proposed work. In near future, the same set of approaches have been also used for the designing a computerized framework for breast density classification on full-field screen mammograms.

Declarations

Ethical approval The reported research was carried out using secondary data. Hence, ethical approval was not required.

Research involving human participants and/or animals This research did not directly involve any human participant or animal.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Boyd N, Martin L, Gunasekara A, Melnichouk O, Maudsley G, Peressotti C, Minkin S. Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. *Cancer Epidemiol Prev Biomark.* 2009;18(6):1754–62. <https://doi.org/10.1158/1055-9965.EPI-09-0107>.
- Brown KA, Simpson ER. Obesity and breast cancer: progress to understanding the relationship. *Cancer Res.* 2010;70(1):4–7. <https://doi.org/10.1158/0008-5472.CAN-09-2257>.
- American Cancer Society. Breast cancer early detection the importance of finding breast cancer early. American Cancer Society. 2014.
- Ramathuba DU, Ratshirumbi CT, Mashamba TM. Knowledge, attitudes and practices toward breast cancer screening in a rural South African community. *Curationis.* 2015;38(1):1–8. <https://doi.org/10.4102/curationis.v38i1.1172>.
- Smith RA, Cokkinides V, von Eschenbach AC, Levin B, Cohen C, Runowicz CD, Eyre HJ. American Cancer Society guidelines for the early detection of cancer. *CA Cancer J Clin.* 2002;52(1):8–22. <https://doi.org/10.3322/canjclin.52.1.8>.
- Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer.* 1976;37(5):2486–92.
- Kumar I, Bhaduria HS, Virmani J, Thakur S. A classification framework for prediction of breast density using an ensemble of neural network classifiers. *Biocybern Biomed Eng.* 2017;37(1):217–28. <https://doi.org/10.1016/j.bbe.2017.01.001>.
- Kumar I, Bhaduria HS, Virmani J, Thakur S. A hybrid hierarchical framework for classification of breast density using digitized film screen mammograms. *Multimed Tools Appl.* 2017;76(18):18789–813. <https://doi.org/10.1007/s11042-016-4340-z>.
- Dontchos BN, Yala A, Barzilay R, Xiang J, Lehman CD. External validation of a deep learning model for predicting mammographic breast density in routine clinical practice. *Acad Radiol.* 2021;28(4):475–80. <https://doi.org/10.1016/j.acra.2019.12.012>.
- Clancy K, Abourabid S, Mohamed A, Sumkin J, Wu S. Deep learning pre-training strategy for mammogram image classification: an evaluation study. *J Digit Imaging.* 2020;33(5):1257–65. <https://doi.org/10.1007/s10278-020-00369-3>.
- Kumar I, Bhaduria HS, Virmani J. A computerised framework for prediction of fatty and dense breast tissue using principal component analysis and multi-resolution texture descriptors. *Int J Comput Syst Eng.* 2018;4(2–3):73–85. <https://doi.org/10.1504/IJCSYE.2018.091386>.
- Huang ML, Lin TY. Considering breast density for the classification of benign and malignant mammograms. *Biomed Signal Process Control.* 2021;67: 102564. <https://doi.org/10.1016/j.bspc.2021.102564>.
- Román M, Louro J, Posso M, Alcántara R, Peñalva L, Sala M, Castells X. Breast density benign breast disease and risk of breast cancer over time. *Eur Radiol.* 2021;31:4839–47. <https://doi.org/10.1007/s00330-020-07490-5>.
- Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, Munishkumaran S. Current status of the digital database for screening mammography. In: Karssemeijer N, Thijssen M, Hendriks J, van Erning L, editors. *Digital mammography.* Dordrecht: Springer. 1998. p. 457–60. https://doi.org/10.1007/978-94-011-5318-8_75.
- Shi P, Wu C, Zhong J, Wang H. Deep learning from small dataset for BI-RADS density classification of mammography images. In: 2019 10th International Conference on Information Technology in Medicine and Education (ITME). IEEE. 2019. p. 102–9. <https://doi.org/10.1109/ITME.2019.00034>.
- Gandomkar Z, Suleiman ME, Demchig D, Brennan PC, McEntee MF. BI-RADS density categorization using deep neural networks. In: Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment (Vol. 10952). Int Soc Opt Photonics. 2019. p. 109520N. <https://doi.org/10.1117/12.2513185>.
- Chang K, Beers AL, Brink L, Patel JB, Singh P, Arun NT, Tilkin M. Multi-institutional assessment and crowdsourcing evaluation of deep learning for automated classification of breast density. *J Am Coll Radiol.* 2020;17(12):1653–62. <https://doi.org/10.1016/j.jacr.2020.05.015>.

18. Kriti, Virmani J. Breast density classification using Laws' mask texture features. *Int J Biomed Eng Technol.* 2015;19(3):279–302. <https://doi.org/10.1504/IJBET.2015.072999>.
19. Mahmud M, Kaiser MS, Hussain A, Vassanelli S. Applications of deep learning and reinforcement learning to biological data. *IEEE Trans Neural Netw Learn Syst.* 2018;29(6):2063–79. <https://doi.org/10.1109/TNNLS.2018.2790388>.
20. Mahmud M, Kaiser MS, McGinnity TM, Hussain A. Deep learning in mining biological data. *Cognit Comput.* 2021;13(1):1–33. <https://doi.org/10.1007/s12559-020-09773-x>.
21. Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A. The state of the art of deep learning models in medical science and their challenges. *Multimed Syst.* 2020;27(4):599–613. <https://doi.org/10.1007/s00530-020-00694-1>.
22. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Lopez MAG. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed.* 2016;127:248–57. <https://doi.org/10.1016/j.cmpb.2015.12.014>.
23. Qiu Y, Yan S, Tan M, Cheng S, Liu H, Zheng B. Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study. In: *Medical Imaging 2016: Computer-Aided Diagnosis* (Vol. 9785). Int Soc Opt Photonics. 2016. p. 978520. <https://doi.org/10.1117/12.2216336>.
24. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: *IEEE conference on computer vision and pattern recognition. IEEE.* 2009;2009:248–55. <https://doi.org/10.1109/CVPR.2009.5206848>.
25. Deng L. A tutorial survey of architectures algorithms and applications for deep learning. *APSIPA Trans Signal Inf Process.* 2014;3(2):1–29. <https://doi.org/10.1017/atsip.2013.9>.
26. Valencia-Hernandez I, Peregrina-Barreto H, Reyes-Garcia CA, Lopez-Armas GC. Density map and fuzzy classification for breast density by using BI-RADS. *Comput Methods Programs Biomed.* 2021;200: 105825. <https://doi.org/10.1016/j.cmpb.2020.105825>.
27. Wu N, Geras KJ, Shen Y, Su J, Kim SG, Kim E, Cho K. Breast density classification with deep convolutional neural networks. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE. 2018. p. 6682–6. <https://doi.org/10.1109/ICASSP.2018.8462671>.
28. Thomaz RL, Carneiro PC, Patrocínio AC. Feature extraction using convolutional neural network for classifying breast density in mammographic images. In *Medical Imaging 2017: Computer-Aided Diagnosis* (Vol. 10134). Int Soc Opt Photonics. 2017. p. 101342M. <https://doi.org/10.1117/12.2254633>.
29. Mohamed AA, Berg WA, Peng H, Luo Y, Jankowitz RC, Wu S. A deep learning method for classifying mammographic breast density categories. *Med Phys.* 2018;45(1):314–21. <https://doi.org/10.1002/mp.12683>.
30. Chan H, Helvie M. Deep learning for mammographic breast density assessment and beyond. *Radiology.* 2019;290(1):466–82. <https://doi.org/10.1148/radiol.2018182116>.
31. Li H, Giger ML, Huo Z, Olopade OI, Lan L, Weber BL, Bonta I. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location. *Med Phys.* 2004;31(3):549–55. <https://doi.org/10.1118/1.1644514>.
32. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Asari VK. The history began from alexnet: a comprehensive survey on deep learning approaches. *arXiv preprint.* 2018. <https://arxiv.org/abs/1803.01164>.
33. Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: comparison of trends in practice and research for deep learning. *arXiv preprint.* 2018. <https://arxiv.org/abs/1811.03378>.
34. Suckling JP. The mammographic image analysis society digital mammogram database. *Digital Mammo.* 1994;375–386. Elsevier Sc. B. V.
35. Wu E, Wu K, Cox D, Lotter W. Conditional infilling GANs for data augmentation in mammogram classification. In: *Image Analysis for Moving Organ, Breast, and Thoracic Images.* Cham: Springer; 2018. p. 98–106. https://doi.org/10.1007/978-3-030-00946-5_11.
36. Lemley J, Bazrafkan S, Corcoran P. Smart augmentation learning an optimal data augmentation strategy. *IEEE Access.* 2017;5:5858–69. <https://doi.org/10.1109/ACCESS.2017.2696121>.
37. Mikolajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In: *international interdisciplinary PhD workshop (IIPHDW).* IEEE. 2018;2018:117–22. <https://doi.org/10.1109/IIPHDW.2018.8388338>.
38. Jiang W, Liu P, Wen F. An improved vector quantization method using deep neural network. *AEU Int J Electron Commun.* 2017;2017(72):178–83. <https://doi.org/10.1016/j.aeue.2016.12.002>.
39. Kocic J, Jovicic N, Drndarevic V. An end-to-end deep neural network for autonomous driving designed for embedded automotive platforms. *Sensors.* 2019;19(9):2064. <https://doi.org/10.3390/s19092064>.
40. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM.* 2017;60(6):84–90. <https://doi.org/10.1145/3065386>.
41. Zhong Z, Jin L, Xie Z. High performance offline handwritten Chinese character recognition using googlenet and directional feature maps. In: *2015 13th International Conference on Document Analysis and Recognition (ICDAR).* IEEE; 2015. p. 846–50. <https://doi.org/10.1109/ICDAR.2015.7333881>.
42. Yoo HJ. Deep convolution neural networks in computer vision: a review. *IEIE Trans Smart Process Comput.* 2015;4(1):35–43. <https://doi.org/10.5573/IEIESPC.2015.4.1.035>.
43. Bottou L. Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010.* Physica-Verlag HD; 2010. p. 177–86. https://doi.org/10.1007/978-3-7908-2604-3_16.
44. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med.* 22(3):276–82. <https://hrcak.srce.hr/89395>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.