How to Cite:
Sridevi, V., & Samath, J. A. (2022). Mlo and CC view of feature fusion and mammogram classification using deep convolution neural network. *International Journal of Health Sciences*, *6*(S7), 5196-5207. https://doi.org/10.53730/ijhs.v6nS7.13106

MLO and CC view of feature fusion and mammogram classification using deep convolution neural network

V. Sridevi
Assistant Professor, Department of Computer Science, PSG College of Arts & Science and Research Scholar of Govt. Arts College, Udumalpet, India
Emal: vissridevi@gmail.com

J. Abdul Samath
Assistant Professor, Department of Computer Science, Chikkanna Government Arts College, Tirupur, India,
Email: abdul_samath@yahoo.com

***Abstract*---**Breast cancer is the most frequent type of cancer in women all over the world. The improvement of computer aided system help the radiologist for the effective analysis and diagnosis of breast cancer. It presents a computational methodology for classifying breast cancer as normal, benign and malignant from CC and MLO views of mammogram image. The proposed strategy consists of feature extraction, multiple view feature fusion and classification. The input images are fed into feature extraction where convolution neural network is applied. The CNN is nicely suitable for both feature extraction, feature fusion and mammogram classification. In this framework, convolution layer, pooling and activation function are used as a feature extraction techniques. After the process of feature extraction, feature fusion is employed by average pooling of CNN. The feature fusion will increase or maximize the relevant information of the breast image. Finally obtained features from the fusion are fed into CNN classifier in which softmax and fully connected layer are employed as a classifier techniques. The proposed work achieves 98.4% of accuracy to classify the breast cancer from MLO and CC views using hybrid feature with CNN classifier.

***Keywords*---**Mammogram, Breast Cancer, CNN, Feature Extraction, Feature Fusion, Convolution, Pooling, MLO and CC Views, Classification.

Manuscript submitted: 9 May 2022, Manuscript revised: 18 July 2022, Accepted for publication: 27 August 2022

Introduction

The breast cancer is found in women and it causes a high mortality rate among women in India. In all countries, this type of disease is the main one in women. It is the second most prevalent type of cancer among all types, and it ranks fifth in terms of inflicting death. (**Ferlay et al., 2015**). The early detection or diagnoses of breast cancer have higher chance for cured. The world government along with various health professionals provides awareness or alerts the female about the risk of this disease. Therefore the early discovery is the perfect way to anticipate this infection. The most common early discovery of breast cancer is done through mammography; this provides the digital image of the breast. Through this, clinical professional specialist distinguishes or visualizes the lesions or non-lesions in the breast (**Junior et al., 2013 and Sampaio et al., 2011**).

The radiologist missed some features in the time of visualization of breast cancer in mammogram which leads to false diagnosis. This provides an error rate of 10% - 20% for the recognition of malignant masses (**Mohamed, 2018**). The suspicious tissues are evacuated from the breast and it given to breast biopsy to test the presence of breast cancer or not (**Shankar Thawkar and Ranjana Ingolikar, 2018**).

To analyse the medical image and offer reliable results, a computer aided diagnosis (CAD) system is used. The CAD system assists the radiologist and lowers the amount of false positive outcomes [**V.Sridevi et al. 2010**]. With the help computer based system, the unnecessary biopsies are reduced by computer based diagnosis system and the radiologist identifies the breast cancer and it helpful to decrease the mortality rate.

Many computer vision applications [**Alex Krizhevsky et al., and Kaiming He et al., 2015]** as well as applications linked to mammography analysis, such as mass segmentation [**Neeraj Dhungel et al., 2015**], mass detection [28], and mammogram classification [**Gustavo Carneiro et al**., **2015**], have delivered state-of-the-art results using deep learning models. The capacity of deep learning models to learn and incorporate low-, mid-, and high-level features by stacking hidden layers in the network architecture is the reason for their success [**Christian Szegedy et al., 2015**].

Therefore, computer aided detection system based on deep learning increases the accuracy rate in the early identification of breast cancer. This method is implemented using MATLAB software as a programming language in which CBIS-DDSM dataset are applied for testing and training purpose.

The proposed framework is primarily based on two-view based feature extraction and feature fusion technique that are evaluated on CBIS-DDSM mammogram images. In this paper, analysis of MLO and CC views are performed through CAD method. The proposed system comprises of three important phases of CAD system are feature extraction, feature fusion of two view features and mammogram classification. Two-view multi set of features are extracted using convolutional and pooling layer which highly increase the classification rate. Method. Then multiple view features are concatenated and reduced by feature fusion technique which uses average pooling method. Finally, the fused features are input into fully connected layer and sofmax classifier for classification of mammogram into normal, benign and malignant case.

1. Related Work

The suspicious region between the left and right mammograms is detected through structural variations between the two regions. Ereceira et al., 2013.

Sampaaio et al., 2015 develop a method for the detection of mass. In this method, first tissues density is detected. The candidate region is classified into mass or non-masses using texture-based characteristics and an SVM classifier. The sensitivity is 94.01%, the specificity is 82.25%, and the accuracy is 84.06% as a consequence.

The automatic classification of mammogram image into normal, benign and malignant classes are proposed by Raghavendra et al. (2016). They extract the features from mammogram image using Gabor Wavelet and select the appropriate features or reduce it using Locality Sensitive Discriminant Analysis method. Finally, the obtained features are fed into K-NN classifier to classify the mammogram image into normal, benign and malignant. They obtained an accuracy of 98.69%.

Khehra and Pharwaha (2017) devised a hybrid technique for the mammogram cancer classification. For feature selection, they apply three different techniques: genetic algorithm (GA), particle swarm optimization (PSO), and biogeography-based optimization. The classification is done using a support vector machine in this method.

S. D. 2017 generated hybrid method for the classification of mammogram image. In this approach, they used MIAS and BCDR standard benchmark breast cancer dataset. The supervised techniques are employed in this method for the extraction of features. The obtained feature vectors are given into hybrid method of SVM and DCNN classifier for the classification purpose. They obtained a sensitivity of 98.75%. The drawback of this structure is high complex valued network.

M. Ciecholewski. 2017 collect the mammogram images from DDSM and MIAS database. They used contour approaches including geometric active contour (GAC) and active contour without edges (ACWE), as well as the damping coefficient (EM) for the classifier. The three types of contour is used it causes lack in the micro calcifications.

Shankar and Ranjana, 2018 developed a biogeography based optimization with adaptive neuro-fuzzy inference system for mass enhancement and classification. They attained the accuracy of 98.91%. A. Rakhlin et al., 2018 used VGG, Inception and ResNet for the classification of mammogram image.

Mohmed et al., 2018 presents a method classified mammogram into mass and non-mass using three phases: image preprocessing, feature extraction and classification. They proposed their method for DDSM and MIAS database and

SVM used for classification. They extract the features using hybrid method is statistical features with texture features. The obtained feature vector is fed in to SVM classifier to classify it. They obtained sensitivity of 98.82% and specificity of 98.69%.

Prabhpreet et al., 2019 established a hybrid technique for breast cancer mammography image detection. The noise is first reduced by combining the work of the median and mean filters. This phase's output is fed into the K-mean clustering algorithm for feature extraction. To categorise the mammography picture into norma1, the recovered feature vectors are fed into a Multi-support vector machine.

V. Sridevi et al, 2020 developed mammogram classification in which the suspicious region is segmented using the Adaptive K-means segmentation algorithm and the KMC-GF feature extraction approach is used. It shown that for all situations, KNN classifier performance using KCM-GF feature extraction outperforms GLCM features for CC-M, CC-N, MLO-M, and MLO-N cases.

2. Methodology

For the proposed methodology, CBIS-DDSM dataset was used. Each image consists of cranio caudal and medio-lateral oblique view of mammogram. Totally 2581 images are collected from CBIS_DDSM cancer data set in which 2197 training images and 384 testing images. The proposed framework consists of feature extraction, feature fusion and classification. Figure 1 depicts the suggested architecture for breast cancer classification in its entirety.

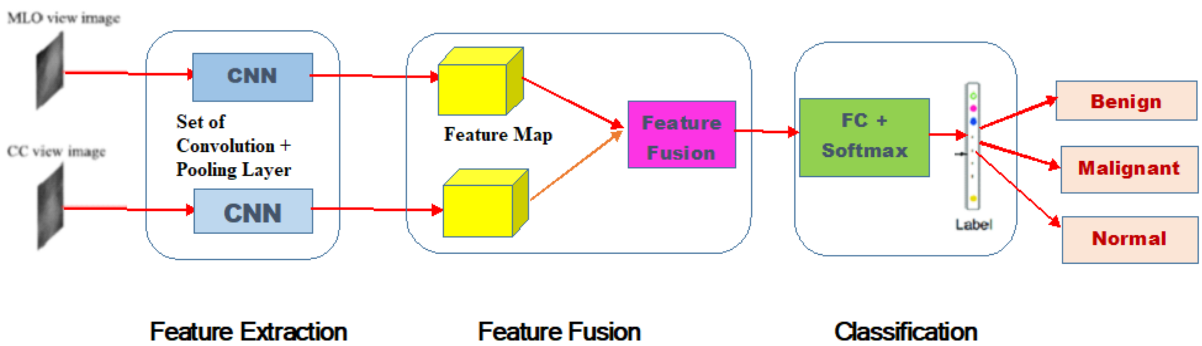

Figure 1: Architecture of breast cancer classification

Convolutional and pooling layers are used in feature extraction to extract features from each view of a mammography image. It extracts the multiple view features from the CC and MLO view mammogram image. So, feature fusion is needed to merge the multiple features into single vector. Finally, the given mammogram image is classified as normal, benign and malignant by fully connected layer of CNN.

3.1 Feature Extraction

The relative features of mammogram images are obtained through feature extraction. In this proposed framework, convolution neural network is employed. During feature extraction, the number of convolution layers is changed. For each view, 4 convolutional layer and 4 max pooling layers are exercised to extract the features. To discover where the neurons output is related to the input local areas, the convolution layer calculates the scalar product between the input and weights. The goal of the pooling layer is to lower the input's spatial dimensionality and the number of parameters in the activation. The feature extraction is evaluated using an activation function such as ReLU. The activation of ReLU is

$$f(x) = \begin{cases} 0 & for \quad x < 0 \\ x & for \quad x \geq 0 \end{cases}$$

Convolution neural network is effective in image classification problem due to spatial correlated features of the image given by convolution operation. Convolution is obtained by initialization of square matrix through specific values. The obtained matrix is applied to each pixel of mammogram image. A maximum pooling layer follows the convolution layer.

To construct the convolution layers entire output volume, each kernel in a CNN has an activation map which is stacked with the depth dimension. By improving the model's output, the convolution layer helps to reduce the model's complexity. The optimizations for depth, stride, and zero-padding are used. The convolution layer generates the output volume depth based on a controlled setup of the number of neurons inside the layer to the same input region. To specify the stride, set the depth around the input's spatial dimensions. The border of the input is padded using the zero-padding method.

To extract convolutional features from a mammographic image, each convolutional neural network employs four convolutional layers. Each convolutional layer has 3×3 convolutional kernels with stride of 1 and is followed by a max pooling layer with kernel filter of 2×2 and stride of 2. A concatenation layer is used at the network's end to integrate the feature maps created by the two convolutional neural subnetworks. The inputted mammographic image is 128 x 128 pixels in size.

The convolution layer takes 2D matrix. The single output matrix of convolution layer is given by below equation.

$$C_j = f\left(\sum_{i=1}^{N} \widetilde{M}_i * K_{i,j} + B_j\right) \quad (24)$$

In the above,

$\widetilde{M}_i$ - represent the input matrix which convolves with kernel matrix $K_{i,j}$

B_j- represents bias which is added to output after the computation of sum of all convoluted matrix f- represents the nonlinear activation function

The objective of CNN neural network is:

- The CNN is having a capable to retrieve the spatial and temporal features of an image with the application of relevant filter.
- It reduces the images without losing their features.
- It extracts the high level features such as edges of the input image

3.2 Feature Fusion

Mammograms are often taken from two different perspectives: CC and MLO. We propose a multi-view feature fusion network model to fully exploit the complementary relationship between the two perspectives. As a result, average pooling is used in this framework to minimize the number of features from each patient's two mammography scans. We employed the global average operation at this point to reduce the picture feature to a one-dimensional vector, which was then fed into the fully connected layer of network. This feature fusion approach significantly increases the system performance. This method's purpose is to lower the prediction error. We stack the two view's feature vectors and is represented as

*

$$Z_v = \lfloor z^{(1)}; \dots ; z^{\langle ny \rangle} \rfloor \in R^{1 \times D \times ny}$$

And apply avg-pooling across them as defined in Equation 4. Again, this operation sacrifices the correspondence between the different views.

3.3 Classification

In a mammography image, the classification attempts to distinguish between normal, benign, and malignant abnormalities. After that, the fused features are fed into a fully connected layer, which is subsequently passed on to the softmax classifier. The number of output classes in the sofmax classifier is equal to the number of the outputs In order to connect our proposed architecture, we used tanh as a non-linear protocol. The softmax function is a squashing function that re-normalizes the k-dimensional input vector, yielding a range of real values between [1,2]. It can be expressed mathematically as

$$\sigma(z)_j = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}}, for\ j = 1, \dots K$$

The overfitting and underfitting phenomena frequently impair deep learning training. We used batch normalization after each layer in our proposed architecture to address these concerns. After the first fully connected layer, the dropout was added.

3.4 Loss Function

We increase the breast mass image classification performance by adjusting the loss function and upgrading the network design. The weighted loss function was used to train the suggested CNN classifier.

$$L(w, x_{i,} y_i) = -\frac{1}{N}\sum_{j=1}^{k} t_{ji} \ln y_{jn}$$

Here, x_i is the input vector, y_i is the classifier's prediction for the nth clinical input data, and t_n is the nth clinical sample's actual response. The number of classes is K, while the total number of clinical samples is N.

3. Result and Discussion

We test our method on publicly available datasets: CBIS-DDSM. The assessment datasets, the experimental setting, and the specifics of our approach's experimental implementation are presented first in this work. Our method's experimental findings are then thoroughly examined.

4.1 Dataset

The Digital Resource for Screening Mammography (CBIS-DDSM), which is currently the largest publicly available mammography resource, is used to train and assess our proposed model. The database contains a total of 10480 mammograms from 2620 cases, with each case including four mammograms from CC and MLO views of bilateral breasts. In this part, a total of 5162 samples from 2581 breasts. We retrieve 1018 normal, 766 benign, and 797 malignant breasts based on radiologists' diagnoses.

All of the data is separated into two categories: training and testing. There are 890 normal breasts, 638 benign and 669 malignant mammogram in the training set. There are 128 normal breasts, 128 benign and 128 malignant mammogram in the testing set. We proportionally partition the training data into ten subsets so that we may test the model's performance using ten-fold cross-validation. Table 1 represents the composition of dataset used for the experimentation. Figure 2 depicts the visualization of each class of mammogram image.

Table 1. Composition of Dataset used for the experimentation

Type	**Normal**	**Benign**	**Malignant**	**Total**
Train	890	638	669	2197
Test	128	128	128	384
Total	1018	766	797	2581

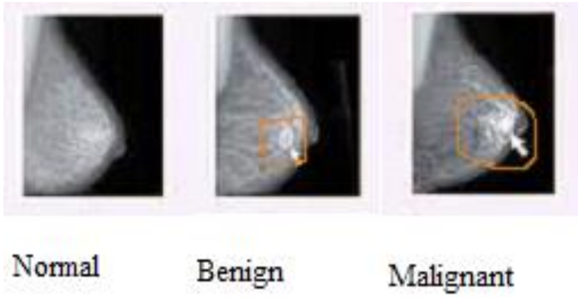

Figure 2. Visualization of each class of mammogram

4.2 Experimental Setup

Modern CNNs are frequently constructed by stacking convolutional layers on top of the input and then connecting the classification output with one or more fully connected (FC) layers. Max pooling layers are widely used between convolutional layers to improve translational invariance and reduce feature map size. In each block 3×3 convolution filter of the same depth, followed by a 2×2 max pooling layer of stride 2 that shrinks the feature map by a factor of two. Table 1 shows the network architecture of our coding system in detail. It is possible to converge after 100 epochs, as indicated in Table 2. Finally, we employed ReLu as an activation function for each convolutional layer. Aside from that, batch normalization was used to speed up deep network training.

Table 2. Network Architecture of Proposed System

Type	Kernel with Stride	Output Size
Convolution1	$3 \times 3 \times 16$	$128 \times 128 \times 16$
Pooling1	$2 \times 2 \times 2$	$64 \times 64 \times 16$
Convolution2	$3 \times 3 \times 32$	$64 \times 64 \times 32$
Pooling2	$2 \times 2 \times 2$	$32 \times 32 \times 32$
Convolution3	$3 \times 3 \times 64$	$32 \times 32 \times 64$
Pooling3	$2 \times 2 \times 2$	$16 \times 16 \times 64$
Convolution4	$3 \times 3 \times 128$	$16 \times 16 \times 128$
Pooling4	$2 \times 2 \times 2$	$8 \times 8 \times 128$
Feature Fusion	$2 \times 2 \times 2$	$4 \times 4 \times 128$
Fully Connected		$1 \times 1 \times 64$
Softmax		$1 \times 1 \times 64$

4.3 Performance Metrics

The feature fusion with CNN system's performance to correctly classify Standard medical metrics are used to evaluate mammograms at all stages. They are sensitivity, specificity, F1 score and accuracy. Sensitivity refers to the ability of a model to anticipate the positive outcome of a test. When the actual class is also positive, a mammogram is recommended. The term "specificity" refers to a model's ability to predict the negative class. The confusion matrix from the classifier output is used to evaluate the proposed system's performance measures.

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

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

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

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

$$AUC = \frac{sensitivity + Specificity}{2}$$

TP (True Positive): Abnormal cases are exactly classified as abnormal.

FP (False Positive): Normal cases are incorrectly classified as abnormal.

TN (True Negative): Normal cases are exactly classified as normal

FN (False Negative): Abnormal cases are incorrectly classified as normal.

4.4 Result Analysis

We used CNN architecture and multiple view Feature Fusion with CNN techniques to investigate and measure network performance. Table 3 shows the outcomes of putting the various network topologies to the test.

Table 3. Comparative performance of CNN without fusion and feature fusion with CNN

Method	**Accuracy(%)**	**Sensitivity(%)**	**Specificity(%)**	**F1-Score(%)**
CNN without fusion	92.2	93.5	92.8	94.4
Feature Fusion with CNN (Proposed Method)	98.4	98.3	97.2	96.9

The feature fusion with CNN classifier provides better recognition rate compared to CNN without fusion. The feature fusion with CNN classifier has an advantage of texture and neural network specification features. Positive and negative likelihood are used to test the ability to predict positive and negative classes on a confusion matrix. Higher positive likelihood and lower negative likelihood indicate better achievement on positive and negative classes. We used the ROC curve and area under the ROC to evaluate our models for a more thorough analysis of the proposed system. The ROC plot is a two-dimensional space with the x-axis representing the false positive rate (1-specificity) and the y-axis representing the true positive rate of the system (sensitivity). The area under receiver operating characteristics curve was used by classifier to assess the imbalance data and also it is more reliable than accuracy.

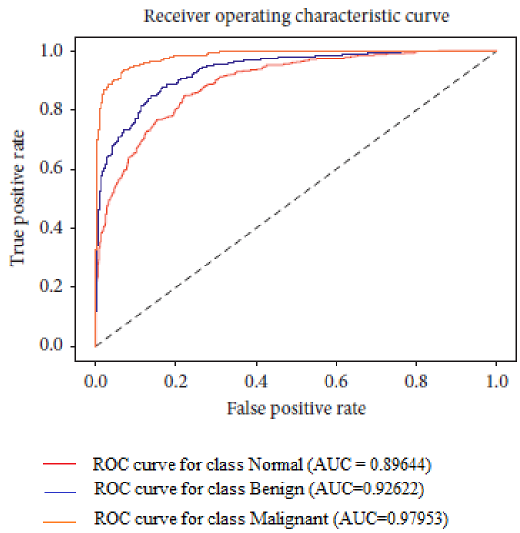

Figure 3. ROC curve of the average performance using fusion after reduction approach

As shown in Figure 3, the area under the ROC curve (AUC) for the CNN and feature fusion with CNN designs was calculated to be 97.9 percent and 95.2 percent, respectively, clearly demonstrating that the suggested network outperformed the CNN architecture.

4. Conclusion

Two views are used in this framework: MLO and CC view models. The information retrieved from these perspectives is utilized to classify them. The proposed two views of MLO and CC normal and abnormal models provide a greater recognition rate than a single-view based system. The experimental results suggest that employing a two-view feature extraction and feature fusion system, the proposed method improves. Furthermore, a comparison analysis demonstrates that the proposed designs outperform other mammography classification systems significantly. The suggested effort will aid in the detection of MLO and CC views of mammogram cancer.

References

1) J. Ferlay , I. Soerjomataram , R. Dikshit , S. Eser , C. Mathers , M. Rebelo , D.M. Parkin , D. Forman , F. Bray , Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012, Int. J. Cancer 136 (5)(2015) E359–E386 .
2) G.B. Junior , S.V. da Rocha , M. Gattass , A.C. Silva , A.C. de Paiva , A mass classification using spatial diversity approaches in mammography images for false positive reduction, Expert Syst. Appl. 40 (18) (2013) 7534–7543 .

3) W.B. Sampaio , E.M. Diniz , A.C. Silva , A.C. De Paiva , M. Gattass , Detection of masses in mammogram images using cnn, geostatistic functions and svm, Comput. Biol. Med. 41 (8) (2011) 653–664 .
4) Mohamed A. Berbar, "Hybrid methods for feature extraction for breast masses classification", Egyptian Informatics Journal 19 (2018) 63–73
5) Shankar Thawkar, Ranjana Ingolikar, "Classification of masses in digital mammograms using Biogeography-based optimization technique", Journal of King Saud University – Computer and Information Sciences (2018),
6) D.R. Ericeira , A.C. Silva , A.C. De Paiva , M. Gattass , Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions, Comput. Biol. Med. 43 (8) (2013) 987–999.
7) Prabhpreet Kaur, Gurvinder Singh, Parminder Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification", Informatics in Medicine Unlocked 16 (2019).
8) W.B. de Sampaio , A.C. Silva , A.C. de Paiva , M. Gattass , Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylo- genetic trees, lbp and svm, Expert Syst. Appl. 42 (22) (2015) 8911–8928
9) Mohamed Abdel-Nasser, Antonio Moreno, Domenec Puig, "Temporal mammogram image registration using optimized curvilinear coordinates", computer methods and programs in biomedicine 127(2016) 1-14.
10) Raghavendra, U., Rajendra Acharya, U., Fujita, H., Gudigar, A., Tan, J.H., Chokkadi, S., 2016. Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images. Appl. Soft Comput. J. 46, 151–161.
11) M. Malignant and benign mass segmentation in mammograms using active contour methods. Symmetry 2017; 9(277):1–22.
12) Khehra, B., Pharwaha, A.,. Comparison of Genetic Algorithm, Particle Swarm Optimization and Biogeography-based Optimization for Feature Selection to Classify Clusters of Microcalcifications. J. Inst. Eng. (India): Ser. B 98 (2), 2017189– 202.
13) R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Transactions on Systems, Man and Cybernetics, vol. 3, no. 6, pp. 610–621, 1973.
14) S. Duraisamy, S. EmperumalComputer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifierDeep Learning in Computer Vision, 11 (8) (2017), pp. 656-662,
15) A. Rakhlin, A. Shvets, V. Iglovikov, A.A. KalininDeep convolutional neural networks for breast cancer histology image analysis A. Campilho, F. Karray, B. ter Haar Romeny (Eds.), Image analysis and recognition. ICIAR 2018. Lecture notes in computer science, vol. 10882, Springer, Chamdoi (2018
16) L. Johnson, D. Ritter, Observation of periodic waves in a pulse-coupled neural network, Optical Letter 18 (1993) 1253.
17) H.S. Ranganath, G. Kuntimad, J.L. Johnson, Pulse coupled neural networks for image processing, in: Proc. of the Southeast Conference on 'Visualize the future', 1995, pp. 37–43.

18) Zhaobin Wang, Yide Ma, Feiyan Cheng, Lizhen Yang, “Review of pulse-coupled neural networks”, Image and Vision Computing 28 (2010) 5–13.
19) Keiron O'Shea, Ryan Nash, “An Introduction to Convolutional Neural Networks”, 2015.
20) Albawi, Saad & Abed Mohammed, Tareq & ALZAWI, Saad. (2017). Understanding of a Convolutional Neural Network. 10.1109/ICEngTechnol.2017.8308186.
21) Jianxin Wu, “Convolutional neural networks”,LAMDA Group, National Key Lab for Novel Software Technology, Nanjing University, February 11, 2020.
22) Dalalyan AS, Hebiri M, Lederer J. On the prediction performance of the lasso. Bernoulli 2017;23:552–81.
23) A. Jouirou, A. Baâzaoui, and W. Barhoumi, ``Multi-view information fusion in mammograms: A comprehensive overview,'' Inf. Fusion, vol. 52, pp. 308_321, Dec. 2019.
24) V. Sridevi, J. Abdul Samath, “A Survey on Breast Cancer Segmentation and Classification Using Several Methods” in International Journal of Scientific Research in Computer Science Applications and Management Studies, Volume 8, Issue 3 (May 2019)
25) Gustavo Carneiro, Jacinto Nascimento, and Andrew P Bradley, “Unregistered multiview mammogram analysis with pre-trained deep learning models,” in MICCAI. Springer, 2015, pp. 652–660.
26) Neeraj Dhungel, Gustavo Carneiro, and Andrew P Bradley, “Deep learning and structured prediction for the segmentation of mass in mammograms,” in MICCAI. Springer, 2015, pp. 605–612.
27) Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks.,” in NIPS, 2012, vol. 1, p. 4.
28) Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” arXiv preprint arXiv:1512.03385, 2015.
29) Neeraj Dhungel Gustavo Carneiro, and Andrew P Bradley, “Automated mass detection in mammograms using cascaded deep learning and random forests,” in DICTA. IEEE, 2015, pp. 1–8.
30) V. Sridevi, J.Abdul Samath, Advancement on Breast Cancer Detection Using Medio-Lateral-Oblique (Mlo) and Cranio-Caudal (CC) Features”, Volume 83, May - June 2020, pp. 85-93.
31) Christian Szegedy and et al., “Going deeper with convolutions,” in ICCV, 2015, pp. 1–9.