AN IMPROVED TRANSFER LEARNING APPROACH TOWARDS BREAST CANCER CLASSIFICATION ON DEEP RESIDUAL NETWORK

Rajasree.P.M

Assistant Professor, EIE Department, RV College of Engineering, Mysuru Road, RV Vidyaniketan Post, Bengaluru, Karnataka, 560059, INDIA.
rajasreepm@rvce.edu.in

https://www.rvce.edu.in/

Dr. Anand Jatti

Associate Professor, EIE Department, RV College of Engineering, Mysuru Road, RV Vidyaniketan Post, Bengaluru, Karnataka, 560059, INDIA.
anandjatti@rvce.edu.in

https://www.rvce.edu.in/

Dr. Divya Santosh

Consultant Breast Radiologist, Sri Shankara Cancer Hospital & Research Centre, Bengaluru, Karnataka,560004, INDIA.
divyasantosh@gmail.com

Abstract

Breast Cancer is considered a significant health threat in women of all ages, and the incidence has been rapidly increasing, particularly in the last decade. Usually, for diagnosis, characteristics such as skin texture, color (redness), size variation, pain, and gene mutation are used. Binary classification is the most frequent classification type to classify benign and malignant tumors. A deep learning-based mechanism has been hugely popular and significant improvisation in binary classification, especially CNN-based mechanism. Moreover, to take CNN advantage, several breakthroughs have been observed for image classification, such as integrating layer-based features with the classifier. However, due to limited images available for medical research, several areas are exposed, such as overfitting issues, memory issues, and network architecture issues. Hence, in this research work, the design and development of an Improvised-DeepResidualNetwork, aka IDRN, for binary classification on mammogram. IDRN adopts the Deep Residual Network and enforces the improved transfer learning approach for training and testing the model. IDRN follows the novel architecture, forcing the transfer learning-based task-specific layer to be integrated to enhance the features and improvise the classification process. Further, IDRN is evaluated considering the INbreast dataset considering the important metrics like accuracy, precision, recall, F1-score, and AUC; a comparative analysis is also carried out with a state-of-art model of deep learning model.

Keywords: **Breast Cancer, Residual Network, Transfer Learning.**

1. Introduction

Breast Cancer is one of the dreadful diseases which affects women [1]; in the general breast, lesions are classified into two distinctive categories, i.e., benign and malignant, depending on their respective stage. Breast lesions develop through a slow process in normal breast tissue, which further causes to generate the cancer lump. These stages are determined following the BI-RADS score [2][3]. Last year, Breast Cancer was also categorized as one of the mischievous diseases that account for nearly 30% of novel cancer patients and almost 15% of death. However, early prediction and classification simultaneously help in an increase in survival rate [4].

The detection and classification for many patients require to face significant issues as it involves radiology experts and other causes massive effort and plenty of time; meantime, radiologists have been careful while examining all mammogram views. Mammogram views exist in two categories, i.e., CC aka craniocaudal and

Medio-Lateral Oblique aka MLO of each breast to inspect the potential tumor for identifying the potential tumor [1,3,5].

Considering the shortcoming of the traditional approach in terms of accuracy, a deep learning approach was adopted to improvise Breast lesion detection and classification accuracy. Thus CAD (Computer-Aided Diagnosis) is designed to automate the method of detection and classification, however, it is observed that the second opinion of radiologists is highly recommended for confirmation. Thus, the work was to focus on the automated mechanism of CAD, which is achieved through the deep learning approach as described in [3] [5], and through the analysis, unanimously, researchers and academic people agreed that deep learning is probably the best approach for detection and classification. Figure 2 presents the typical process of the CAD model.

The Deep learning approach follows three distinctive stages i.e., pre-processing, utilization of parameters, and feature extraction with recognition [3] [5]. Through the deep learning approach, high-level feature maps and low-level features can be obtained directly from input breast image [9]; initial adaption of deep learning for breast lesion recognition and detection were focused in [10] [11] and performed better than the traditional approach. An automated approach for breast lesion detection is one of the important tasks which improvises the diagnosis of breast cancer; other absolute detection of breast lesion plays an important role in achieving the high-TPR (True Positive Rate), lesion detection is a challenging task due to varies in size, location, shape, and texture. Classification is considered as the final step of the CAD mechanism; the main aim of this step is to identify the detected and segmented ROI as malignant or benign; the classification approach depends on the derived features which present the primary characteristics of the lesion [12] [13]. To recognize benign or malignant, a deep learning mechanism requires fair characteristics distribution. Figure 1 shows the benign and malignant breast images, which are classified according to the BI-RADS score.

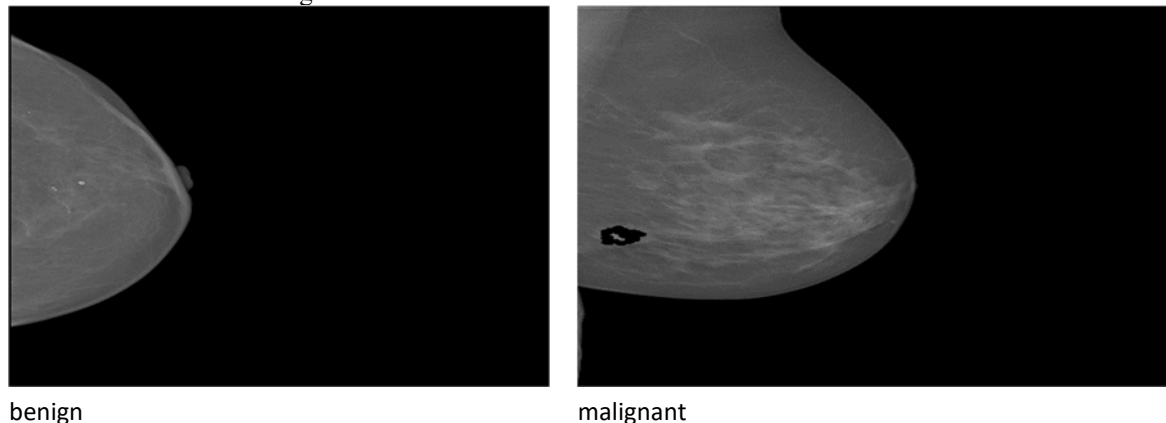benign   malignant

Figure 1 benign and malignant image

Further research was carried out on the breast density classification, and a group of researchers develops the BI-RADS score, which parts the breast into four distinctive categories. BI-RADS 1 represents the lowest densities, and BI-RADS 4 represents the highest densities; in BI-RADS 1, the fat percentage remains between 0 to 25%, BI-RADS 2 have a fat percentage with fibro glandular tissue of 26 to 50%, BI-RADS 3 presents the heterogeneous breast of density 51 to 75%, at last BI-RADS 4 is exceptionally dense with 76 to 100%. Patients that have BI-RADS 4 have a higher risk of developing breast cancer than BI-RADS1 [14] [15].

Hence, considering the phenomena, Breast density plays an important role in early detection; However, radiologists must inspect vast numbers of mammograms, which is timeconsuming and subjective, with low intra- and interradiologist reproducibility. It is critical to developing an automated system that can reliably identify mammographic densities.
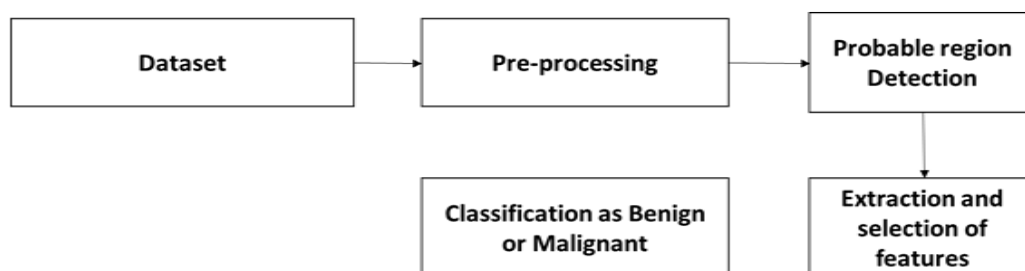

Figure 2 typical CAD process

2. Literature Survey.

In this section of the research, a review of the various existing classification model of breast cancer is proposed, with research having focused on computer-aided tools over the last two decades. Early research in [16]-[19] aimed to ease the workload of radiologists by avoiding the fine needle aspiration. In general, these CAD model comprises four distinguish steps which include pre-processing, probable ROI (Region of Interest) Identification, FE (Feature Extraction) and task-based on these extracted features; out of these steps feature extraction is considered to be more vulnerable. Hence considering the importance of feature extraction, a plethora of work has been carried out and novel methodologies and networks were designed. Further, the development of the Deep learning domain, especially neural network architecture like Convolutional Neural networks, provided a significant revolution in the feature extraction domain. However, it required a fair amount of data and an ideal mathematical model where CNN learns through an iterative approach. In recent, Deep Brief Network [1] was one of the improvised networks designed; this was a basic CAD-based model developed to identify and distinguish the type of cancer into benign and malignant; however, this mechanism requires a conventional CAD model. DBN mechanism was compared with traditional classifier, LDA (Linear Discriminant analysis) based classifier and other DNN (Deep Neural Network) based classifier and prove to be better than these classifiers. DBN-classifier were improvised through integrating the convolution layer in [17]; here, deep features were extracted through the convolutional layer from given input images instead of manual FE.

The extracted features were enforced into the DBN-classifier and perform better performance than the other mechanism, still possesses various challenges such as the construction of feature map. Thus [20] adopted feature map-based extraction instead of a unit feature vector. They focused on two types of feature maps, i.e., high level and low level, through various convolution layers, and further were integrated for training. These derived features were passed through the SVM classifier and possess impressive accuracy. To take advantage of the multilayer feature, [22] adopted a SoftMax layer along with a multilayer perceptron classifier. Although this mechanism is claimed to be deep learning-based CNN architecture adoption, it still requires manual detection of ROI. Hence, the research was focused on an automated model for ROI for whole breast images, and these ROI detections are directly passed for FE and classification steps.

With the rise in the competition of MIA (Medical Image Analysis), several architectures like VGG-16, Alex-Net, GoogleNet, Inception, ResNet, and DenseNet were successfully deployed for breast classification [23]-[28] respectively. [8] used the VGG-16 model and performed a detailed comparative analysis with ResNet-50 and Inception-V3 model; here, FCN was adopted as the classifier for an above-discussed mechanism. Also, the probable region was detected through a manual approach, and later it was passed to the classification model.

Deep Learning domain has proved to be better than shallow mechanism; however, two major challenge remains major issue first is with the manual process, and another one is complexity in the training process. Only a few mechanisms could develop fully integrated models as in [2] [3]and [6]; these models were designed with a three-step approach, namely pre-processing, segmentation, and classification. To detect the lesion cascade DL, which uses GMC and modified-DBN, another model used dual R-CNN, and the model uses CRF (Conditional Random Forest). Extracted lesions were enforced to optimize False Positive per Image, aka FPI. Nevertheless, segmentation was carried out through complex –DL, including contour refinement and dual DBNS integrated with CRF classifier. Furthermore, other work like [29] used dilated convolution for image classification instead of classic CNN used by the above model; the main issue with the classification model is a redundancy of the model in a large number of feature maps. Thus, to avoid attention mechanism were used which highlights the desired information and suppress the redundant; [30] uses spatial attention for essential regions. The improvisation is carried out in [31] and [32]. Also, another recent model in [33] stratification approach is used for the classification and deep learning-based stacked ensemble model in [33] and random projection algorithm in [34] and improvised version of dilated convolution in [35].

Through the detailed review, it was observed that CNN was heavily adopted for classification of breast cancer; there are several research gap which needs to be focused such as classification results in terms of performance metrics, memory complexities, real-time application, and long prediction is a significant challenge. Thus this research analyses the importance of residual network and adopts the DeepResidual-Network, and enforces the improvised transfer learning for improvisation in performance metrics and flexibility.

The importance of network depth is prominent in designing a neural network-based architecture; however, the deeper network remains challenging to train due to the network complexities, and this can be overcome through the approach of residual learning and eases the network training deeper, which improvises the image processing task like breast cancer classification.
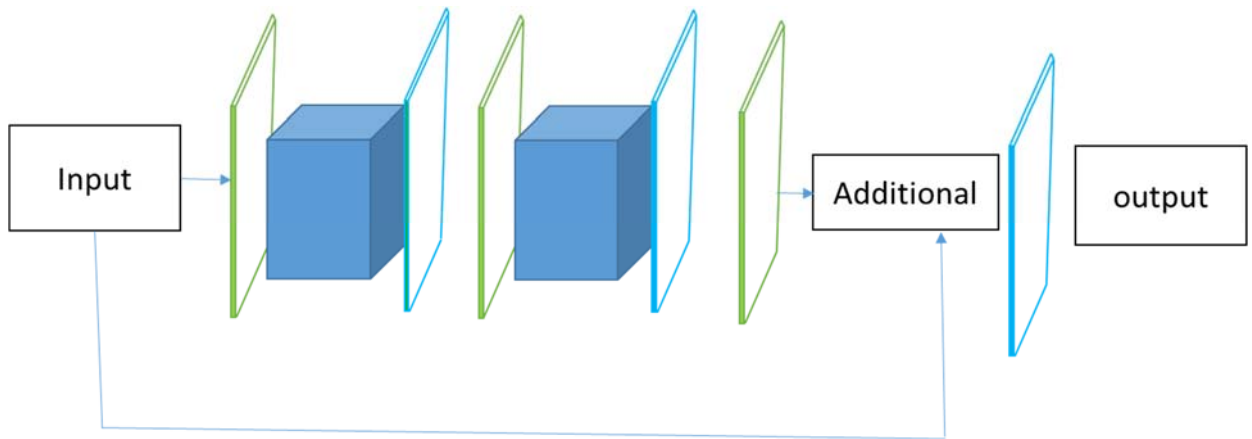
Figure 3 Residual block

The main advantage is they require small parameters to train as parameter optimization is carried out along with task-specific layers; when the models are trained deeper, their accuracy gets saturated and reduces shockingly, which is not caused through the overfitting and adding further layer causes the higher training error. Deep Residual Network is an improvised version of convolution neural network that has the residual network connection and deep architecture; as shown in figure 3, this residual network comprises the lesser convergence without overfitting.

2.1. *Formation of residual blocks*

Here, we consider a residual network with series of residual blocks; these blocks comprise various convolutional layers, and the residual block can be written as:

$$G^{k+1} = UnitFunc(mf(G^k, V^k) + V^k) \tag{1}$$

In equation(1), G^k indicates the input of residual block whereas G^{k+1} represents the output of the residual block. $UnitFunc(.)$ indicates the unit function. mf represents the function that maps the residual layer and V^k Indicates the parameter of residual. It was observed that $mf(G^k, V^k)$ can be either equal or unequal; so when they are unequal, the equation (1) can be reformulated as:

$$G^{k+1} = UnitFunc(func(G^k, V^k) + Z^k G^k) \tag{2}$$

The convolution layer comprises many neurons, and these neurons consist of bias and weights; these convolution layers consider the receptive fields of a previous layer as the input and given as:

$$X_i^h = Af\left(\sum_{l=1}^{L} a_i^h + (v_{il}^h * x_l^{h-1})\right) \tag{3}$$

In equation (3), v_{il}^h represents the weight of feature map, Af as the activation function and a_i^h as bias, also rectified linear is used as the activation function, and the network is replaced from two convolution layers to $1 \times 1 + 3 \times 3 + 1 \times 1$, i.e., dimension is reduced and restored later. Thus, this phenomenon can be formulated as:

$$G^K = G^K + \sum_{h=1}^{K-1} mf\left(G^h, v_{il}^h\right) \tag{4}$$

In equation(4), a sum of input of complex mapping and residual input is expressed through variable K; the backpropagation is derived as:

$$\frac{\partial \alpha}{\partial G^k} = \frac{\partial \alpha}{\partial G^K}\left(\frac{\partial}{\partial G^K} \sum_{h-1}^{K-1} mf(G^K, V^k) + 1\right) \tag{5}$$

2.2. Enhanced-Deep Residual network

- In this section, we present the enhancement of deep residual as unsuitable gradient possesses adverse effects, we propose supervision of the hidden layer training layer; in here, if features are hidden in a designed network, then features maps are trained in discriminative fashion, and feature map of this layer is used as the final classification. The intermediate and low feature maps are considered as additional de-convolutional layers, and other full-feature maps are given for prediction.

The w_l is considered as the weight of the main network and $r(S_h|w_h; V)$ indicates the prediction probability and loss can be formulated as:

$$\varpi(V; W) = \sum_{w_h \in W} -logo(V; S_h|w_h) \tag{6}$$

In the equation (6), W indicates training dataset and S_h indicates target label of an input image w_h.

$$\varpi_c(V_c; v_c'; W) = \sum_{w_h \in W} -logo(S_h|w_h; V_c; v_c') \tag{7}$$

An objective function is designed as per equation (8).

$$\varpi = \varsigma \left(\sum_{c \in C} \|v_c'\|^2 + \|V\|^2 \right) + \sum_{c \in C} \varrho_c \varpi_c (W; V_c; v_c') + \varpi(W; V) \tag{8}$$

To convolute the absolute path for the output layer

Transfer Learning

Transfer learning is an approach where a particular model developed for the task is again used for the initial of the second task; in this research work, the focus is on developing the improved transfer learning approach for enhancement in terms of metrics.

2.3. Domain adaption

Let's consider the source domain and target domain as $C^r = \{(w_h^r, x_h^r)\}_i^{m_r}$ and $C^r = \{W_h^r\}_{i=1}^{m_s}$ with labeled and unlabeled as m_s and m_r; these two are characterized with estimated distribution with variables as o and p. To construct the network which can learn the features and form the classifier $x = \Upsilon(w)$ that optimizes $\eta_S(\Upsilon) = \Gamma_{(w,x) \simeq}[\Upsilon(w)]$ through supervision.

2.4. Network Design

To design the network, the assumption is made that there is no loss of generality; also, let's consider Υ_i^r and Υ_i^s is a parameter for weights and bias respectively of a given filter. The model is trained by optimizing the loss function.

$$K(W^r, W^s, X^s|\Upsilon^r, \Upsilon^s) = K_{IDRN} + K_r + K_s \tag{9}$$

$$K_r = \sum_{h=1}^{M^r} b(\Upsilon^r |W_h^r, x_h^r) \left(\frac{1}{M^r}\right) \tag{10}$$

$$K_s = \sum_{h=1}^{M_k^s} b(\Upsilon^s|W_h^r, x_h^r) \left(\frac{1}{M_i^s}\right) \tag{11}$$

$$K_v = \sum_{i \in \rho} q_v(\Upsilon_i^r, \Upsilon_i^s) (\varphi_v) \tag{12}$$

$$K_{IDRN} = ((\Upsilon^r, \Upsilon^s|W^r, W^s) (\varphi_v q_t) \tag{13}$$

In equation (13), $b(\Upsilon^r |W_h^r, x_h^r)$ indicates the is classification loss.

2.5. Improved domain adaption

Let's consider an architecture that has the three fully connected layer and five convolutional layers where each FCN-layer is trained through non-linear mapping $g_h^k = af^k(a^k + V^k g_h^{k-1})$ with its layer l where af^k is the activation along with V^k weights and a^k bias with hidden layers unit and rectifier units for the output layer as shown in figure 4.introducing a general parameter $\varrho = \{a^k, V^k\}_{k=1}^k$

$$\min_{\varrho} \sum_{h=1}^{m_z} I\left(x_h^z, \varrho(w_h^z)\right) \left(\frac{1}{m_z}\right) \tag{14}$$

In the equation(14), I indicates the loss function; $\varrho(w_h^z)$ is the probability where label assignment is carried out, the initial three convolutions learn the generic features which are transferable, whereas the fourth and fifth convolution is domain-based. In here, the first three blocks are freeze and fine-tuned the fourth and fifth.
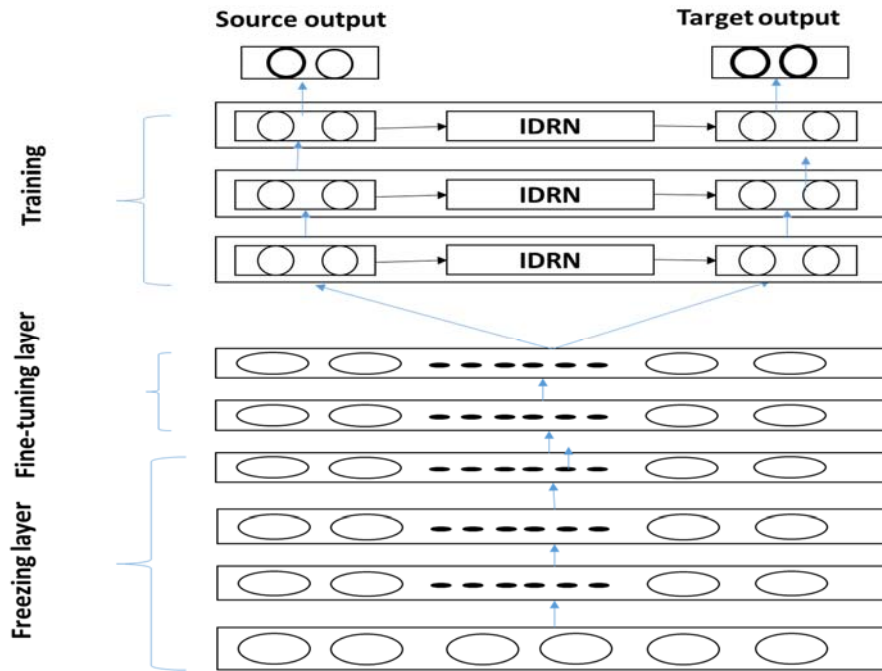

Figure 4 Improvised adaption

Figure 4 presents improvised adaptions; it comprises a three-layer module, i.e., freezing layer module, fine-tuning layer module, and training module. The freezing layer module shall consist of three convolution layers. A fine-tuning layer is made less transferable; hence, these are trained through fine-tune approach; the training layer comprises a complete connected layer deployed for the task. Therefore, they are made non-transferable and should be adopted.

$$\min_{\varrho} \quad \varphi \sum_{k=k_1}^{k_2} c_j^2(C_r^k, C_s^k,) + \frac{1}{m_z} \sum_{h=1}^{m_z} I\left(x_h^z, \varrho(w_h^z)\right) \tag{15}$$

2.6. *Parameter optimization*

In this section, optimization of the parameter is done. At first, the filter function is computed and given through the equation(16):

$$C_j^2(o,p) = D_{w^s w'^s} j(W^s, W'^s) + D_{w^r w'^r} j(W^r, W'^r) - 2D_{w^s w'^s} j(w^r, w^s) \tag{16}$$

In equation (16) w^r is approximately equal to o and the computation causes the time complexity of $O(m^2)$ which is not acceptable for deep learning-based architecture, thus unbiased estimation is carried out, which is given as:

$$c_h^2(o,p) = \sum_{h=1}^{m_r/2} fj\ (\mathbf{y}_h)\ (2(m_r)^{-1}) \tag{17}$$

The equation (17) quad tuple $\mathbf{y}_h \triangleq (w_{2h=1}^r, w_{2h}^r, w_{2i=1}^r, w_{2h}^r)$ is evaluated on multiple filters through the approach:

$$fj(\boldsymbol{y}_h) \triangleq j(w^r_{2h-1}, w^r_{2h}) + j(w^s_{2h-1}\, w^s_{2h}) - j(w^r_{2h-1}\, w^s_{2h}) - j(w^r_{2h}, w^s_{2h-1}) \tag{18}$$

The designed architecture is trained to consider the only gradient for an earlier designed tuple with given data points and the computation of $\frac{\partial_{fh}(\mathbf{y}^k_j)}{\partial \varrho^m}$. A gradient of designed parameter concerning the defined layer is given as:

$$\nabla_{\varrho^k} = \varphi\,\frac{\partial_{fh}(\mathbf{y}^k_h)}{\partial \varrho^k} + \frac{\partial I(\mathbf{y}_h)}{\partial \varrho^k} \tag{19}$$

For a given filter j, as integration of adopted filters denoted as $\{j_t(\mathbf{w}_h, \mathbf{w}_j) = e^{-\|\mathbf{w}_h, -\mathbf{w}_i\|^2/t}\}$; gradient is computed through mathematical chain rule:

$$\frac{\partial l(\mathbf{g}^{rk}_{2j-1,}, \mathbf{g}^{rk}_{2h})}{\partial \mathbf{V}^l} = -\sum_{t=1}^{l} H\left(\left[, \mathbf{g}_{2j\,2h-1}^{\,r(k-1)}\right] - H\left[\mathbf{g}^{r}_{2h}{}^{(k-1)}\right]\right)^R (\frac{2\alpha_t}{\beth_t}j_t(\mathbf{g}^{rk}_{2h-1,}, \mathbf{g}^{rk}_{2h}))(\mathbf{g}^{rk}_{2h-1} - \mathbf{g}^{um}_{2h}) \tag{20}$$

In the equations (20), $H\left(\left[, \mathbf{g}_{2j\,2h-1}^{\,r(k-1)}\right] - H\left[\mathbf{g}^{r}_{2h}{}^{(k-1)}\right]\right)^R$ computes the gradient of a defined layer with g as an indicator such that two considered are considered as given in the following equation:

$$\begin{cases} H\left[\mathbf{g}^{k-1)}_{hi}\right] = \mathbf{g}^{k-1}_{hi} & \text{if } \mathbf{V}^k_i, \mathbf{g}^{k-1}_h + \mathbf{a}^k_j \geq \mathbf{0} \\ H\left[\mathbf{g}^{k-1)}_{hi}\right] & \mathbf{0} \end{cases} \tag{21}$$

2.7.*Error minimization*

Proposed architecture tends to minimize the error by integrating the test power and error observed; this error can be formulated as:

$$\max_{j\in J} c^2_j\,(C^k_s, C^k_r)\vartheta^{-2}_l \tag{22}$$

Where the variance is estimated and computation of same is through the given equation:

$$\vartheta^{-2}_l = \mathbf{D}_y f^2_y(y) - \left[\mathbf{D}_{yfj}(y)\right]^2 \tag{23}$$

Considering the number of filters, covariance is computed as $\mathbb{R} = \text{cov}(h_l) \in \mathbb{S}^{n\times n}$ and equation (23) is minimized to the mathematical equation given as:

$$\min_{\xi^v\,\alpha=1,\alpha\geq 0} \alpha^R\,(\mathbf{P} + \iota\mathbf{H})\alpha \tag{24}$$

2.8. *Analysis*

Considering the basic theory of transfer learning, analysing the proposed mechanism through two distinctive scenarios. The first scenario is that let ϱ be some hypothesis and then estimated risk of source and target domain is given as:

$$\boldsymbol{\iota}_s(\varrho) \leq \boldsymbol{\iota}_s(\varrho) + 2c_j(o, p) + B \tag{25}$$

Where B is the complexity constant, thus, through optimization in an explicit manner for multiple layers, features and classifier learned through the proposed mechanism reduces the upper bound and provides the best-achieved accuracy to date.
Improvised-DeepResidual-Network is evaluated considering the various performance metrics, and evaluation is carried out in the next section.

3. Performance Analysis

Breast Cancer classification is a final and significant task of CAD-based mechanism; in this section of the research, evaluating the IDRN model considering the INbreast dataset. Using CNN based approach, the system's computation and configuration play a vital role; the system configuration model includes 16GB of CUDA-enabled graphics with 2TB of storage and 16 GB of RAM. Machine learning libraries like TensorFlow and CUDA are used for processing; also, python is used as the base language, and other graphs like AUC is plotted through MATLAB 2020.

3.1. *Dataset and other details.*

INbreast is considered to be one of the finest publicly available datasets which can be requested from the [37]; this particular dataset has the ground truth annotations regarding the breast class as benign or malignant along with their coordinates. INbreast Mammograms are considered as the FF(Full Field)-Digital Mammograms, which are obtained through follow-up cases, diagnostic, and screening. The screening process has been carried out following regional and national standards. This dataset comprises 410 Full Field (FF)- Digital Mammograms with three distinctive classes benign, malignant, and normal. It follows the format of DICOM and includes the CC(Craniocaudal) and MLO(Mediolateral oblique) from 115 patients; pixel values have the 14-bit contrast resolution, and the dimension of mammogram is 2560×3328 and 3328×4084.

For evaluation, all mammograms with CC and MLO views are used in a total of 112 cases; also, IDRN uses the BI-RADS score for binary classification of mammograms into benign and malignant. BI-RADS score is a standard numerical scale range of cancer classification between 0 to 6. The INbreast dataset has 36 mammograms which belong to score of 2 and 3 as benign, whereas 76 mammograms with a score of 4,5 and 6 as malignant.

3.2. *Performance Metrics and comparison.*

Any deep learning-based approach is evaluated through a set of metrics which includes accuracy, sensitivity, specificity, F1-Score and AUC, and MCC (Mathews Correlation Coefficient) through confusion matrix; confusion matrix summarizes the various combination of higher metrics suggests the efficiency of the model. The model is evaluated with the same parameter.

The deep learning domain is an ever-growing domain, and in medical image analysis, mechanisms are proposed often; hence to prove the model efficiency, comparative analysis plays an important role. Thus the comparison is made with IDRN models like an improvised version of U-net, which shows fine potential for classification as MS-UNet [38], MS-FCN-8s[39], MS-U-SegNet [40], MS-SegNet [41], MS-Res-CNet [42] and InceptionResNet-V2[43]; these models are considered as the state-o-art model, a further comparison is carried out in the next sub-section.

3.3. *Accuracy*

In general, accuracy can be described as the model efficiency; it is one of the significant metrics to evaluate any classification mechanism; it is computed as the number of all correct prediction which can be formulated as:

$$accuracy = (true_{positive} + true_{negative})/(postitive + negative) \qquad (26)$$

Figure 5 presents a comparison of various mechanisms like MS-U-net, MS-FCN-8s achieves reasonable accuracy of 80.88 and 86.76%; further improvised mechanism of these methods MS-U-SegNet and MS-SegNet achieves good accuracy of 88.24. Similarly inspired by these mechanisms, MS-ResCU-Net achieves an impressive result of 94.12% of accuracy. Further, using the ResNet architecture, InceptionResNet-V2 performs an accuracy of 95.32 %; this method has computed only accuracy and AUC and omitted other metrics, whereas IDRN (Improvised DeepResidual Network) achieves massive accuracy of 98.21%.
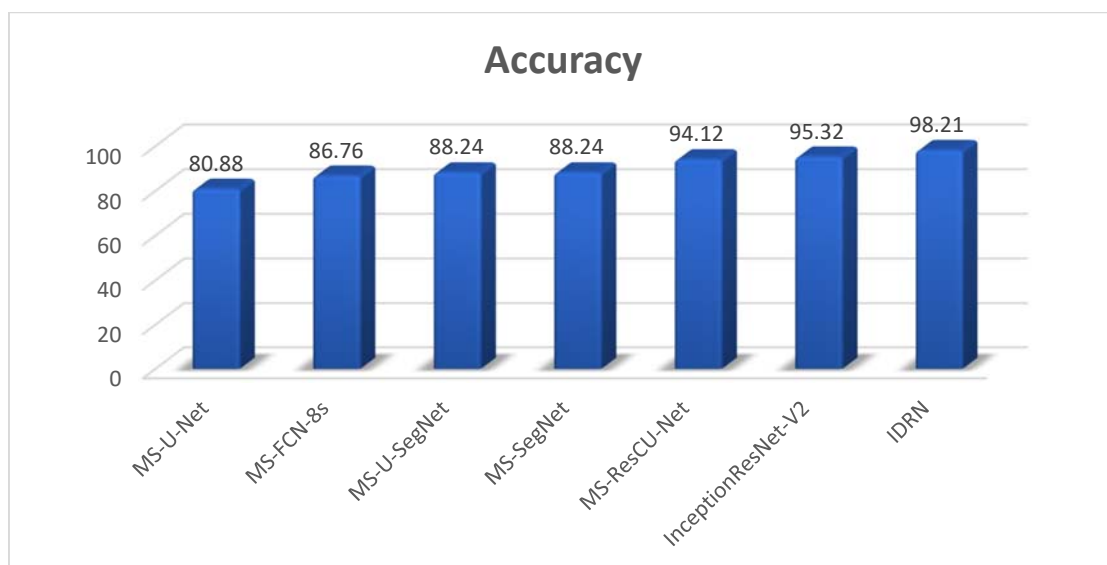

Figure 5. Accuracy Comparison

[1] *F1-Score*

F1-score is defined as the test's accuracy measure; it is one of the major performance metrics for evaluation; it is computed through the equation (27); figure 6 shows the comparison of various algorithms on the F1-score. CNN architecture model MS-U-Net achieves 83.12% and MS-FCN-8s achieves F1-score of 88.89%; another improvised version of U-Net MS-U-SegNet achieves good F1-score of 90.7%, further improvisation is made, and existing model MS-ResCU-Net achieves 95.24% whereas proposed model i.e. IDRN achieves massive F1-score of 98.77%.

$$F1 - Score = \frac{2(true_{positive})}{2(true_{positive}) + false_{negative} + false_{positive}} \qquad (27)$$

Figure 6. F1-score comparison.

Sensitivity

Sensitivity is the statistical measure of binary classification test performance, it measures the actual proportion identified correctly, and the value should be higher. Figure 7 shows the comparison of IDRN with another state-of-art technique model; In here, MS-U-Net and MS-SegNet achieve fair sensitivity of 78.05% and 85.37%, respectively, whereas the improvised architecture of U-net achieves 87.8 and 95.12 % of sensitivity. In comparison with these models, MS-ReSCUNet and achieves similar sensitivity of 97.56%.
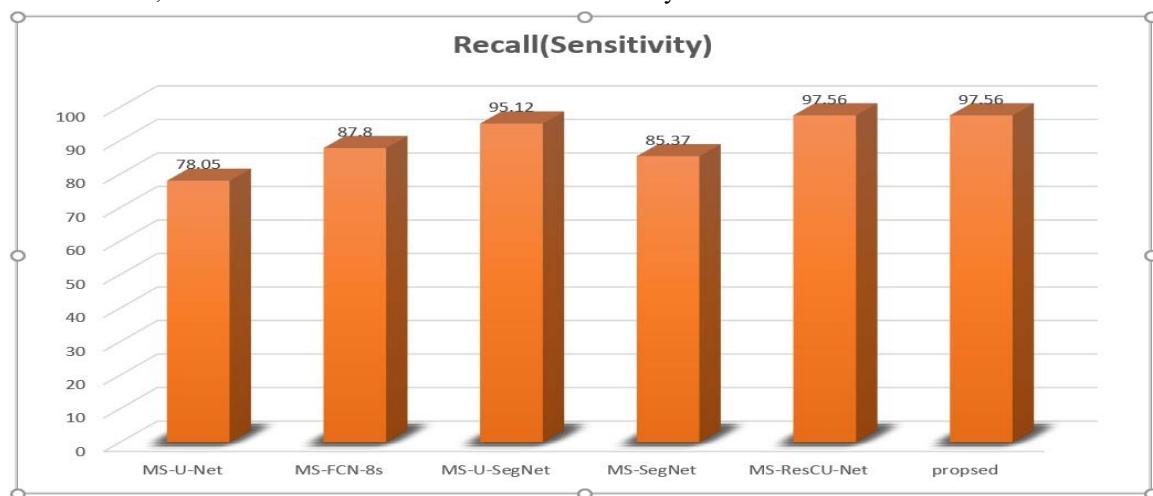

Figure 7: sensitivity comparison

Specificity

Specificity is the measure of actual negative, which is identified correctly, and it is computed as in the given formula concerning the confusion matrix.

$$specificity = \frac{true_{negative}}{true_{negative} + false_{positive}} \qquad (28)$$

Figure 8 presents the comparison of various techniques with the proposed IDRN model; to observe that MS-U-SegNet achieves the lowest 77.78 % of specificity; MS-FCN8s and MS-UNet achieve 85.19% of specificity. MS-RESCU-Net achieves 88.89 % specificity, whereas MS-SegNet achieves 92.59 %; in comparison with all these, the proposed model outperforms all these with a specificity of 93.75 %.
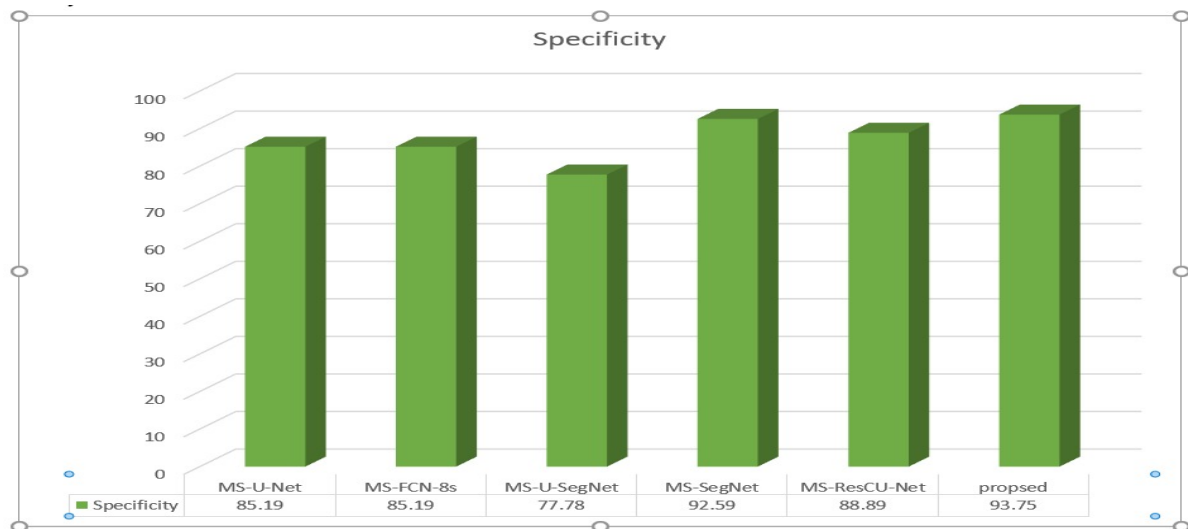

	MS-U-Net	MS-FCN-8s	MS-U-SegNet	MS-SegNet	MS-ResCU-Net	propsed
Specificity	85.19	85.19	77.78	92.59	88.89	93.75

Figure 8: Specificity comparison

AUC (Area under Curve)

AUC-ROC is considered as the major metric for any classification mechanism, and AUC is plotted considering various thresholds; figure.9 presents the AUC graph of a proposed model, which represents the degree of separability. The graph in figure.9 is plotted as true positive against false positive rate.
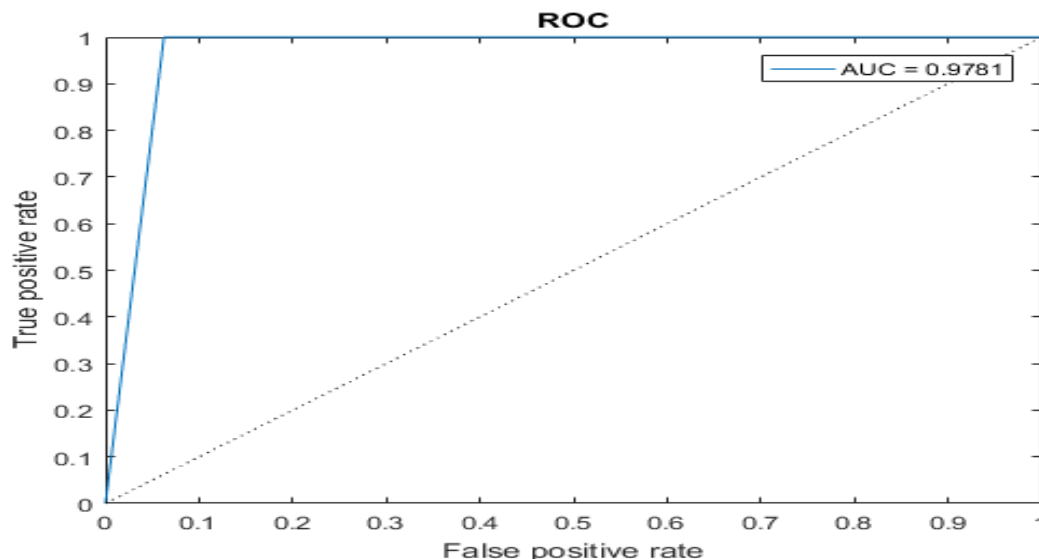

Figure 9: Proposed AUC model

3.4. *Comparative analysis and discussion*

This section analyses the performance of the proposed model IDRN (Improvised DeepResidual Network) through various models, i.e., Accuracy, Sensitivity, Specificity, F1-score, and AUC. To prove the model efficiency, we compare the model with several state-of-art mechanisms, and it is observed that in terms of accuracy, IDRN achieves nearly 3% than the Inception-V-Net model and 4% than the existing model. Moreover, Inception-VNet computed only accuracy and AUC and ignored the other parameter. In terms of F1-score, the proposed model IDRN achieves 3.03% better than the existing model; in terms of sensitivity, IDRN achieves a similar value of 97.56%; in terms of specificity, IDRN achieves 1.19% to the MS-SegNet model. At last, AUC was considered for evaluation and observe that IDRN performs impressive AUC improvisation of 2.8% than the Inception-V3 model

4. Conclusion

MIA (Medical Image Analysis) through Deep Learning has a significant impact in the research world in comparison with the traditional ML-based mechanism; recent development shows the major growth for learning medical images through discovering the feature patterns in the given raw image. The application of the deep learning-based approach focuses on feature extraction, segmentation, detection, and classification; thus, this research adopts the DeepResidual-Network and proposes IDRN for mammography diagnosis; in the previous research of the same work, we have proposed segmentation, which achieved remarkable performance. In this research work, a classification technique is proposed, which classifies the region into benign or malignant i.e. non-cancerous or cancerous, based on the BI-RADS score. IDRN is evaluated considering the INbreast dataset and five important measuring metrics namely Accuracy, precision, specificity, Recall, and F1-score, and outperforms the state-of-art technique and existing methodology. Classification technique plays an essential part in cancer detection as based on the BI-RADS score it decides whether a patient is suffering from cancer or not. Hence to adopt in a real-time environment it must be designed very carefully by considering several circumstances, through this research work a step ahead is taken to detect cancer; the proposed model achieves higher accuracy than any other model to date the research has been carried out. However, the model performs on a lower side in terms of specificity and further optimization can be carried out; several constraints must be considered to see the model performance. Meanwhile, this research work can be concluded by the fact that it is one step towards the classification technique.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] M. A. Al-antari, M. A. Al-masni, S.-U. Park, J. Park, M. K. Metwally, Y. M. Kadah, et al., "An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network," Journal of Medical and Biological Engineering, vol. 38, pp. 443-456, 2018.
[2] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," Medical Image Analysis, vol. 37, pp. 114-128, Apr 2017.
[3] M. A. Al-Antari, M. A. Al-Masni, M. T. Choi, S. M. Han, and T. S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," International Journal of Medical Informatics, vol. 117, pp. 44-54, Sep 2018.
[4] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," CA: a cancer journal for clinicians, vol. 69, pp. 7-34, Jan 2019.
[5] M. A. Al-Masni, M. A. Al-Antari, J. M. Park, G. Gi, T. Y. Kim, P. Rivera, et al., "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system," Computer Methods and Programs in Biomedicine, vol. 157, pp. 85-94, Apr 2018.
[6] G. Carneiro, J. Nascimento, and A. P. Bradley, "Automated analysis of unregistered multi-view mammograms with deep learning," IEEE transactions on medical imaging, vol. 36, pp. 2355-2365, 2017.
[7] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, et al., "Large scale deep learning for computer-aided detection of mammographic lesions," Medical image analysis, vol. 35, pp. 303-312, 2017.
[8] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep Convolutional Neural Networks for breast cancer screening," Computer Methods and Programs in Biomedicine, vol. 157, pp. 19-30, Apr 2018.
[9] M. A. Al-Masni, M. A. Al-Antari, M. T. Choi, S. M. Han, and T. S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," Computer Methods and Programs in Biomedicine, vol. 162, pp. 221-231, Aug 2018.
[10] Rampun, B. Scotney, P. Morrow, H. Wang, and J. Winder, "Breast density classification using local quinary patterns with various neighborhood topologies," J. Imaging, vol. 4, no. 14, pp. 1–23, 2018.
[11] J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern." Cancer, vol. 37, no. 5, pp. 2486–2492, 1976.
[12] N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, and M. J. Yaffe, "Quantitative classification of mammographic densities and breast cancer risk: Results from the canadian national breast screening study." J. Natl. Cancer Inst., vol. 87, no. 9, pp. 670–675, 1995.
[13] M. Kallenberg, K. Petersen, M. Nielsen, A. Ng, P. Diao, C. Igel, C. Vachon, K. Holland, N. Karssemeijer, and M. Lillholm, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring." IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1322–1331, 2016.
[14] W. Huo, G. L. Chew, K. L. Britt, W. V. Ingman, M. A. Henderson, J. L. Hopper, and E. W. Thompson, "Mammographic density—a review on the current understanding of its association with breast cancer," Breast Cancer Res. Treat., vol. 144, no. 3, pp. 479–502, 2014.

[15] W. A. Berg, C. Campassi, P. Langenberg, and M. J. Sexton, "Breast imaging reporting and data system: Inter- and intraobserver variability in feature analysis and final assessment." Am. J. Roentgenol., vol. 174, no. 6, pp. 1769–77, 2000.

[16] Van Ginneken, B.M. ter Haar Romeny, M. Viergever, Computeraided diagnosis in chest radiography: a survey, IEEE Trans. Med. Imaging 20 (12) (2001) 12281241.

[17] Y. Jiang, R.M. Nishikawa, R.A. Schmidt, et al., Improving breast cancer diagnosis with computer-aided diagnosis, Acad. Radiol. 6 (1) (1999) 2233.

[18] H.-P. Chan, K. Doi, C.J. Vybrony, et al., Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computeraided diagnosis, Invest. Radiol. 25 (10) (1990) 11021110.

[19] Tang, Jinshan, Rangaraj M. Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang. " Computer-aided detection and diagnosis of breast cancer with mammography: recent advances." IEEE transactions on information technology in biomedicine 13, no. 2 (2009): 236-251.

[20] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," Communications of the ACM, vol. 54, pp. 95-103, 2011.

[21] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification," Neurocomputing, vol. 197, pp. 221-231, 2016.

[22] Y. Qiu, S. Yan, R. R. Gundreddy, Y. Wang, S. Cheng, H. Liu, et al., "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," Journal of X-ray Science and Technology, vol. 25, pp. 751-763, 2017.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," presented at the arXiv preprint arXiv:1409.1556 2014.

[24] Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[25] Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[26] Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," presented at the arXiv:1602.07261v2 [cs.CV] 23 Aug 2016,, 2016.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778. [25] G. Huang, Z. Liu, and L. Maaten, "Densely Connected Convolutional Networks," presented at the IEEE conference on computer vision and pattern recognition, 2017.

[28] P. Rahmati, A. Adler, and G. Hamarneh, "Mammography segmentation with maximum likelihood active contours," Medical image analysis, vol. 16, pp. 1167-1186, 2012.

[29] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in IEEE CVPR, 2017. pp. 472–480.

[30] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2921–2929.

[31] G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel' Squeeze & Excitation' in fully convolutional networks," in MICCAI, 2018, pp. 421–429.

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in IEEE CVPR, 2018. pp. 7132–7141.

[33] J. Xing et al., "Using BI-RADS Stratifications as Auxiliary Information for Breast Masses Classification in Ultrasound Images," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2020.3034804.

[34] N. Arya and S. Saha, "Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, doi: 10.1109/TCBB.2020.3018467.

[35] M. Heidari et al., "Applying a random projection algorithm to optimize machine learning model for breast lesion classification," in IEEE Transactions on Biomedical Engineering, doi: 10.1109/TBME.2021.3054248.

[36] Li et al., "Multi-View Mammographic Density Classification by Dilated and Attention-Guided Residual Learning," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, doi: 10.1109/TCBB.2020.2970713.

[37] C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," Academic radiology, vol. 19, pp. 236-48, Feb 2012.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015, vol. 9351 of LNCS, pp. 234–241.

[39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, 2015, pp. 3431–3440.

[40] P. Kumar, P. Nagar, C. Arora, and A. Gupta, "U-segnet: Fully convolutional neural network based automated brain tissue segmentation tool," in ICIP. IEEE, 2018.

[41] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481–2495, 2017.

[42] T. Shen, C. Gou, J. Wang and F. -Y. Wang, "Simultaneous Segmentation and Classification of Mass Region From Mammograms Using a Mixed-Supervision Guided Deep Model," in IEEE Signal Processing Letters, vol. 27, pp. 196-200, 2020, doi: 10.1109/LSP.2019.2963151.

[43] Mugahed A. Al-antari, Seung-Moo Han, Tae-Seong Kim,Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms,Computer Methods and Programs in Biomedicine,Volume 196,2020,105584,ISSN 0169-2607,https://doi.org/10.1016/j.cmpb.2020.105584

Authors Profile:

Prof. Rajasree.P.M, is working as Assistant Professor, Dept of Electronics & Instrumentation Engineering, RV College of Engineering, Mysuru Road, Bengaluru-560 059, Karnataka, India. Her areas of interest include Digital Image Processing, Biomedical Instrumentation, Virtual Instrumentation.

Dr. Anand Jatti is working as Associate Professor, Dept of Electronics & Instrumentation Engineering, RV College of Engineering, Mysuru Road, Bengaluru-560 059, Karnataka, India. He holds a Ph.D in (Electrical & Electronics Engineering) from VTU, Belgaum.

His research areas of interest include biomedical signal processing, biomedical image processing, biomedical instrumentation, Digital signal processing, signals and systems, etc.
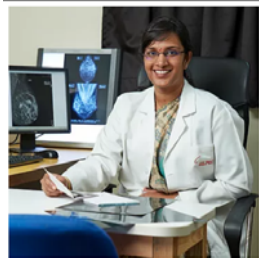
Dr. Divya Santosh FRCR, CCT (UK), Fellowship (Ca), is currently working as Radiologist & Breast Diseases Specialist, at Sri Shankara Cancer Hospital & Research Centre, 1st Cross, Shankaramutt Premises, Shankarapuram, Basavanagudi, Bangalore – 560004.