



Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection

Shivaji D. Pawar^{1,2}, Kamal K. Sharma^{3*}, Suhas G. Sapate⁴, Geetanjali Y. Yadav⁵, Roobaea Alroobaea⁶, Sabah M. Alzahrani⁶ and Mustapha Hedabou⁷

¹ Department of Computer Science and Engineering, Lovely Professional University, Jalandhar, India, ² SIES Graduate School of Technology, Navi Mumbai, India, ³ School of Electronics and Electrical Engineering, Lovely Professional University, Jalandhar, India, ⁴ Department of Computer Science and Engineering, Annasaheb Dange College of Engineering and Technology, Sangli, India, ⁵ NMC Royal Medical Centre Karama, Abu Dhabi, United Arab Emirates, ⁶ Department Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, ⁷ School of Computer Science, Mohammed VI Polytechnic University, Ben Guerir, Morocco

OPEN ACCESS

Edited by:

Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:

Clement Nyirenda,
University of the Western Cape,
South Africa
Xiao-Zhi Gao,
University of Eastern Finland, Finland

*Correspondence:

Kamal K. Sharma
kamal.23342@lpu.co.in

Specialty section:

This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: 27 February 2022

Accepted: 14 March 2022

Published: 25 April 2022

Citation:

Pawar SD, Sharma KK, Sapate SG, Yadav GY, Alroobaea R, Alzahrani SM and Hedabou M (2022) Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection.
Front. Public Health 10:885212.
doi: 10.3389/fpubh.2022.885212

Percentage mammographic breast density (MBD) is one of the most notable biomarkers. It is assessed visually with the support of radiologists with the four qualitative Breast Imaging Reporting and Data System (BIRADS) categories. It is demanding for radiologists to differentiate between the two variably allocated BIRADS classes, namely, "BIRADS C and BIRADS D." Recently, convolution neural networks have been found superior in classification tasks due to their ability to extract local features with shared weight architecture and space invariance characteristics. The proposed study intends to examine an artificial intelligence (AI)-based MBD classifier toward developing a latent computer-assisted tool for radiologists to distinguish the BIRADS class in modern clinical progress. This article proposes a multichannel DenseNet architecture for MBD classification. The proposed architecture consists of four-channel DenseNet transfer learning architecture to extract significant features from a single patient's two a mediolateral oblique (MLO) and two craniocaudal (CC) views of digital mammograms. The performance of the proposed classifier is evaluated using 200 cases consisting of 800 digital mammograms of the different BIRADS density classes with validated density ground truth. The classifier's performance is assessed with quantitative metrics such as precision, responsiveness, specificity, and the area under the curve (AUC). The concluding preliminary outcomes reveal that this intended multichannel model has delivered good performance with an accuracy of 96.67% during training and 90.06% during testing and an average AUC of 0.9625. Obtained results are also validated qualitatively with the help of a radiologist expert in the field of MBD. Proposed architecture achieved state-of-the-art results with a fewer number of images and with less computation power.

Keywords: breast cancer, BIRADS Density Classification, DenseNet, deep learning, multichannel architecture, mammographic breast density

INTRODUCTION

Breast cancer has its reputation as a deadly disease and it is the second most frequent event of departure for the women community in society. It is frequently diagnosed in women with 12.3% compared to the average population (1–4). The incidence of breast cancer is associated with many biomarkers such as calcifications, masses, mammographic breast density (MBD), and architectural distortion. Advanced apprehension and proactive strategy are the unique alternatives to protect the lives of breast cancer cases and regress their subconscious shock (5–8). MBD is an essential biomarker in interpreting digital mammograms and preparing a systematic mammography screening program. According to epidemiological investigations, females with highly dense breast tissue risk developing breast malignancy (9, 10). MBD is a radiographically visible density on the mammogram, consisting of lobular elements, ducts, and fibrous connective tissue compared to the lucent fatty tissue in the breast (11). The reports are prepared during the digital mammography screening program as per the American College of Radiology Breast Imaging Reporting and Data System (ACR BIRADS) catalog. This catalog was last modified in November 2015 (12, 13). As per the BIRADS, MBD gets classified into the four significant groups of mammographic breast density as class A, class B, class C, and class D, which reduces the sensitivity of digital mammography (14). **Figure 1** depicts all the MBD BIRADS classes.

Different scientific studies have revealed that digital mammograms' sensitivity strongly depends on the density class of the breast tissue. In dense breasts, the sensitivity of mammograms is as low as 63%, while in the low-density breast, there is an exponential rise of 87%. Hence, patients with high-density breasts have to go for additional imaging such as tomosynthesis, ultrasound, or breast MR to enhance cancer detection chances (15). Researchers have proposed many semiautomatic and automatic approaches for measuring breast density from the last two decades. However, the assessment of mammographic breast density (MBD) is subjective, which expert radiologists do (16). Despite this labeling, due to poor interreader and intrareader reproducibility, MBD classification has many limitations (17).

For an automatic objective assessment of MBD classification, many study efforts have been in progress from the last few decades. The initial study focuses on image processing techniques such as area-based thresholding, region growing, and clustering algorithms. A significant step formulated was the emergence of the machine learning (ML) algorithms based on different extracted image topographies from the histogram, texture intensities, patterns, and image acquisition parameters. Nowadays, deep learning algorithms provide yet another leap forward in MBD classification. Deep learning algorithms can go substantially deeper and discover all the significant features from the image. Due to system architecture and hardware improvements, it is possible to train deep learning architecture intensely. This improvement makes deep learning architecture an excellent tool for medical image analysis. Many deep learning architectures

such as LeNet, (visual geometry group) VGG19, highway networks, residual networks, and DenseNet are recorded in literature.

Nevertheless, before every deep learning network lets more intelligence, a new study intricacy happens, the “vanishing gradient problem.” DenseNet (dense convolutional network) provides unique insight to secure the best data flow between layers to solve the connectivity problem. This interface immediately combines all the layers, agreeing on characteristic map dimensions in feed-forward type; hence, the individual layer receives input from all the previous layers also transfers its distinct map to all the farther layers. Thus, the DenseNet concatenate feature map passes through all the subsequent layers instead of summarizing features such as ResNet. This concept includes $L(L+1)/2$ connections instead of L connections, identifying a dense connectivity pattern (18). The consequence of this connectivity guide, i.e., DenseNet architecture, provides the following advantages:

1. Effective solution for gradient vanishing.
2. Consolidation in characteristic distribution.
3. Provision of feature recycle.
4. Significant reduction in training parameters.
5. Easy to train and offers better parameter efficiency.

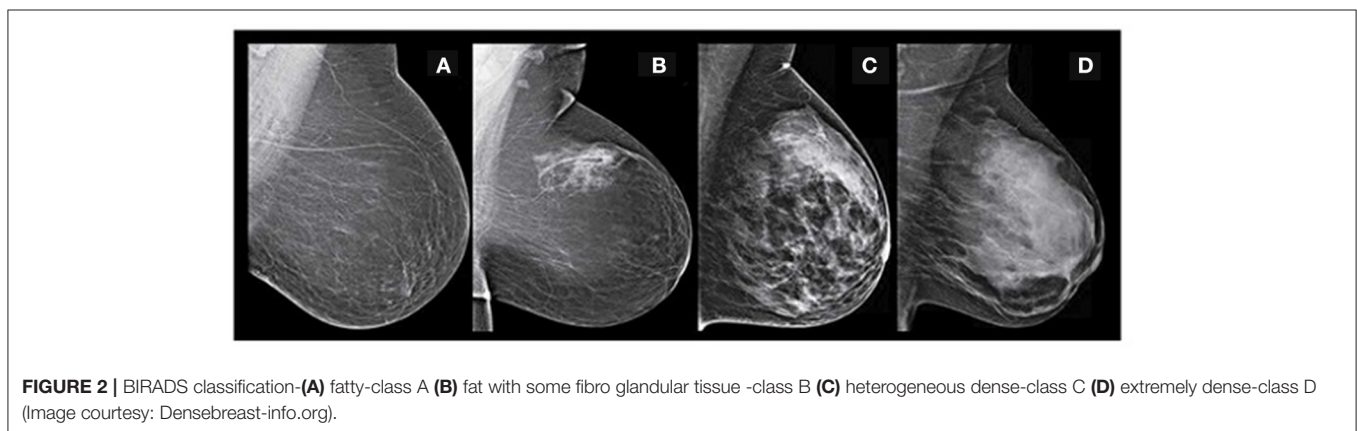
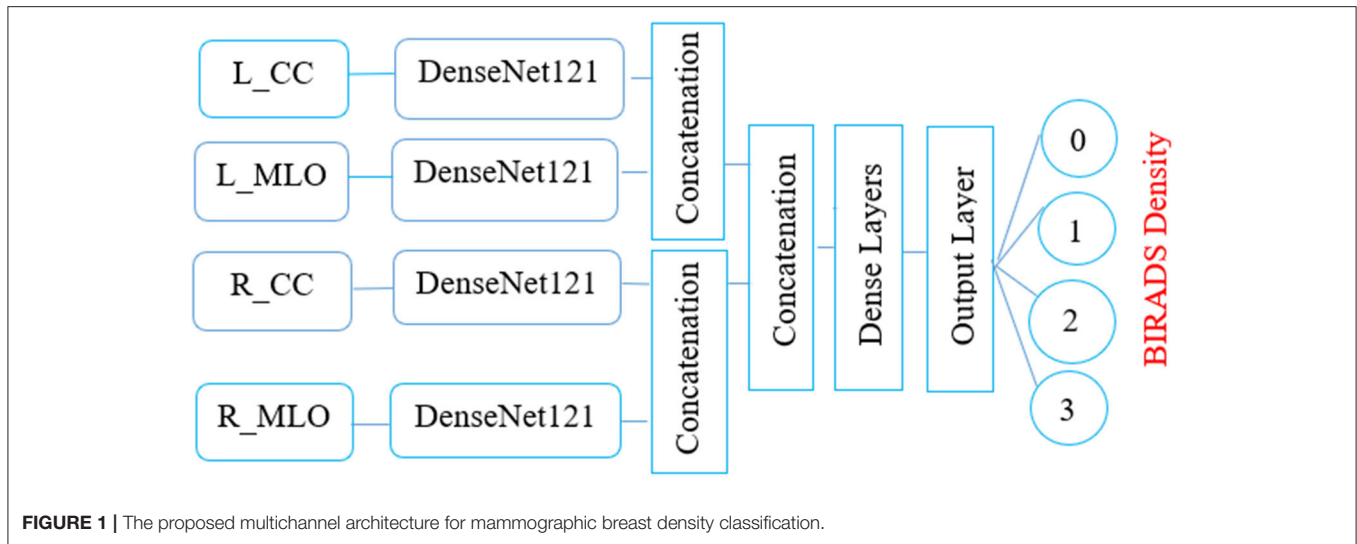
Due to these advantages, it is helpful to use this model without pretraining for medical image analysis.

The elemental aspiration behind this study is to investigate the use of multichannel DenseNet architecture for MBD classification and analyze the proposed architecture compared with different existing methods. **Figure 2** depicts the proposed multichannel architecture for MBD classification.

The key feature of this architecture is classifier performance that is evaluated with 800 digital mammograms among the diverse BIRADS density categories. Results are confirmed exclusively by expert radiologists and objectively with classification accuracy and the area under the curve (AUC). Various sections described in this article are as below. Related study section covers unique existing deep learning algorithms for MBD classification, Study dataset section presents the aspects of the source dataset, Proposed methodology section represents the proposed architecture, and Analysis of experiment and outcomes section introduces the empirical outcomes. Finally Discussion section discusses the future scope and Segment 7 concludes this study.

RELATED STUDY

Mammographic breast density classification is a long-lasting study area due to more thought-provoking and challenging image preprocessing, segmentation, and classification tasks. Successful deep learning classifiers such as ResNet, VGGNet, and GoogLeNet give brand-new pathological imaging and investigation prospects. This section describes the remarkable of such existing methods used for MBD classification. In summation, these classifiers offer better results on different imaging modalities for image classification. Intrinsically, high interreader fluctuations are the prime problem in MBD



assessment. To overwhelm this problem, Ciritsis et al. (19) suggested deep learning architecture with 11 convolutional layers, three fully connected layers, and an output SoftMax layer to distinguish breast density into two classes and four classes. Two expert radiologists marked density ground truth during this classification task. A combination of mediolateral oblique (MLO) and craniocaudal (CC) (left or right) views (20,578) digital mammograms were used to train the proposed model. During the training, to enhance the model’s performance and desist overfitting, the convolutional layers are zero-padded and used a dropout rate of 50%. Batch size and the highest number of epochs used during training are 40 and 120, respectively. Input data segmented as 70% for training and 30% for validation. This model is validated independently on CC and MLO views for two-class classification and achieved an overall classification accuracy of 89.9 and 86.6%.

Predominantly, fat appears darker than fibroglandular tissue in MBD assessment. Hence, the pixel intensity of the histogram acts as an essential feature for classifier training. Wu et al. (20) suggested this technique for MBD classification. In this method, the SoftMax regression classifier is used to learn pixel intensity histograms. Four regular views [left CC (LCC), right CC

TABLE 1 | Input dataset used for testing and validation of proposed algorithm.

BI_RADS Density Class	Number of images
Class-A	200
Class-B	200
Class-C	200
Class-D	200
Total	800

(RCC), left MLO (LMLO), and right MLO (RMLO)] of 2,00,000 screening images were given individually to train the prototype. The proposed architecture is deep and consists of 100 hidden units between input and output layers. All the hidden layers use Rectified Linear Unit (ReLU) as an activation unit. This model is tested for two- and four-class classification and achieved 81.1 and 82.5% classification accuracy. Thus, this model provides moderate classification accuracy despite an extensive dataset. Lizzi et al. (21) suggested a residual convolutional network for MBD classification. The recommended design consists of 41

convolutional layers formed in residual blocks with 2 million parameters. The first block of this architecture consists of a convolutional layer, a batch normalization layer, a leaky ReLU as an activation function, and a two-dimensional (2D) max pooling. A series of four sections consisting of three residual modules uses the output features of the input block. Leaky ReLU with $\alpha = 0.2$ activation functions was used to train the architecture. Categorical cross-entropy as a loss function validates the performance of the model. Maximum accuracy with four-class classification is 78% and the two-class accuracy is 89.4%.

In a comparative study for evaluating the performance of deep learning and transfer learning on a similar dataset, Lee et al. (17) proposed both the approaches on 22,000 mammographic images. In this approach, expert radiologists marked density ground truth for the input images. Initially, the model's training starts with 500 images with the AUC of 0.942 and the final value of the AUC is 0.9882 reported on the whole dataset. Then, the proposed model is cross-validated with the transfer learning (ImageNet) application and obtained the overall AUC of 0.9857. Thus, this study shows that both the deep learning and transfer learning applications provide almost identical results on the equivalent dataset. The requirement of a larger dataset for training is an essential need of deep learning architecture. Shi et al. (22) proposed optimized lightweight deep learning architecture to optimize deep learning performance on smaller datasets. This architecture combines three convolutional neural networks (CNNs), one dense layer, and an output layer with the SoftMax function. Data augmentation is done with additional image processing to increase the numbers in the dataset. This architecture was trained and tested on the 322 mini-Mammographic Image Analysis Society (MIAS) dataset. This architecture provides overall accuracy of 83.6% on four-class classification. The main limitation of lightweight architecture on smaller datasets is the low stability of the network, which may occasionally cause large and significant variations in the accuracy. Data augmentation can enlarge the dataset in this method, but it is still challenging to get well-trained convolutional layers due to little diversity between the original and generated datasets. In the literature, there are two ways that are recorded to enhance the model performance with a smaller dataset, which are generative adversarial network (GAN) and another is transfer learning (23).

Recently, different researchers proposed the concept of transfer learning on a smaller dataset. For example, Kaiser et al. (24) proposed new architecture that can take all the four views of single patients to classify MBD into two-class classification (dense and nondense). For this purpose, the author proposed four-channel VGGNet architecture to extract all the features with average global pooling from input mammograms. Before the classification layer, to concatenate all the input layer features, two dense layers are used. Then, the proposed model is trained with 5-fold cross-validation and recorded 88% classification accuracy with the AUC of 0.954. Finally, subjective assessment is done with a panel of 32 radiologists to compare interobserver variability. In this approach, interobserver variability for breast density assessment is observed even high in two-class classification. Thus, the automated processes for MBD can help to

minimize interobserver variability. Despite different automated approaches, MBD assessment is subjective and consists of intra- and interobserver variations. Objective evaluation of other commercially viable methods consists of mixed evaluation results. Another fundamental limitation is that most of existing deep learning methods need a higher dataset and validated density ground truth; hence, data acquisition becomes difficult for researchers. In addition, mammographic images are vendor specific, making deep learning more robust; training the deep learning model through different vendor-specific samples is required, another bottleneck in MBD classification. All the limitations mentioned above result in moderate objective MBD classification accuracy.

The primary motivation behind this study is to investigate the transfer learning application of DenseNet architecture toward enhancing the classification accuracy of MBD. A significant contribution is the design and development of multichannel architecture to utilize four mammographic views of a single patient.

STUDY DATASET

The intended study utilizes the openly available dataset from digital database for screening mammography (DDSM) (25), consisting of 2,620 samples of various classes labeled as benign, normal, and malignant with confirmed pathogeny data. The aimed algorithm uses 200 Right-MLO, 200 Left_MLO, 200 R_CC, and 200 L_CC views. A total of 800 mammograms are used for training and testing purposes. The ground truth of each class is labeled with the help of specialist radiologists team into four classes as 0, 1, 2, and 3 as a four MBD classes. All the density groups consist of 200 cases of different images (MLO, LMO, R_CC, and L_CC). **Table 1** presents the details of the ground truth input dataset used in this proposed study.

PROPOSED METHODOLOGY

This segment explains the proposed MBD classification technique and divided into three subsections as presented subsequently.

Segmentation of Pectoral Muscle

Instead of using raw images for models' training, the proposed method preprocesses both the LMO and MLO views. Input mammographic images consist of high-intensity marks of artifacts and tags and pectoral muscle. Those fields in the breast region can reduce the MBD classification accuracy (26, 27). The proposed method uses the depth-first search algorithm, our previous study (28), to remove pectoral muscle, artifacts, and tags from all the MLO and LMO views. This preprocessing study helps the model to classify all the BIRADS classes correctly with less input images. Depth-first search (DFS) algorithm identify all the unwanted high-intensity areas (artifacts and pectoral muscle) of MLO and CC views and removes them, which further help the model to make the correct decision. **Figures 3, 4** depict the input and output images after preprocessing.

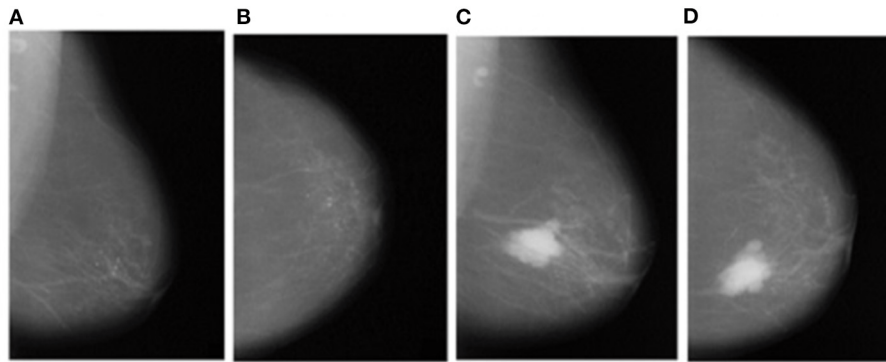


FIGURE 3 | Input raw images- (A) Left_MLO (B) Left_CC (C) Right_MLO (D) Right_CC.

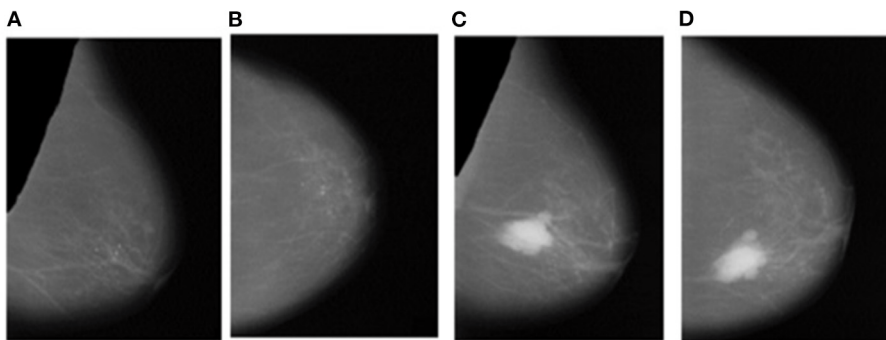


FIGURE 4 | Output Images after segmentation and cropping - (A) Left_MLO (B) Left_CC (C) Right_MLO (D) Right_CC.

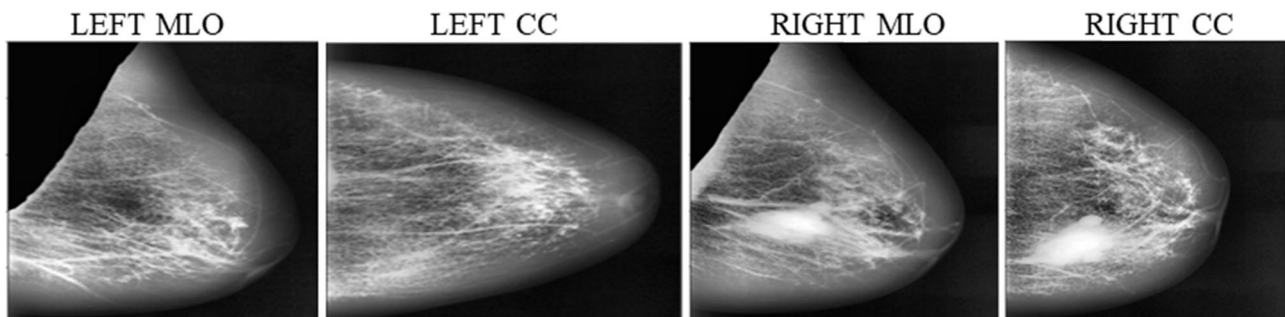


FIGURE 5 | Contrast enhancement of input images.

Contrast Enhancement

Mammographic breast density classification is a function of the density of fibroglandular tissues inside the breast. Contrast enhancement helps to improve the visibility of fibroglandular tissues. Subsequently, it helps to improve the classification accuracy of deep learning models. In literature, many contrast enhancement methods are recorded to enhance the quality of medical images. These methods are “histogram equalization (HE),” “adaptive histogram equalization (AHE),” and “wavelet transform (WT) coefficients” (29–31). But, these documented

practices take ample processing time and are less effective in noise reduction. Another recorded method is the “unsharp masking (USM) method,” which enhances the local contrast inside the image by limiting the global contrast. Still, this technique creates artifacts in the image. Due to this, the image looks artificial; therefore, it is not fitting for the enrichment of the medical images (32–34).

In contrast to gain agreement for the high-frequency element of the image, which is the basic principle behind “adaptive contrast enhancement (ACE),” it consists of limitation in terms of

high processing time (32–38). In MBD classification, local details are more important than global features and reduce processing time; the proposed architecture uses the contrast limited adaptive histogram equalization (CLAHE) algorithm for limited contrast enrichment of fibroglandular tissues (39–41). The first merit of this method helps to minimize the edge shadowing effect and

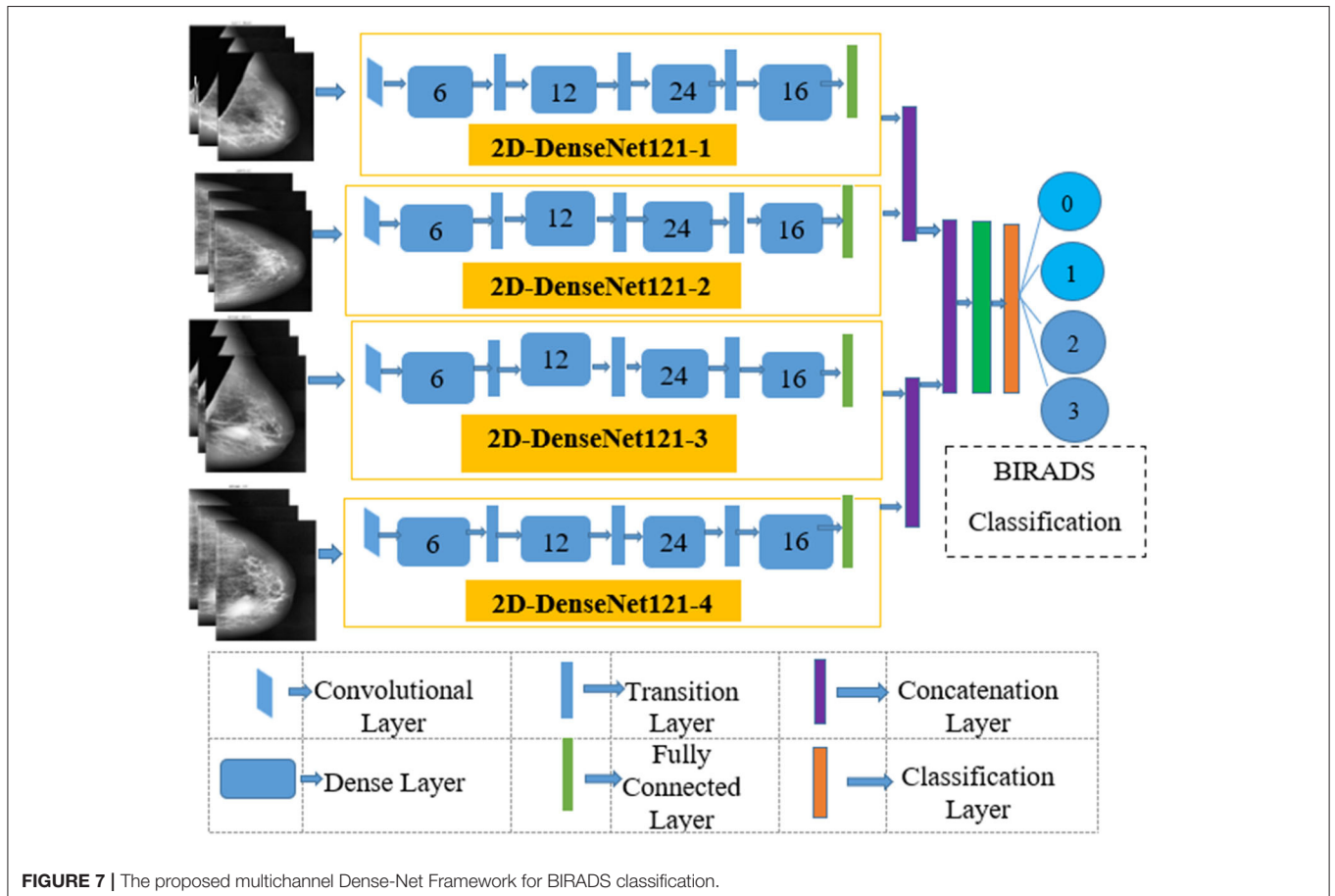
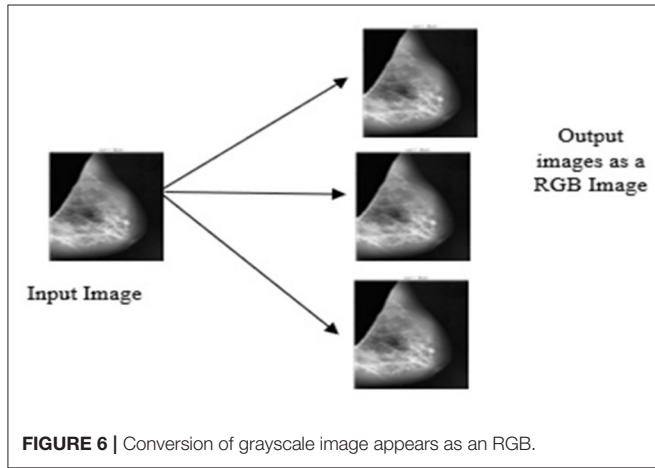
noise produced in homogeneous input digital mammograms. Second, small images known as tiles are used instead of the entire image to perform the CLAHE operation. Hence, contrast enhancement of each tile histogram matches with exponential distribution or Rayleigh distribution. Moreover, to overcome artificial-induced borders, neighboring tiles are connected by bilinear insertion. This advanced CLAHE technique is outlined under:

1. Initially, all the input mammograms are divided into 8×8 non-overlapping contextual fields of equal sizes and later a histogram of various contextual regions is calculated.
2. The clip limit (β) is the threshold parameter used to alter the contrast of the image, which is calculated by Equation (1).

$$\beta = \frac{M \times N}{L} \left(1 + \frac{\alpha}{100} (S_{\max} - 1) \right) \quad (1)$$

Where β is the clip termination, calculated as eight by several experiments, $M \times N$ is the number of pixels in each field, L is the number of gray scales, and α is a clip factor (0–100). It is the highest allowable slope, which is set to be 4 for this analysis.

3. Each histogram is aligned, so that its maximum does not surpass further than the clip boundary.



4. The transformation function, which is described below, is used to modify the histogram.

$$t(r_k) = \sum_{j=0}^k Pr(r_j) \quad (2)$$

$$\text{Where } p_r(r_j) = \frac{n_j}{n} \quad (3)$$

Equations (2) and (3) describe the probability function of input image gray scale value j , n is a total number of pixels in input mammogram image, and n_j is input pixel number of gray value j .

5. The adjacent tiles were joined by bilinear interpolation and the image gray scale values were altered according to the revised histograms.

In this method, the contrast factor is limited to 0.01 to prevent oversaturation of the image, specifically inhomogeneous areas for optimized output. Furthermore, the number of bins for the histogram structure is restricted to 64 over the uniform distribution for contrast enhancing transformations. **Figure 5** depicts the result of contrast enhancement of the CLAHE algorithm on input images.

Multichannel Model Development

This article proposes the feature learning ability of multichannel DenseNet architecture presented by Huang et al. (18) toward MBD classification. The proposed method uses four independent DenseNet architecture as four-channel architecture known as multichannel architecture. This architecture is competent in taking all the four views of an individual patient for the classification of MBD.

Conversion of Gray Scale Image Into RGB

The DenseNet model is pretrained on red, green and blue channel (RGB) images, but the proposed study uses the gray scale image as the input image. To appear gray scale image as an RGB, perform repetition of image array three times due to which the same image appears on the channels. Then, after duplication of the input image, all the input images are resized into $320 \times 320 \times 3$. **Figure 6** depicts the conversion of the gray scale image into a three-channel RGB.

Input Convolutional Layer

The four input channels of the proposed architecture are marked as L_CC, L_MLO, R_CC, and R_MLO. The fundamental merit of this combination is that all the four views of digital mammography are processed concurrently. Each input layer of DenseNet architecture consists of a convolution layer of the kernel of 7×7 with a stride of 2. This convolution operation reduces the input size of the images to $112 \times 112 \times 3$. Input image further passes through a pooling layer of 3×3 maximum pooling with stride 2×2 . Thus, the input layer's convolution and pooling operation reduce the input image size to $56 \times 56 \times 3$ and before passing to the dense blocks. **Figure 7** depicts the proposed multichannel architecture.

Design of DenseNet Neural Structure

The DenseNet architecture consists of different design variants such as DenseNet121, DenseNet169, DenseNet 201, and DenseNet 264. The DenseNet architecture's fundamental merit is the structure of dense layers precisely designed to take care of downsampling and feature concatenation. Therefore, out of four variants, the proposed architecture uses the DenseNet121 network structure, consisting of a combination of dense block layers and transition layers. Thus, the proposed model uses 58 convolutional layers and a growth rate ($k = 12$), including four dense and two transition layers. In addition, the proposed model consists of comparatively fewer parameters hence, saving computational memory and reducing the overfitting.

Dense Block Layer

In four dense blocks, the individual layer is responsible for forming a k -characteristic map after convolution, which also maintains feature maps of each layer are in the same size. K convolution kernels extract all the features from the layers. Parameter k is known as a hyperparameter in DenseNet, which is the growth rate of the network. Each dense layer receives the different inputs from previous layers to reduce computation and enhance the efficiency of the dense block. The dense block internally uses the bottleneck layer (1×1 convolution layer between batch normalization, ReLU, and 3×3 convolution layer).

Transition Layer

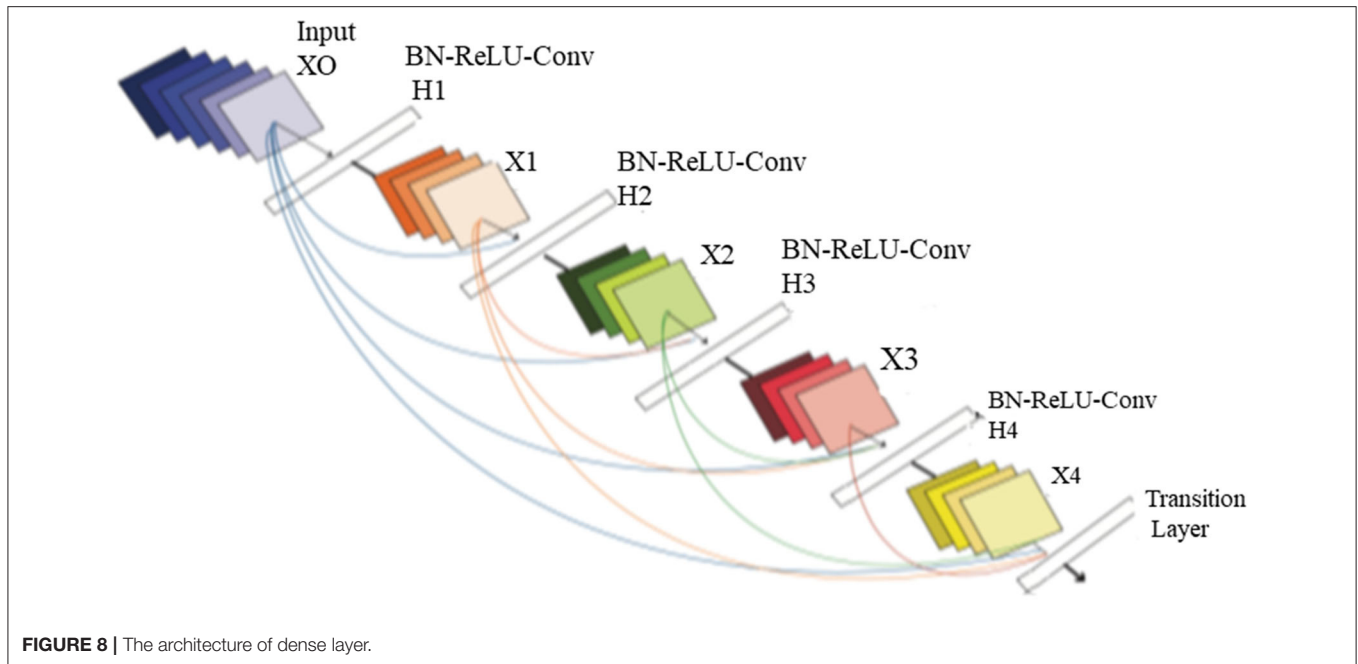
This section consists of a batch normalization layer and a one \times one convolution layer followed by a two \times two average pooling layer. This layer combines two nearby dense block layers to reduce the feature map size. A combination of 4 dense blocks and transition layers converts the image size into $7 \times 7 \times 3$, further provided to the output layer. Each layer connects to the previous stage as an input described by Equation (4):

$$X_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (4)$$

A non-linear transformation function $H_l(.)$ is responsible for combining series output of batch normalization, ReLU, pooling, and convolution operation. **Figure 8** depicts the design architecture of dense layer.

Output Classification Layer

The output layer of the proposed architecture consists of a specific average pooling layer for each channel to extract meaningful features. Extracted features are flattened by the flatten layer and are given to the individual dense layer. MBD features received from four-channel are concatenated together with two concatenation blocks. Subsequently, the third concatenation block joins all the proposed method features together. The three dense layers accept all the features collected together and, finally, the classification layer receives the output of three dense layers for classification. The proposed method uses the SoftMax classifier to classify output into four classes as per the BIRADS Density Classification. **Table 2** presents the specifications of the proposed method.



ANALYSIS OF EXPERIMENT AND OUTCOMES

The experimentally proposed design is trained and tested upon the PyTorch framework on Google Colaboratory, a free online cloud-based Jupiter notebook environment. The proposed method input dataset is not sufficient to split into the train, validate, and test data sets; hence, training and testing of the model are done in two phases. **Figure 9** depicts the image data distribution during training (phase I) and training and testing (phase II).

Phase I

The entire model is trained with stochastic gradient descent (SGD) algorithm using batch sizes 4 and 30 epoch on the whole dataset. SGD is an optimization algorithm that estimates the error gradient for the model's current state with an example of a training set; after this, it updates the weights of the model using backpropagation (35–37). Equation (5) describes weight updating mechanism in SGD algorithm.

$$w_{new} = W_{old} - n \nabla Q_i(w_{new}) \quad (5)$$

Where w_{new} is new weight, W_{old} is weight at previous iteration, n is learning rate and ∇Q_i is weight gradient. The primary merit of this algorithm is that it updates parameters for each training example and performs one update at one time. Thus, SGD is faster and can also learn online. The weight updating step size is the learning rate of the model. The learning rate is a configurable hyperparameter that controls the speed by which the model learns. The initial learning rate for this model is 0.1 (default value) and further divided by ten at 50 and 75% of the total training epochs. The categorical cross-entropy acts as a loss function in

this model, quantifying the difference between four probability distributions. This loss function works well with the SoftMax activation function in multiclass classification. Equation (6) describes the categorical cross-entropy mathematically, which is:

$$C.E. = - \sum_i t_i \log(s_i) \quad (6)$$

Where $C.E.$ is cross-entropy, t_i and s_i are ground truth and the convolutional neural network (CNN) score for each class i in c . **Table 3** presents the setting of different hyperparameters used to obtain the optimized results of the proposed architecture.

During the training on the entire dataset, the best classification accuracy score was 96.35% at 18 epochs with a loss factor of 0.1344. **Figure 10** depicts the training phase results on the dataset as a whole.

Phase II

After training the model on the entire dataset, the proposed model performance is validated by spitting the image dataset in a ratio of 80% as training and 20% as testing. The proposed model performed significantly well on all the BIRADS density classes during the testing phase and recorded the best classification accuracy, 90.00%, with a validation loss of 0.3814. **Figure 11** depicts the outcomes of validation over the training model.

Results Evaluation

The proposed multichannel DenseNet architecture performance is analyzed from the confusion matrix of the model on the test dataset. **Figure 12A** shows the proposed architecture's heat map (confusion matrix) on the test dataset. The heat map helps to analyze which category is correctly classified by the proposed architecture. The main diagonal darker

TABLE 2 | Technical specification of proposed architecture.

No. of channels		04
Layers/Channel	Output size/channel	Block description/Channel
Convolution	112 by 112	Kernel 7 × 7 stride 2
Pooling	56 by 56	3 × 3 Max.Pooling, stride 2 × 2
Dense block 1	56 by 56	[1 × 1 Conv] × 6 [3 × 3 Conv] × 6
Transition 1	56 × 56	Batch normalization layer and a 1 × 1 convolution layer followed by 2 × 2 average pooling layer
	28 × 28	
Dense block 2	28 × 28	1 × 1 Conv] × 12 [3 × 3 Conv] × 12
Transition 2	28 × 28	Batch normalization layer and a 1 × 1 convolution layer followed by 2 × 2 average pooling layer
	14 × 14	
Dense block 3	14 × 14	1 × 1 Conv] × 24 [3 × 3 Conv] × 24
Transition layer 3	14 × 14	Batch normalization layer and a 1 × 1 convolution layer followed by 2 × 2 average pooling layer
	7 × 7	
Dense block 4	7 × 7	1 × 1 Conv] × 16 [3 × 3 Conv] × 16
Classification layer	1 × 1	7 × 7 global average pool

1000D fully connected, SoftMax

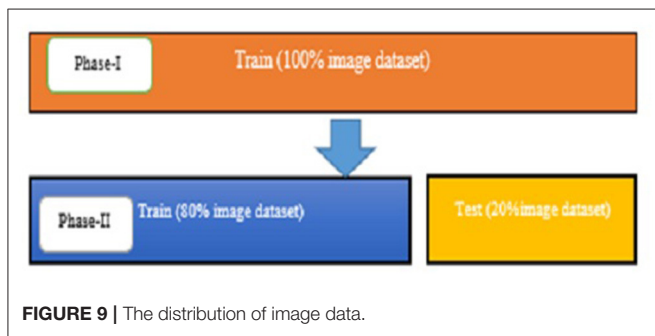


FIGURE 9 | The distribution of image data.

version indicates a better classification rate. Although it is clear that the model is working well in classes A, C, and D, there is some confusion in classifying category class B. Still, this model correctly classifies heterogeneous dense (C) and extremely dense (D), which is the essential bottleneck behind MBD classification. **Figure 12** depict the heat map and the AUC curve, respectively, of the proposed model.

Evaluation of the classification results of the intended architecture is performed in terms of precision, recall, the F1-score, and classification accuracy. Among those parameters, precision is the proportion of samples with optimistic predictions concerning the total number of correct positive samples. The recall ratio of correctly predicted samples to the whole samples

TABLE 3 | Setting of hyperparameters used during experiment.

Hyper parameters	Value
Model	Multichannel-Dense Net
No. of channel	04
Model initial learning rate	0.1
Image size	320 × 320 × 3
Batch size	04
Target labels	Ground Truth
Data augmentation	Not used
Loss function	Categorical cross-entropy
Optimization algorithm	Stochastic gradient decent
Validation parameter	Classification accuracy

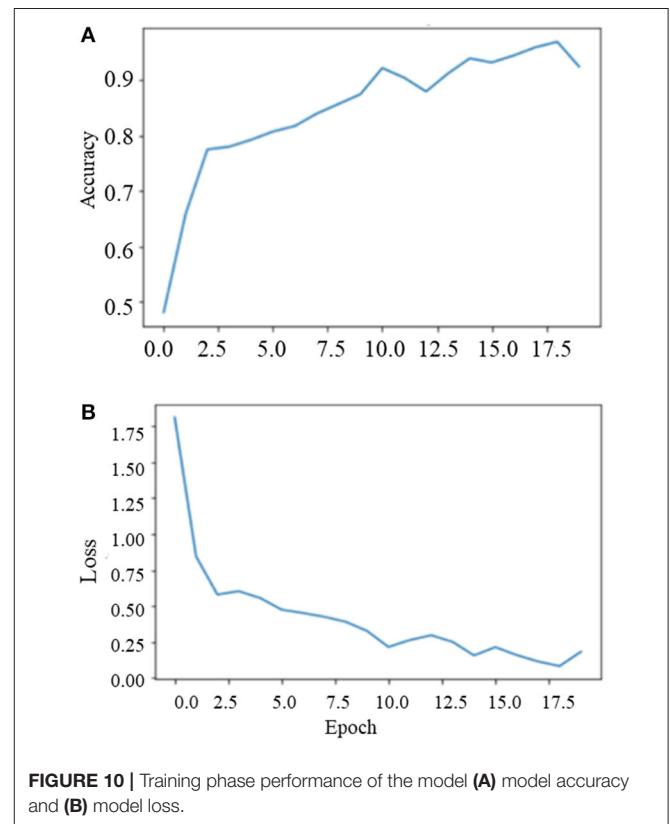


FIGURE 10 | Training phase performance of the model (A) model accuracy and (B) model loss.

and the F1-score are the precision and recall weight. Finally, classification accuracy is the total correct predictions to the total number of samples. Equations (7) to (10) define precision, recall, the F1-score, and classification accuracy, respectively:

$$Acc = \frac{N_A + N_B + N_C + N_D}{N_{All}} \tag{7}$$

$$F1 - score = \frac{2PR}{P + R} \tag{8}$$

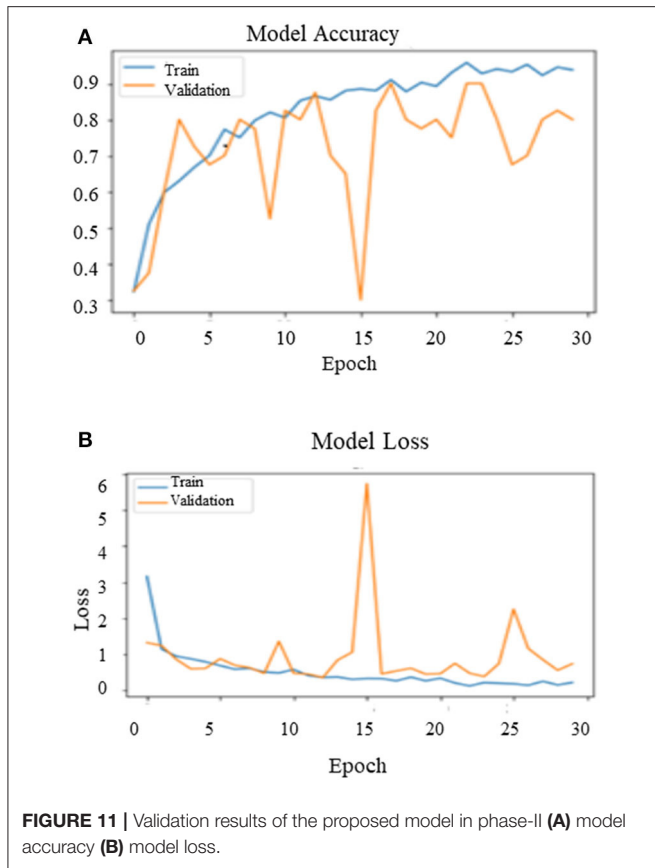


FIGURE 11 | Validation results of the proposed model in phase-II (A) model accuracy (B) model loss.

Where N_{All} is the total number of images and $N_a, N_b, N_c,$ and N_d signify the number of images in MBD classes A, B, C, and D.

$$P = \frac{N_t}{N_i} \times 100 \tag{9}$$

$$R = \frac{N_t}{N_f} \times 100 \tag{10}$$

Where, N_t is the correct number of predictions of a specific category and N_i is all the number of forecasts of a class and indicates the actual number of the category, among them are precision and recall, respectively. The model accuracy is the ratio of the sum of the diagonal elements to all the elements. Thus, it acts as an indicator of the overall prediction of the model. **Table 4** presents the model’s overall performance in detail and variation in precision rate, the recall rate, and the F1-score rate under different categories. The number (or percentage) of accurate positive samples for all the four BIRADS density classes are 92, 75.5, 92.2, and 94.7% of their respective totals. From the results shown in **Table 4**, there is no confusion between classes A and C, B and C, and C and D. The proposed algorithm results are consistent with the results evaluated by the radiologists, which are a positive sign that indicates that deep learning models are helpful for the classification of MBD. Another graphical technique utilized to investigate the realization of computer-aided diagnostic methods is the receiver operating characteristic (ROC), as shown in **Figure 12B**. This curve analysis performance

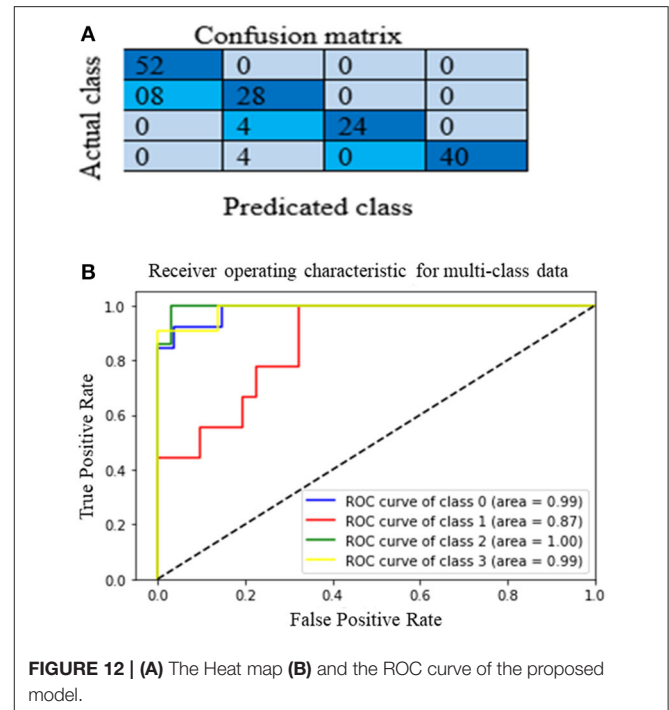


FIGURE 12 | (A) The Heat map (B) and the ROC curve of the proposed model.

of the prototype should be in terms of true positive rate (TP) and false positive rate (FP). The AUC value refers to the area enclosed by the ROC curve in the [0, 1] period and the X-axis. The higher the AUC content, the more reliable the realization of the model. It additionally highlights the ability of the model to differentiate among the classes.

DISCUSSION

Inconsistency in MBD assessment is the fundamental reason behind unnecessary extra screening procedures and cause for patient anxiety (41–43). Due to improved system architecture and hardware capability, the deep learning model can be an alternative for medical image classification. Still, the need for a larger dataset and vanishing gradient are the primary bottleneck issues to obtain state-of-the art results from deep learning models (44).

Advantages of Proposed Method

This study article proposes multichannel DenseNet architecture for MBD classification to investigate the performance of DenseNet architecture on a smaller dataset. The proposed architecture has recorded good classification performance in four-class classification. Furthermore, the visualization results show that the model can distinguish between all the BIRADS density classes, especially in “scattered density” and “heterogeneously dense category of the BIRADS.” Thus, this model can help the radiologists to classify the BIRADS density classes quickly. The main reasons for the excellent performance of this model are as follows:

TABLE 4 | Performance parameter of the proposed method.

BIRADS density classes	Precision	Recall	F1-score	Overall classification accuracy	Overall AUC
Predominantly fatty-class A	1	0.866	0.92	0.9006	0.9625
Fat with some fibro glandular tissue class B	0.77	0.77	0.755		
Heterogeneous dense-class C	0.857	1	0.922		
Extremely dense-class D	0.90	1	0.947		

TABLE 5 | Comparative status of the proposed method with current state-of-the-art methods.

References	Dataset	Proposed method	Classification accuracy
Wu et al. (20)	2,00,000	A deep convolutional network with 100 layers.	0.825 on Four views
Cirtsis et al. (19)	20,578	A deep convolutional network with 11 layers and performed analysis separately on CC and MLO views.	0.897 On CC views and 0.866 on MLO views.
Kaiser et al. (24)	8,150	A multichannel architecture with transfer learning by VGG-Net.	0.88 on all four views
Shi et al. (22)	322	A light-weight deep learning architecture with 3 convolutional layers.	0.836 On MLO views.
Deng et al. (36)	18,157	A single channel architecture with transfer learning by Dense Net 121 combined with SE-Attention network.	0.9179 on all Four views
Proposed method	800	A multichannel architecture with transfer learning with Dense Net 121	0.90 on Four views

1. Instead of using raw images, the proposed architecture uses preprocessed images. Hence, there are no high-density areas such as pectoral muscle and tags on mammograms, which are helpful to increase the classification accuracy of the proposed architecture.
2. The proposed method uses a contrast enhancement technique to improve the quality of training data.
3. The model contains four DenseNet branches, which extract the features of mammograms from four views of the single patient, so that the network can focus on a broader range of spatial information.
4. Due to multichannel architecture, it is possible to process all the views of a single patient simultaneously.
5. Subsequently, with multichannel and multiview architecture, it is probable to consolidate all the features collectively; consequently, the performance of the intended model is found more reliable than single-view classification.

Comparison With Existing Methods

While there are variations in related datasets and evaluation methods, the straightforward comparison is challenging to researchers. This section compiles the proposed algorithm's comparison state with existing classifications. To study the interobserver variation in MBD assessment, N Kaiser et al. (24) proposed the novel multichannel VGG architecture. This approach uses a total of 8,150 digital mammograms, divided into 600 cases. This method recorded 88% two-class classification accuracy (dense and nondense) with the AUC of 0.954. Besides, these results are also compared with the 32 individual radiologist's panel's density ground truth. This study reveals that the deep learning approach performs better than average radiologists. Thus, we only need to refine the deep learning model for MBD classification. However, the

fundamental limitation of this method is that the gradient flows from the final layer to the initial layer; hence, vanishing gradient problem takes place, which increases training time and reduces the classification accuracy. In the proposed method, due to DenseNet architecture, all the layers are directly connected in feed-forward nature, acting as an effective solution for vanishing gradient and reducing training time. Thus, the results of the proposed algorithm outperform this method on a smaller dataset.

Another method directly comparable to the proposed method is the optimized lightweight deep learning architecture proposed by Shi et al. (22). The elemental focus of this method is to overcome the requirement of the larger dataset and vanishing gradient problem of the deep learning algorithm. This method combines three CNN layers, one dense layer, and an output layer to classify MBD. This architecture is tested on the 322 mini-MIAS dataset with different data augmentation techniques and recorded 83.6% classification accuracy. However, due to the smaller dataset, this architecture has limitations regarding moderate classification accuracy and low stability of the network. Therefore, the proposed method used the concept of transfer learning and multichannel architecture to overcome these limitations. As a result, the proposed model outperforms this method in classification accuracy on a smaller dataset. **Table 5** provides the comparative state of the proposed method with other different existing methods.

Limitations and Future Study

Even though the proposed method has improved the classification performance of the BIRADS density classes, some issues still need to be addressed. First, this study uses a smaller amount of image data and no image enhancement strategies are used to expand the dataset. Hence, model performance, especially stability during validation, is affected

due to limited image data and model found it a little confusing to classify classes A and B. Therefore, in future study, data enhancement techniques will improve the model's performance. Second, the proposed study addresses only one type of dataset; hence, this approach does not address the robustness of the model. Future study will address the robustness of the model by training the model with different vendor-specific image datasets and testing results of all the mentioned state-of-the-art methods with the proposed method with the same work environment. The proposed study will be undoubtedly helpful in addressing the issues mentioned above.

CONCLUSION

In summary, the primary objective behind this study is to classify MBD as per the BIRADS classification. This study proposes the novel approach of multichannel architecture with DenseNet121 for the objective assessment of MBD. The proposed framework uses the four views of a single patient to enhance feature learning ability through a multiview approach. In the method, image contrast enhancement and preprocessing of the input image are implemented to enhance the condition of the training image data. The input images are processed through multichannel architecture to extract and fuse all the features. Analysis of the results suggests that the proposed model successfully distinguishes between all the BIRADS density classes, but is predominantly found superior in the two most distinctive and challenging BIRADS categories: "BIRADS_C" and "BIRADS_D." Classification

REFERENCES

- Advani P, Moreno-Aspitia A. Current strategies for the prevention of breast cancer. *Breast Cancer Targets Therapy*. (2014) 6:59–71. doi: 10.2147/BCTT.S39114
- Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol*. (2006) 24:2137–50. doi: 10.1200/JCO.2005.05.2308
- Sapate S, Talbar S, Mahajan A, Sable N, Desai S, Thakur M. Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms. *Biocybern Biomed Eng*. (2020) 40:290–305. doi: 10.1016/j.bbe.2019.04.008
- Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee J, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020). 8:218–25. doi: 10.3389/fpubh.2020.00357
- Dhanamjayulu C, Nizhal U, Maddikunta P, Gadekallu T, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process*. (2021) 16:647–58. doi: 10.1049/ipr.2.12222
- Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia System*. (2021) 1:1–15. doi: 10.1007/s00530-021-00774-w
- Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng*. (2021) 18:8444–61. doi: 10.3934/mbe.2021418
- Dungan JS. Mammographic density and the risk and detection of breast cancer. *Obstet Gynecol Women Health*. (2008) 1:214–5. doi: 10.1016/S1090-798X(08)79014-3

accuracy of the proposed model is recorded at 96.67% during training and 90.06% during testing and the average AUC of 0.9625. The introduced design consists of some weaknesses discussed and will be addressed in future study; with certain modifications, the proposed method is suitable for application in clinical workflow in breast cancer screening to avoid false recalls.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

The authors are grateful to the Taif University Researchers Supporting Project number (TURSP-2020/36), Taif University, Taif, Saudi Arabia and gratefully acknowledge Dr. Yogesh G. Yadav and Dr. Atul E. Sawant, Specialist Radiologist, Lifecare Hospital, Musaffah, Abu Dhabi for their valuable and timely help in marking ground truth on some difficult cases of mammogram from DDSM dataset and providing subjective evaluation of same.

- Lam PB, Vacek PM, Geller BM, Muss HB. The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. *Cancer*. (2000) 89:369–75. doi: 10.1002/1097-0142(20000715)89:2<369::aid-cncr23>3.0.co;2-j
- Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S, Thakur M. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. *Comput Methods Programs Biomed*. (2018) 163:1–20. doi: 10.1016/j.cmpb.2018.05.017
- Rice MS, Bertrand KA, Lajous M, Tamimi RM, Torres G, López-Ridaura R, et al. Reproductive and lifestyle risk factors and mammographic density in Mexican women. *Ann Epidemiol*. (2015) 25:868–73. doi: 10.1016/j.annepidem.2015.08.006
- Pawar S, Sapate S, Sharma K. Machine learning approach towards mammographic breast density measurement for breast cancer risk prediction: an overview. *Proc ICAST*. (2020) 2020:1–14. doi: 10.2139/ssrn.3599187
- Alyami H, Alosaimi W, Krichen M, Alroobaea R. Monitoring social distancing using artificial intelligence for fighting COVID-19 virus spread. *Int J Open Source Softw Proc*. (2021) 12:48–63. doi: 10.4018/IJOSSP.2021070104
- Arul R, Alroobaea R, Tariq U, Almulih AH, Alharithi FS, Shoaib U. IoT-enabled healthcare systems using block chain-dependent adaptable services. *Pers Ubiquitous Comput*. (2021) 1:1–15. doi: 10.1007/s00779-021-01584-7
- Devi SS, Vidivelli S. Classification of breast tissue density in digital mammograms. In: *Proceedings of ICIECS*. Coimbatore (2018). p. 1–7. doi: 10.1109/ICIECS.2017.8276139
- Srivastava RK, Greff K, Schmidhuber J. *Training Very Deep Networks*. *Advances in Neural Information Processing Systems*. (2015). p. 2377–85.
- Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. *Med Phys*. (2018) 45:1178–90. doi: 10.1002/mp.12763

18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of CVPR*. Los Alamitos, CA (2017). p. 2261–9. doi: 10.1109/CVPR.2017.243
19. Ciritis A, Rossi C, De Martini I, Eberhard M, Marcon M, Becker AS, et al. Determination of mammographic breast density using a deep convolutional neural network. *Br J Radiol.* (2019) 92:20180691. doi: 10.1259/bjr.20180691
20. Wu N, Geras K, Shen Y, Su J, Kim SG, Kim E, et al. Breast density classification with deep convolutional neural networks. In *Proceedings of ICASSP*. Calgary, AB (2018). p. 6682–6. doi: 10.1109/ICASSP.2018.8462671
21. Lizzi F, Laruina F, Oliva P, Retico A, Fantacci ME. Residual convolutional neural networks to automatically extract significant breast density features. *Commun Comput Inform Sci.* (2019) 1089:28–35. doi: 10.1007/978-3-030-29930-9_3
22. Shi P, Wu C, Zhong J, Wang H. Deep learning from small dataset for bi-rads density classification of mammography images. In: *Proceedings of ITME*. Qingdao (2019). p. 102–9. doi: 10.1109/ITME.2019.00034
23. Bengio Y. Learning deep architectures for AI. *Found Trends Mach Learn.* (2009) 2:1–27. doi: 10.1561/22000000006
24. Kaiser N, Fieselmann A, Ritschl L, Kappler S, Ravikumar N, Ritschl L, et al. Mammographic breast density classification using a deep neural network: assessment on the basis of inter-observer variability. *SPIE-Intl Soc Opt Eng.* (2019) 1:23. doi: 10.1117/12.2513420
25. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. *International workshop on digital mammography*. Toronto, ON: *Med Physics in Proceedings of IWDM* (2001).
26. Sapate S, Talbar S. An overview of pectoral muscle extraction algorithms applied to digital mammograms. *Stud Comput Intell.* (2016) 651:19–54. doi: 10.1007/978-3-319-33793-7_2
27. Pawar SD, Sharma KK, Sapate SG. *Design of Intelligent Applications Using Machine Learning and Deep Learning Techniques* 1st ed. London: Francis and Taylor; CRC Press. (2021). p. 125–43. doi: 10.1201/9781003133681-8
28. Pawar SD, Sharma KK, Sapate SG, Yadav GY. Segmentation of pectoral muscle from digital mammograms with depth-first search algorithm towards breast density classification. *Biocy Biomed Eng.* (2021) 41:1224–41. doi: 10.1016/j.bbe.2021.08.005
29. Zeng M, Li Y, Meng Q, Yang T, Liu J. Improving histogram-based image contrast enhancement using gray-level information histogram with application to X-ray images. *Optik.* (2012) 123:511–20. doi: 10.1016/j.ijleo.2011.05.017
30. Öktem H, Egiazarian K, Niittylahti J, Lemmetti J. An approach to adaptive enhancement of diagnostic X-Ray images. *Eur J Appl Sig Process.* (2003) 5:430–6. doi: 10.1155/S1110865703211069
31. Deng G. A generalized unsharp masking algorithm. *IEEE Transac Image Process.* (2011) 20:1249–61. doi: 10.1109/TIP.2010.2092441
32. Chang DC, Wu WR. Image contrast enhancement based on a histogram transformation of local standard deviation. *IEEE Transac Med Imaging.* (1998) 17:518–31. doi: 10.1109/42.730397
33. Bourouis S, Channoufi I, Alroobaea R, Rubaiee S, Andejany M, Bouguila N. Color object segmentation and tracking using flexible statistical model and level-set. *Multimedia Tool Appl.* (2021) 80:5809–31. doi: 10.1007/s11042-020-09809-2
34. Hedabou M. *Computer and Cyber Security: Principles, Algorithm, Applications and Perspective* 1st ed. London: Francis and Taylor Publisher; CRC Press (2018).
35. Toulis P, Airoidi EM. Asymptotic and finite-sample properties of estimators based on stochastic gradients. *Ann Stat.* (2017) 45:1694–727. doi: 10.1214/16-AOS1506
36. Deng J, Ma Y, Li D-A, Zhao J, Liu Y, Zhang H. Classification of breast density categories based on SE-Attention neural networks. *Comput Method Program Biomed.* (2020) 193:105489. doi: 10.1016/j.cmpb.2020.105489
37. Iggaramen Z, Hedabou M. FADETPM: Novel approach of file assured deletion based on trusted platform module in lecture notes in networks and systems. *Springer Verlag.* (2017) 49:49–59. doi: 10.1007/978-3-319-97719-5_4
38. Azougaghe A, Hedabou M, Belkasm M. An electronic voting system based on homomorphic encryption and prime numbers. In: *Proceedings of ICIAS*. Kuala Lumpur (2016). p. 140–5. doi: 10.1109/ISIAS.2015.7492759
39. Bentajer A, Hedabou M, Abouelmehdi K, Iggaramen Z, El Fezazi S. An IBE-based design for assured deletion in cloud storage. *Cryptologia.* (2019) 43:254–65. doi: 10.1080/01611194.2018.1549123
40. Mukhtar H, Rubaiee S, Krichen M, Alroobaea R. An Iot framework for screening of covid-19 using real-time data from wearable sensors. *Int J Environ Res Public Health.* (2021) 18:56–62. doi: 10.3390/ijerph18084022
41. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm. *J Circuits Syst Comput.* (2022) 31:2250010. doi: 10.1142/S0218126622500104
42. Arikumar KS, Prathiba SB, Alazab M, Gadekallu TR, Pandya S, Khan JM, et al. FL-PMI: federated learning-based person movement identification through wearable devices in smart healthcare systems. *Sensors.* (2022) 22:1377. doi: 10.3390/s22041377
43. Rajput DS, Basha SM, Xin Q, Gadekallu TR, Kaluri R, Lakshmana K, et al. Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. *J Ambient Intell Humaniz Comput.* (2021) 10:1–12. doi: 10.1007/s12652-021-03154-4
44. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput Sci.* (2021) 7:e390. doi: 10.7717/peerj-cs.390

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Pawar, Sharma, Sapate, Yadav, Alroobaea, Alzahrani and Hedabou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.