Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-023-01871-x

ORIGINAL ARTICLE

A combined deep CNN-lasso regression feature fusion and classification of MLO and CC view mammogram image

V. Sridevi[1] · J. Abdul Samath[2]

Received: 11 September 2021 / Revised: 26 January 2022 / Accepted: 24 January 2023

Abstract Breast cancer is the most frequent disease among women, and it is a serious threat to their lives and well-being. Due to high population expansion, automatic mammography detection has recently become a critical concern in the medical industry. The emergence of computer-assisted systems has aided radiologists in making accurate breast cancer diagnoses. An automated detection and classification system should be implemented to prevent breast cancer from spreading. Breast densities vary widely among women, which causes missed cancers. In the case of breast density, the deep CNN algorithms can significantly reduce radiologist workload and improve risk assessment. The goal of this paper is to offer a deep learning strategy for identifying MLO and CC views of breast cancer as malignant, benign, or normal using an integration of deep convolutional neural networks (CNN) and feature fusion of LASSO (Least Absolute Shrinkage and Selection Operator) regression. The proposed method comprises pre-processing, data augmentation, feature extraction, feature fusion, and classification. The generated features were fed into LASSO regression for the best prediction in this system, which utilized CNN for feature extraction. The fused features were then transferred to CNN's fully connected layer for mammography classification. In our experiment, the publically available dataset CBIS-DDSM (Curated Breast Imaging Subset of DDSM) was employed. The proposed work gained an accuracy of 99.2%, specificity of 98.7%, AUC of 99.8%, sensitivity of 99.4%, and FI-score of 98.7%, which is higher than multi view CNN without a feature fusion based system.

Keywords CNN · LASSO · Mammogram cancer classification · Regression · Feature fusion · Breast cancer

Abbreviations

CNN	Convolution neural network
MLO	Medio-lateral
CC	Cranio-caudal
DDSM	Digital database of screening mammography
CBIS-DDSM	Curated breast imaging subset of DDSM
LASSO	Least absolute shrinkage and selection operator
CAD	Computer aided diagnosis
ROC	Receiver operating characteristic
AUC	Area under ROC curve
SVM	Support vector machine
ReLU	Rectified linear unit
ADASYN	Adaptive synthetic sampling
TP	True positive
TN	True negative
FP	False positive
FN	False negative
FC	Fully connected layer

1 Introduction

Due to the architecture of the human body, Women are easily influenced to breast cancer. According to the American Cancer Society, the number of new breast cancer

✉ V. Sridevi
vissridevi@gmail.com

J. Abdul Samath
abdul_samath@yahoo.com

1 Department of Computer Science, PSG College of Arts and Science and Research Scholar of Govt. Arts College, Udumalpet, India

2 Department of Computer Science, Chikkanna Government Arts College, Tirupur, India

Published online: 27 February 2023

cases in women will reach 276,480 in 2020, accounting for around 30% of new cancer cases (Siegel 2020; Sridevi 2019). Furthermore, statistics report that new incidences of breast cancer have increased at a rate of roughly 0.3 percent each year since 2004, posing a major threat to women's lives and health.^#^#^#^#

Breast cancer is one of the main causes of mortality for women worldwide (Bray et al 2011; Gao et al. 2002; Munir et al. 2019; Cancer facts and figures 2019). It happens as a result of unregulated breast cell proliferation. These cells are generally found in clusters. Different imaging methods can detect malignancies in the breast area. Masses and calcifications in the breast can be signs of breast cancer (Dheeba 2015). Masses can be classified as benign or malignant masses based on their shape. It are possible for a mass to be benign or cancerous. The shapes of benign and malignant tumours differ, with benign tumours having a round or oval shape and malignant tumours having a somewhat rounded shape with an uneven contour. Furthermore, the cancerous mass will seem whiter than the surrounding tissue. (Tang et al. 2009).

Digital mammography is a common breast cancer diagnostic method used all over the world. Computer-aided diagnosis (CAD) is deployed to accurately diagnose a variety of ailments. It aids medical experts in analyzing and determining the phases of various diseases (Zhang et al. 2020a, b). In general, deep learning CAD systems concentrate on CNNs, that are the most used models for medical image analysis and identifying cancer accurately (Xu 2015; Swiderski 2016). In CNNs, the feature extraction may be automated, which is an essential representation of the network, reducing human involvement.

A screening mammogram is normally performed by scanning the two views of the breast: the medio-lateral oblique (MLO) projection and the cranio-caudal (CC) projection. The MLO projection shows the breast at an angle of 45-degrees. The CC projection exposes the breast from the top down. The diagnostic method includes both views.

The process of fusing two or more objects into a single entity is known as fusion. Multisource information fusion, in particular, strives to generate precise and comprehensive unified estimations of complex circumstances. By efficiently aggregating key evidence among the huge, diversified, and often contradictory information gathered from many comparable or dissimilar sources, the estimates could be more accurate in general (Amira Jouirou et al. 2019).

Researchers have identified uses for deep learning-based CAD systems, such as lung cancer (Kumar et al. 2015; Hua et al. 2015), breast cancer (Wang et al. 2016), and Alzheimer's disease (Suk and Shen 2013; Liu et al. 2014; Suk et al. 2014). In terms of detection and classification, they showed the best results. Breast cancer has been identified, detected, diagnosed, and its risk assessed using deep learning (Kallenberg et al. 2016; Cheng et al. 2016).

The goal of this study is to develop a deep learning based system that integrates CNN and LASSO regression to diagnose breast cancer from mammography pictures automatically. The proposed framework is based on two-view based feature extraction that is evaluated on DDSM mammogram images. The proposed system comprises five important phases of the CAD system: preprocessing, data augmentation, feature extraction, feature fusion, and classification. Preprocessing is used to improve the quality and remove pectoral muscle of the mammogram image. In order to improve model generalization and training set, augmentation approaches have been applied to images. One of the key stages in image processing is feature extraction, which greatly improves the classification rate. The efficiency of the classification and CAD systems is improved by the extraction of relevant features from the mammogram image in an efficient manner. Here, deep CNN is used separately for MLO and CC views of same breast for feature extraction. Next the combining of all features extracted from the two views was performed using the feature fusion strategy. LASSO regression is depicted for feature fusion. The fused features were given into the fully connected layer of CNN to classify the MLO and CC views of mammogram images as malignant and benign. In our experiment, the publically available dataset CBIS-DDSM was employed. In this system, a dataset of 3362 breast images was employed, including 974 mammography cancer images. To test the performance of the suggested system, thorough experimental research is presented in respect of accuracy, sensitivity, specificity, F1-score, and receiver operating characteristic (ROC) curve. The performance of the proposed combined deep CNN and LASSO regression system was calculated with the following parameters: accuracy, specificity, sensitivity, ROC curve, and F1 Score.

2 Related works

There have been various methods developed to address mammogram cancer classification problems, which has two categories: traditional methods and deep learning methods. Wu et al. (2019) developed deep convolutional neural network trained and tested on over 200,000 examinations The strategy employed in this paper is to generate heatmaps as adding extra input channel to the breast-level classifier. This model is capable of learning both local and macroscopic properties such as breast symmetry. When evaluated on the screening population, their network gets an AUC of 0.895 in prediction.

Lai (2018) presented amodel that integrates a deep CNN with certain classical features, called the Coding Network with Multilayer Perceptron. They tested the method on medical imaging datasets, HIS2828 and ISIC2017. They achieved 90.2% overall classification accuracy, which is greater than the recent approaches.

Nasir Khan et al. (2019a, b) offered a Multi-View Feature Fusion based CADx system for mammography classification. This system outperformed the single view-based approach. They gained an AUC of 0.84 for malignant and benign cancers. Saleem et al. (2020) provided a cheat sheet with the data augmentation to the proposed CNN. Before handing the mammography to CNN, the cheat sheet recognizes artificial patterns easily. The results showed an accuracy of 92.1, a sensitivity of 91.4, and a specificity of 96.8.

Sun et al. (2019) developed an architecture based on CNN that incorporates two subnetworks of MLO and CC view of mammogram images to extract features. They also included a dilated convolution layer to boost the diversity of breast mass by expanding the respective fields. They achieved an accuracy of 0.8156, which is 1.38% higher than that of the Multi view Convolution Neural Network.

Abdullah-Al Nahid et al. (2018) categorized a series of mammography images using novel deep neural network techniques. They used combined CNN and Long Short Term Memory for classification of mammogram cancer and SVM for decision-making purposes. They gained 91% accuracy and a 91% precision value.

Sridevi et al. (2020) utilized the adaptive K means clustering method to segment the two views of a mammogram image. In the feature extraction stage, the features of CC and MLO views are extracted using a mix of k-means clustering and the Gabor filter. Finally, the Knn classifier is used to classify the mammography image. The AUC for MLO-N cases using KCM-GF features was 95.24, while the AUC for GLCM features was 90.01.

Taha Muthar et al. (2021) focused on the ADASYN (Adaptive Synthetic Sampling) algorithm to eliminate the uneven data problem and also lower the error rate. Then the feed forward network is employed for data classification. The proposed technique demonstrated its robustness with an increasing accuracy of 99.1%.

3 Methodology

Our aim is to design an automatic mammogram cancer classification algorithm that is based on two views of the mammography image. We propose the benign/malignant decision task in a three step approach. In first step, distinct CNN architectures are involved to the CC and MLO views to extract the relevant features. In the second step, the fused features are determined by concatenating the feature vectors of both the views. In the third step, the classifier is resolved to make a final decision of cancer. The overall system architecture for the detection and classification of mammogram images is depicted in Fig. 1.

Raw MLO and CC View mammography images were first run through a preprocessing that included normalization in order to increase image quality by deleting the pectoral muscle region. The preprocessed data was then augmented using affine transformation techniques which include different operations such as flipping, rotation, or translations, to raise the number of training data. Then, the augmented data was portioned into a training set and a testing set. Then the training data set is given for feature extraction, which includes the convolutional and pooling layers of CNN. The extracted features are concatenated using feature fusion, which includes LASSO regression to reduce the discriminative feature and redundancy. The output of fusion was fed into a fully connected layer with a softmax

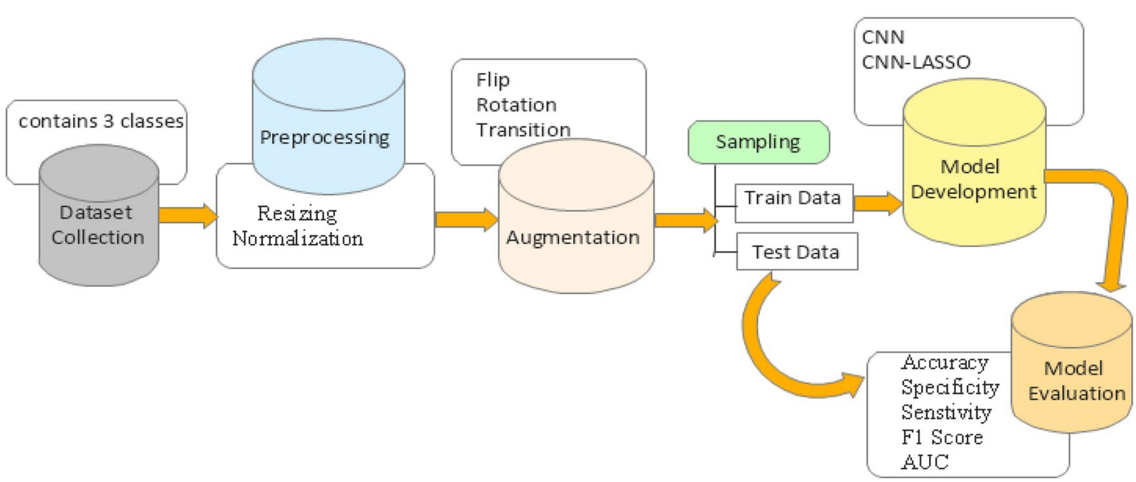

Fig. 1 Overall system architecture of proposed framework

activation function for the classification of mammogram images. There are 23 layers in the network, including ten convolutional layers with a kernel, ten max pooling layers, one LASSO regression layer, and one FC layer with the softmax activation function.

The proposed architecture of CNN-LASSO model development is depicted in Fig. 2. We employed CNN training for each mammographic view to discover masses in the image. In order to detect matching pairings of masses on the CC and MLO perspectives, the results of two CNNs are fused. Each CNN is divided into five sections, each of which has one convolution layer followed by a maxpooling layer. The convolutional layer is frequently referred to as a filter or a kernel that is convolved with the input as a set of weights. The filtering is carried out on a region-by-region basis. The activation function of the Rectified linear unit (ReLU) is applied to the convolution layer to incorporate nonlinearities into the model via the resulting output map. The pooling layer separates the input image patch into a series of rectangles that do not overlap. Through the pooling layer, the number of parameters and the quantity of processing in the network can be minimized, allowing over-learning to be controlled.

The dataset is divided into training and validation sets for CNN training. The validation set is used to measure how well the trained CNN performs after each epoch, whereas the training set is used to train the network and update its weights. The number of times the algorithm analyses the complete dataset is referred to as an epoch. There were 150 epochs in the training procedure. We adjusted the initial learning rate to 0.0001 based on training performance.

The size of the RGB image of the input layer of the CNN architecture is $256 \times 256 \times 1$. The first convolutional layer's input dimension is $256 \times 256 \times 1$. In this case, each block consists of one convolutional layer with 16 channels and the same padding, followed by a 3×3 maxpooling layer with a stride 2×2. The outputted feature map from the last maxpooling layer is $8 \times 8 \times 256$ pixels after the stack of convolutional and maxpooling layers. The flatten layer was added to create a $1 \times 16{,}384$ feature vector from each view of the image. After the fifth block of CNN training, we add a feature fusion layer to create our modified model, CNN-LASSO. A feature fusion layer was added to create a 1 25,088 feature vector. In this framework, the L1 regularization method with the penalty of loss function is used. The output of a feature fusion layer is a $1 \times 16{,}384$ feature vector. The fully connected layer (FC) is the final layer of CNN-LASSO. It's commonly referred to as a decision layer. The classification vector acquired from FC is normalized using a softmax activation function at the end. For each of the three classes, the FC Layer outputs three channels. The normal output category is 0, the benign output category is 1, and the malignant output category is 2.

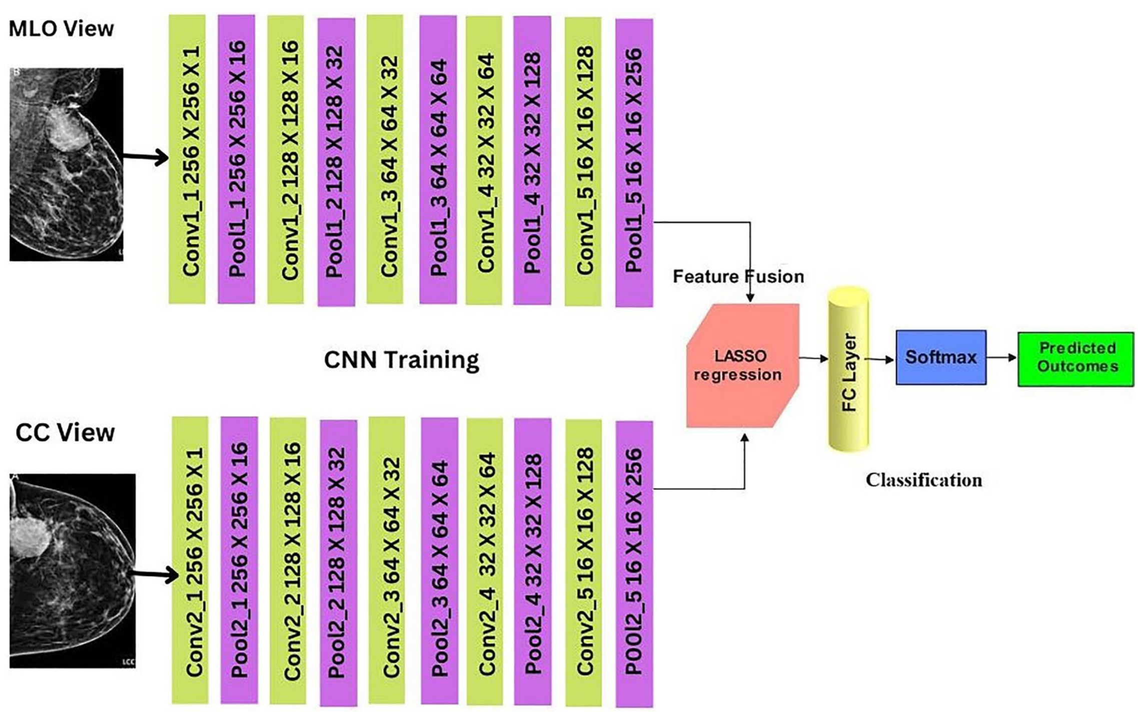

Fig. 2 Architecture of CNN-LASSO model development

3.1 Preprocessing

Preprocessing operations are needed, which include normalization to improve the quality of the image by removing the pectoral muscle region from the image. Compared to other neural networks, CNN requires much fewer preprocessing operations. The collected input image is converted into a grey scale image and applied with a spatial linear filtering algorithm for the removal of noise from the mammogram image. Linear spatial filters allow enhancing images and removing noise from them in a variety of ways. The filters include Mean and Gaussian for smoothing, gradient operators such as sobel, prewitt and canny filters for edge detection and basic highpass spatial filter and laplacian filter for sharpening. The contrast enhancement approach is then used on the denoised image, which is then input into histogram equalization and the sobel edge mask. This process is called gradient magnitude. Finally, a morphological operation is performed on the gradient magnitude image to remove the pectral muscle from the mammogram image. The visualization of preprocessed image is depicted in Fig. 3.

3.2 Data augmentation

Increasing the amount of data used to train deep learning neural network models may result in more skillful models, and augmentation techniques can allow the fitted models to generalize their learning to new images by creating variations of the images. In order to improve model generalization and raise the number of training sets, augmentation approaches have been applied to images. Here, affine image transformations, that are rotation, flipping, or transitions, are followed in the augmentation of the data set. The primary parameters validated for image augmentation were random horizontal and vertical flips with a range of 0.2, random rotation of0-360degrees, zoom with a 0.2 ranges, and transition of 0.2.

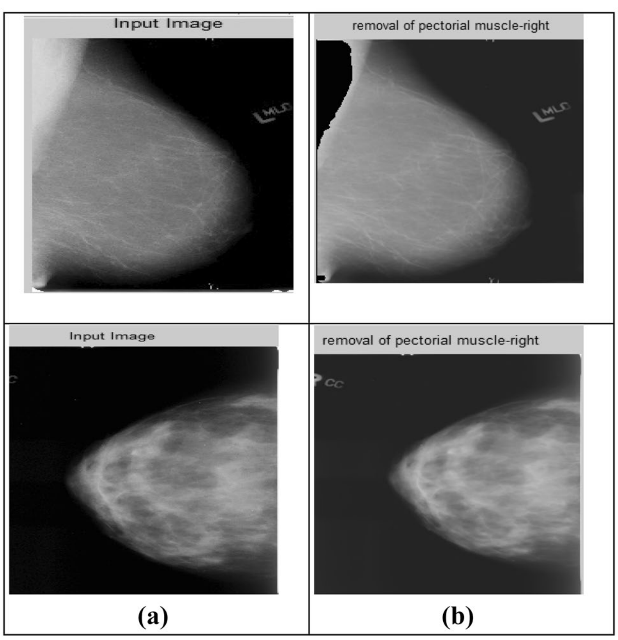

Fig. 3 Preprocessing output

3.3 Feature extraction

Feature extraction can be performed separately for MLO and CC views of same breast. By collecting high-level features from the mammography picture, a convolutional layer constricts the given input image and makes feature maps corresponding to each feature detector. Many factors influence each convolutional layer, including the input size, kernel size, padding, and stride.

In each view of the mammography image, a stack of five convolutional layers and five pooling layers is used to extract features. The size of the input picture supplied to the CNN was set to 256 × 256. The ReLU function activates the convolutional layer with a size of 3 × 3 kernels, which is used for feature extraction. The convolution operation is expressed in Eq. (1).

$$Conv^{k}(\mathrm{i,j}) = \sum_{x,y} W^{k.l}(p,s).input^{i}(i-p,j-s) + b^{k,l} \tag{1}$$

where $W^{\mathrm{k.l}}$ means the k^{th} kernel and $b^{\mathrm{k,l}}$ means the bias of k^{th} layer.

For each feature map, a rectified linear unit (ReLU) is applied to get nonlinearity in the feature maps. ReLU estimates activation by retaining the threshold input at zero. It can be stated numerically as Eq. (2).

$$f(x) = \mathrm{max}imum(0,x) \tag{2}$$

It returns the input if it is positive, otherwise it returns zero. It is easier to train and achieve good performance.

A max-pooling layer is coupled to the output of the first convolution layer. For down sampling the dimensions of an input image, the max-pooling layer with a size of 2 × 2 kernel is used. Each layer has a set stride size of 1 pixel. In this feature extraction, the same convolutional block was used for each view separately.

3.4 Feature fusion

The next step is the merging of numerous characteristics retrieved from the MLO and CC perspectives of the mammography image. We integrated all the feature vectors from two mammography scans into a single vector in this fusion procedure, which was then used for final classification. This technique strengthens overall efficiency and model interpretability by achieving finer precision and diagnostic accuracy.

The LASSO regression module fuses the features from two-view mammograms with the benefit of shrinking. Data points are shrunk to the absolute mean in shrinking. The number of features accessible is reduced using Lasso regression. The L1 regularization method with the penalty of loss function is performed in this framework. The Eq. (3) shows that the penalty loss with the feature set absolute value.

$$\theta_{LASSO} = min \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 + \beta \sum_{j=0}^{k} \left|\theta_j\right| \tag{3}$$

In the equation, β represents the tuning factor which controls the penalty strength. The following are the standard outcomes of.

- If $\beta = 0$, then similar features are chosen from the feature set
- If $\beta = \infty$, then no features are selected, set zero to all coefficients.
- If $0 < \beta < \infty$, then the features are reduced between 0 and m where m is the coefficient of linear regression.

3.5 Mammogram classification based on CNN

The architecture sorts across a fully connected layer accompanied by a ReLu function where all the features are flattened and then connected to the softmax classifier. Then the end layer is mathematically expressed in Eq. (4).

$$\mathcal{F}_k^{end} = \sum_{j=1}^{end-1} w_{k,j}^{end} \mathcal{F}_g^{end-1} + L_k^{end-1} \tag{4}$$

Softmax classifiers provide the probabilities for each class label. Probabilities are considerably easier to interpret. The cross-entropy loss is used by the softmax classifier, which is a binary variant of Logistic Regression. It converts an input into a vector of positive values based on a probability distribution with a total sum of one. The output values are between 0 and 1. The dropout ratio is 15%.

In this case, the loss function should minimize the proper class's negative log likelihood. It's the proportion of the input parameters exponential to the sum of all the inputs' exponential parameters. The mathematical representation of the loss function for a single data point is given in Eq. (5).

$$L_i = -\log\left(e^{s_{y_i}} | \sum_j e^{s_j}\right) \tag{5}$$

The average is then used to calculate the cross-entropy loss for a whole input and is given in Eq. (6).

Table 1 The partitioning specification of dataset

Data/type	Normal	Benign	Malignant	Total
Train	1240	738	769	2747
Test	205	205	205	615
Total	1445	943	974	3362

$$L = \frac{1}{N} \sum_{i=1}^{N} L_i \tag{6}$$

4 Experimental results and discussion

4.1 Dataset collection and description

In this experiment, we used the CBIS-DDSM dataset (Shin et al. 2016) which contains biopsy-proven annotated mammograms. The collection contains scanned mammography pictures of bilateral breasts from the cranio-caudal (CC) and medio-lateral (MLO) viewpoints.

From the database, we retrieved 1418 normal mammograms, 852 benign mammograms, and 897 malignant mammograms based on radiologists' diagnoses. There are two mammograms from the CC and MLO views in each of these breasts, for a total of 6334 mammograms from 3167 breasts. We divide the data into two sets: training and testing. 1240 normal mammograms and 1507 abnormal mammograms (738 benign and 769 malignant) are included in the training. In the testing, there were 178 normal mammograms and 242 abnormal mammograms (114 benign and 128 malignant). Table 1 shows the partitioning specification of the dataset. The visualization of mammogram images of normal, benign, and malignant classes is shown in Fig. 4.

4.2 Experimental setup

MATLAB is utilized to create the model in this study. The tests were also performed on an NVIDIA GTX 1050 Ti GPU. The combined CNN-LASSO architecture is summarized in Table 2. As shown in Table 2, the proposed network has 23 layers: 10 combined convolutional and pooling layers for each view of mammogram, one feature fusion layer, one fully connected layer, and one softmax classifier. There were 150 epochs in the training procedure. We adjusted the initial learning rate to 0.0001 based on previous experience.

4.3 Performance evaluation metrics

The proposed system's performance is measured using the following metrics: TP (True Positive) signifies accurately classified the malignant mammogram, FP (False Positive)

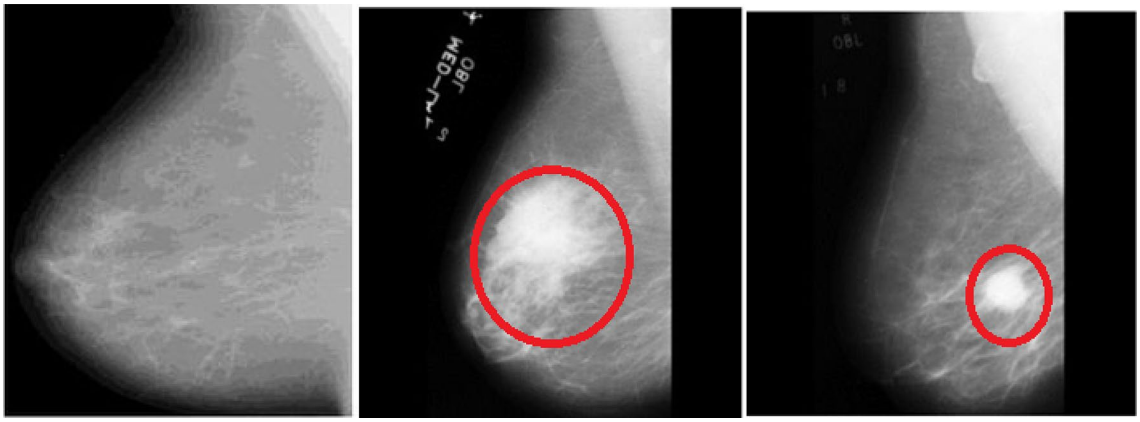

Fig. 4 Shows sample images of normal, benign and malignant cases respectively

signifies the normal or benign mammograms misclassified as malignant, TN (True Negative) signifies truly classified normal or benign mammograms, and FN (False Negative) signifies the malignant misclassified as normal or benign mammograms.

$$Accuracy = (TP + TN)/(TN + TP + TP + FN)$$

$$Sentitivity = TP/(TP + FN)$$

$$Specificity = TN/(TN + FP)$$

$$F1/;Score = (2 * TP)/(2 * TP + FP + FN)$$

4.4 Experimental results

In all experiments, we used the CBIS-DDSM dataset (Shin et al. 2016) which contains biopsy-proven annotated mammograms. The collection contains scanned mammography pictures of bilateral breasts from the cranio-caudal (CC) and medio-lateral (MLO) viewpoints. We retrieved 1418 normal mammograms, 852 benign mammograms, and 897 malignant mammograms from the database based on radiologists' diagnoses (see Table 1).

Experiment #1: The benign/malignant mammogram decision is performed in a three step approach. In the first step distinct CNN architectures are involved in the CC and MLO views to extract the relevant features. In the second step, the fused features are determined by concatenating the feature vectors of both the views. The extracted features are concatenated using feature fusion, which includes LASSO regression to reduce the discriminative feature and redundancy. In the third step, the classifier is resolved to make a final decision on cancer. Figure 5b represents the confusion matrix of

Table 2 Details of combined CNN-LASSO architecture

Layer type	Input size	Kernel size	Stride	Output size
Convolution1(1)	$256 \times 256 \times 1$	$3 \times 3 \times 16$	1	$256 \times 256 \times 16$
Convolution1(2)	$256 \times 256 \times 1$	$3 \times 3 \times 16$	1	$256 \times 256 \times 16$
Pooling1 (1)	$256 \times 256 \times 16$	2×2	2	$128 \times 128 \times 16$
Pooling1(2)	$256 \times 256 \times 16$	2×2	2	$128 \times 128 \times 16$
Convolution2(1)	$128 \times 128 \times 16$	$3 \times 3 \times 32$	1	$128 \times 128 \times 32$
Convolution2(2)	$128 \times 128 \times 16$	$3 \times 3 \times 32$	1	$128 \times 128 \times 32$
Pooling2(1)	$128 \times 128 \times 32$	2×2	2	$64 \times 64 \times 32$
Pooling2(2)	$128 \times 128 \times 32$	2×2	2	$64 \times 64 \times 32$
Convolution3(1)	$64 \times 64 \times 32$	$3 \times 3 \times 64$	1	$64 \times 64 \times 64$
Convolution3(3)	$64 \times 64 \times 32$	$3 \times 3 \times 64$	1	$64 \times 64 \times 64$
Pooling3(1)	$64 \times 64 \times 64$	2×2	2	$32 \times 32 \times 64$
Pooling3(2)	$64 \times 64 \times 64$	2×2	2	$32 \times 32 \times 64$
Convolution4(1)	$32 \times 32 \times 64$	$3 \times 3 \times 128$	1	$32 \times 32 \times 128$
Convolution4(2)	$32 \times 32 \times 64$	$3 \times 3 \times 128$	1	$32 \times 32 \times 128$
Pooling4(1)	$32 \times 32 \times 128$	2×2	2	$16 \times 16 \times 128$
Pooling4(2)	$32 \times 32 \times 128$	2×2	2	$16 \times 16 \times 128$
Convolution5(1)	$16 \times 16 \times 128$	$3 \times 3 \times 256$	1	$16 \times 16 \times 256$
Convolution5(2)	$16 \times 16 \times 128$	$3 \times 3 \times 256$	1	$16 \times 16 \times 256$
Pooling5(1)	$16 \times 16 \times 256$	2×2	2	$8 \times 8 \times 256$
Pooling5(2)	$16 \times 16 \times 256$	2×2	2	$8 \times 8 \times 256$
Lasso fusion layer	$8 \times 8 \times 256$	–	–	$8 \times 8 \times 256$
FC layer	16,384	64	–	–
Softmax classifier	64	3	–	–

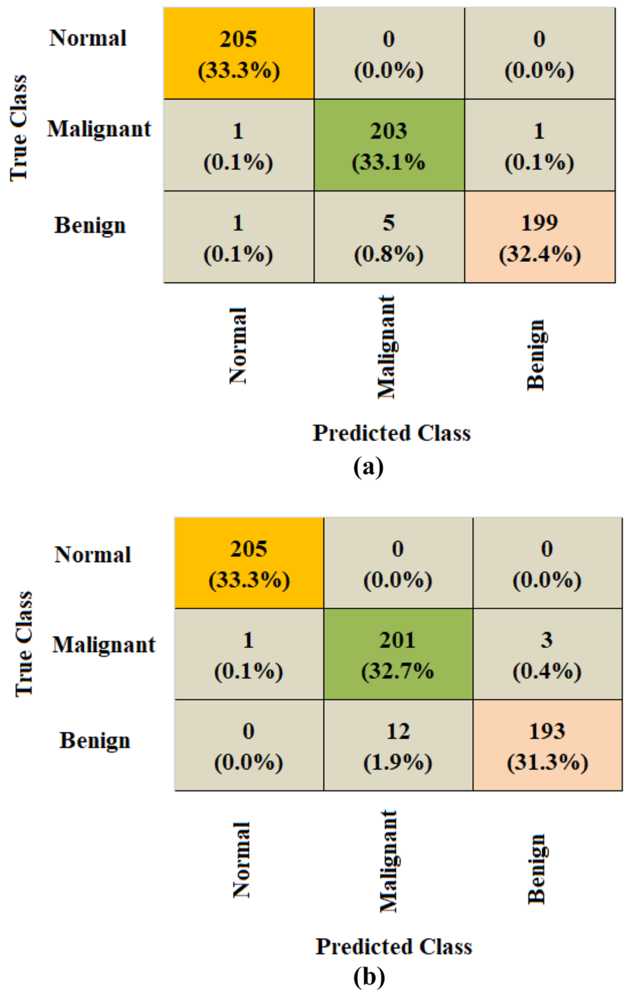

Fig. 5 Confusion Matrix of the mammogram classification **a** combined CNN-LASSO **b** CNN

Table 3 CNN and CNN-LASSO Regression performance comparison

Class	CNN-LASSO (with Feature Fusion)			CNN (without Feature fusion)		
	Normal (%)	Benign (%)	Malignant (%)	Normal (%)	Benign (%)	Malignant (%)
Accuracy	99.7	98.9	98.9	97.9	97.9	97.4
Specificity	99.5	99.8	98.8	97.8	99.3	97.1
Sensitivity	100	97.1	99	100	94.1	98
F1 Score	99.5	98.3	98.3	97.8	96	96.2

combined CNN-LASSO regression for the classification of mammogram cancer.

Experiment #2: It comprises three important phases of the CAD system: feature extraction, feature fusion of two view features, and mammogram classification. Two-view multi set of features are extracted using convolutional and pooling layers, which greatly increase the classification rate. Then multiple view features are concatenated without feature fusion. Finally, the flattened features are input into the fully connected layer and the sofmax classifier for classification of mammograms into normal, benign, and malignant cases. Figure 5a represents the confusion matrix of CNN without feature fusion for the classification of mammogram cancer.

For result analysis, we put up a set of comparative studies of two experiments to prove the efficiency. The confusion matrix of CNN without feature fusion and combined CNN-Lasso regression feature fusion for the classification of mammogram cancer is shown in Fig. 5. The CNN architecture misclassified 16 of the 615 images, including four for malignant cases. Meanwhile, the suggested CNN-Lasso regression architecture misclassified only eight images, including two for malignant cases. As a result, the suggested method is capable of accurately classifying malignant situations.

The suggested CNN-LASSO Regression feature fusion algorithm's performance is compared to that of the CNN architecture without feature fusion in Table 3. For the malignant cases, the CNN network obtained 97.4% accuracy, 97.1% specificity, 98.0% sensitivity, and 96.2% F1-score. It achieved 97.8% accuracy, 99.3% specificity, 94.1 percent sensitivity, and 96% F1-score of the benign classification. It achieved 99.7% accuracy, 99.8% specificity, 99.8 F1-Score and 100% sensitivity of normal classification. Furthermore, the proposed CNN-LASSO regression architecture classified the malignant cases with 98.9% accuracy, 99% sensitivity, 98.8% specificity, and 98.4% F1-score. For the benign classification, it achieved 98.9% accuracy, 99% sensitivity, 99.8% specificity, and 98.8% F1-score. It attained 99.8% accuracy, 100% sensitivity, 99.5% specificity, and 99.5% F1-score of normal classification. As a result, the suggested method is capable of accurately classifying malignant situations.

The ROC curves between the true positive rate (TPR) and the false positive rate (FPR) are additionally incorporated to assess the overall performance in Fig. 6. The area under ROC curve (AUC) for the CNN without feature fusion and CNN-LASSO Regression feature fusion architectures, respectively, was calculated to be 98.3% and 99.8%, indicating that the suggested network outperformed the CNN architecture.

The results show that combining CNN and LASSO regression has a considerable influence on the detection of

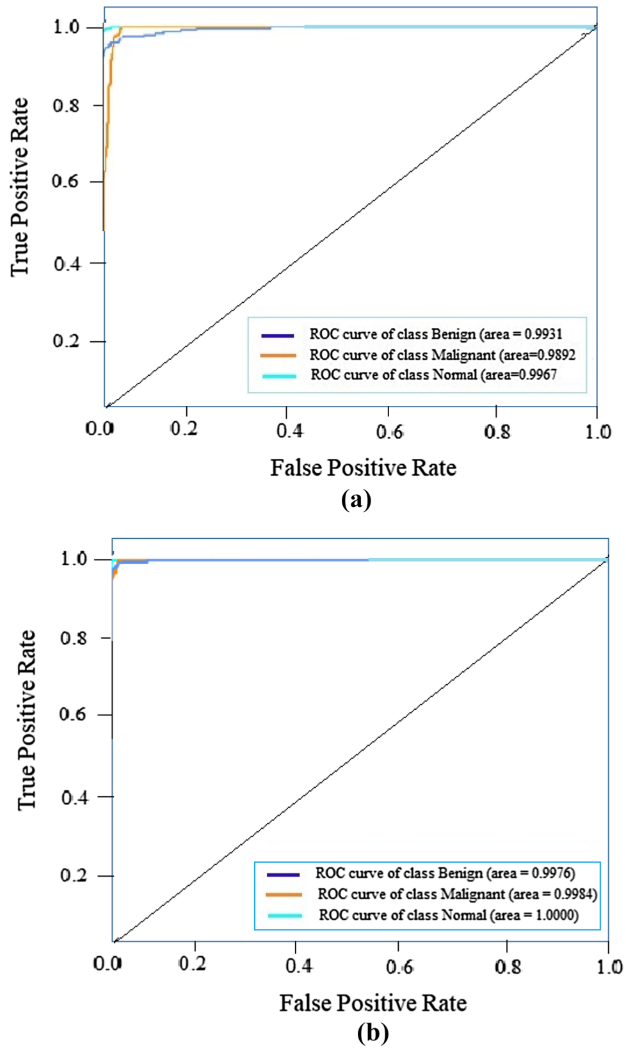

Fig. 6 ROC plotting for normal, benign and malignant classification **a** CNN **b** CNN-LASSO regression

4.5 Comparison with state-of-the-art mammogram classification

In this study, we developed a deep learning based system that integrates CNN and LASSO regression to diagnose breast cancer from mammography pictures automatically. To demonstrate the advancement of our proposed strategy, we compare it to a deep learning-based system for classifying mammograms published in recent years. Table 4 compares our results with those of other deep learning-based methods based on the classification of multiview mammograms.

Hua Li et al. (2020) used gate recurrent unit (GRU) structures of RNN to fuse the features of the two images based on the spatial correlation between different views. To extract breast-mass properties of mammograms from cranio-caudal (CC) and medio-lateral oblique (MLO) views, the model is made up of two branch networks and two modified ResNet.

Zhang et al. (2020a, b) employed two CNN branches to extract mammography characteristics from two different MLO and CC views. They focused on the development of a multi-scale convolution module and an attention module which can be used to extract features from different views of mammograms. The multi-scale convolution module allows the network to extract visual information at various scales, and the attention module can assist the model in focusing on useful information selectively.

Hua Li et al. (2019) developed a DenseNet-II model for classifying whole mammograms as benign or malignant by adding an Inception structure after the DenseNet model to extract multi-scale signals. The creation of a new DenseNet-II neural network model keeps the network structure sparse. It also prevents overfitting of the model and boosts computer performance.

Table 4 Comparison with state-of-the-art mammogram classification

Method	Data set	Accuracy (%)	Sensitivity (%)	AUC (%)
Hua Li et al. (2020)	DDSM	94.7	94.1	96.8
Zhang et al. (2020b)	DDSM	95.24	96.11	95.03
Li et al. (2019)	DDSM	94.55	95.6	91.2
Khan et al. (2019b)	CBIS-DDSM	77.66	81.82	76.9
Our Proposed	CBIS – DDSM	99.2	99.4	99.8

mammogram cancer using automatic feature extraction from mammogram images. With the proposed approach, malignant cases may be distinguished from benign and normal cases with good accuracy.

Nasir Khan et al. (2019a, b) devised a model that merges the features from four views of a mammogram of each patient and significantly improves classification performance. Early fusion is applied, which is the process of concatenating multiple feature vectors into a single feature vector.

According to Table 4, our proposed model's effectiveness was improved. Our model outperforms the methods outlined above, with an accuracy of 99.2%, sensitivity of 99.4% and an AUC of 99.8%. Our suggested approach is able to obtain an outcome that is superior to recent state-of-the-art studies by performing LASSO regression feature fusion. This achievement proves the potential of CNN-LASSO in classifying breast cancer from mammograms with the highest accuracy.

5 Conclusion

In this research, we primarily concentrated on the classification of mammogram cancer and proposed a novel feature, named "multi view CNN-LASSO regression," for identifying breast cancer. We discussed the performance of multi view CNN and multi view CNN with feature fusion. We conducted experiments with two views of the data, and the results show that feature fusion helped us achieve the highest testing accuracy rate. Furthermore, the comparison analysis reveals that the suggested model achieves better classification accuracy while also reducing the computing complexity of the system. In the future, we can explore the way to achieve the most efficient feature subset choice to spice up the performance further through utilizing some optimization procedures and also examine the effect of different sources for the enhancement of the proposed system.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Human or animal rights This article does not contain any studies involving animals performed by any of the authors.

References

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2011) Global cancer statistics. Cancer J Clin 68:394–424

Cheng J-Z, Ni D, Chou Y-H, Qin J, Tiu C-M, Chang Y-C et al (2016) Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Re Ports 6(1):24454. https://doi.org/10.1038/srep24454

Dheeba J, Singh NA (2015) Computer aided intelligent breast cancer detection: second opinion for radiologists—a prospective study. Computational intelligence applications in modeling and control. Springer Cham, Switzerland, pp 397–430

Gao F, Chia K-S, Ng F-C, Ng E-H, Machin D (2002) Interval cancers following breast cancer screening in Singaporean women. Int J Cancer 101:475–479

Hua K-L, Hsu C-H, Hidayati SC, Cheng W-H, Chen Y-J (2015) Computer- aided classification of lung nodules on computed tomography images via deep learning technique. Onco Targets Ther 8:2015–2022. https://doi.org/10.2147/OTT.S80733

Jouirou A, Baâzaoui A, Barhoumi W (2019) Multi-view information fusion in mammograms: a comprehensive overview. Inform Fusion 52:308–321. https://doi.org/10.1016/j.inffus.2019.05.001

Kallenberg M, Petersen K, Nielsen M, Ng AY, Diao P, Igel C et al (2016) Unsupervised deep learning applied to breast density segmentation and mam- mographic risk scoring. IEEE Trans Med Imaging 35(5):1322–1331

Khan HN, Shahid AR, Raza B, Dar AH, Alquhayz H (2019) Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. IEEE Access 7:165724–165733. https://doi.org/10.1109/ACCESS.2019.2953318

Khan HN, Shahid AR, Raza B, Dar AH, Alquhayz H (2019) Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. IEEE Access 7:165724–165733. https://doi.org/10.1109/ACCESS.2019.2953318

Khan TM, Shengjun Xu, Khan ZG, Uzair M, chishti. (2021) Implementing multilabeling, ADASYN, and relieff techniques for classification of breast cancer diagnostic through machine learning: efficient computer-aided diagnostic system, journal of healthcare. Engineering. https://doi.org/10.1155/2021/5577636

Kumar D, Wong A, Clausi DA (2015) Lung nodule classification using deep features in CT Images. In: Proceedings of the 12th conference on computer and robot vision (pp 133–138). https://doi.org/10.1109/CRV.2015.25

Lai ZF, Deng H (2018) Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. Comput Intell Neurosci. https://doi.org/10.1155/2018/2061516

Li H, Zhuang S, Li D-a, Zhao J, Ma Y (2019) Benign and malignant classification of mammogram images based on deep learning. Biomed Signal Process Control 51:347–354. https://doi.org/10.1016/j.bspc.2019.02.017

Li H, Niu J, Li D, Zhang C (2020) Classification of breast mass in two-view mammograms via deep learning. IET Image Proc. https://doi.org/10.1049/ipr2.12035

Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D (2014) Early diagnosis of Alzheimer's disease with deep learning. In: Proceedings of the IEEE 11th international symposium on biomedical imaging (ISBI) (pp 1015–1018), https://doi.org/10.1109/ISBI.2014.6868045

Munir K, Elahi H, Ayub A, Frezza F, Rizzi A (2019) Cancer diagnosis using deep learning: a bibliographic review. Cancers 11:1235

Nahid A-A, Mehrabi MA, Kong Y (2018) Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. BioMed Res Int. https://doi.org/10.1155/2018/2362108

Nan W, Phang J, Park J, Shen Y, Huang Z, Zorin M, Jastrzebski S, Fevry T, Katsnelson J, Kim E, Wolfson S, Parikh U, Gaddam S, Lin LLY, Ho K, Weinstein JD, Reig B, Gao Y, Toth H, Pysarenko K, Lewin A, Lee J, Airola K, Mema E, Chung S, Hwang E, Naziya Samreen S, Kim G, Heacock L, Moy L, Cho K, Geras KJ (2020) Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 39(4):1184–1194. https://doi.org/10.1109/TMI.2019.2945514

Online Document Breast Cancer Facts and Figures (2019) American cancer society. Atlanta, GA, USA

Saleem Z, Ramadan Fantacci, Maria E (2020) Using convolutional neural network with cheat sheet and data augmentation to detect breast cancer in mammograms. Comput Math Methods Med. https://doi.org/10.1155/2020/9523404

Shin H-C, Roth HR, Gao M et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imag 35(5):1285–1298

Siegel RL, Miller KD, Jemal A (2020) Cancer statistics. Cancer J Clin 70(1):7–30. https://doi.org/10.3322/caac.21590. (**PMID: 31912902**)

Sridevi V, Dr J, Samath A (2020) Advancement on breast cancer detection using medio-lateral-oblique (Mlo) and cranio-caudal (CC) features. Test Eng Manag 83:85–93

Sridevi V, Abdul Samath J (2019) A survey on breast cancer segmentation and classification using several methods. International journal of Scientific Research in computer science applications and management studies 8(3)

Suk H, Lee SW, Shen D (2014) Hierarchical feature representation and mul- timodal fusion with deep learning for AD/MCI diagnosis. Neuroimage 101:69–582. https://doi.org/10.1016/j.neuroimage.2014.06.077

Suk H, Shen D (2013) Deep learning-based feature representation for AD/MCI classification. In: Lecture notes in computer science (including subseries Lecture notes in artificial intelligence and lecture notes in bioinformatics): 8150 (pp 583–590), LNCS https://doi.org/10.1007/978-3-642-40763-5_72

Sun L, Wang J, Zhijun H, Yong X, Cui Z (2019) Multi-view convolutional neural networks for mammographic image classification. IEEE Access 7:126273–126282. https://doi.org/10.1109/ACCESS.2019.2939167

Swiderski B, Kurek J, Osowski S, Kruk M, Barhoumi W (2016) Deep learning and non-negative matrix factorization in recognition of mammograms. In: Proceedings of the eighth international conference on graphic and image processing (ICGIP 2016), Tokyo, Japan, pp 29–31

Tang J, Member S, Rangayyan RM, Xu J, El Naqa I (2009) Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE Trans Inf Technol Biomed 13:236–251

Wang D, Khosla A, Gargeya R, Irshad H, Beck AH (2016) Deep learning for identifying metastatic breast cancer, 1–6. arXiv Preprint https://people.csail.mit. edu/khosla/papers/arxiv2016_Wang.pdf

Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activations in convolutional network. arXiv 2015, arXiv:1505.00853

Zhang C, Zhao J, Niu J, Li D (2020) New convolutional neural network model for screening and diagnosis of mammograms. PLoS ONE 15(8):e0237674. https://doi.org/10.1371/journal.pone.0237674

Zhang C, Zhao J, Niu J, Li D (2020) New convolutional neural network model for screening and diagnosis of mammograms. PLoS ONE 15(8):e0237674. https://doi.org/10.1371/journal