Estimation of the ambit of breast cancer with a modified ResNet analysis using machine learning approach

Narayanappa C K[1], Poornima G R[2], Basavaraj V Hiremath[3]
{[1]c_k_narayanappa@msrit.edu,[2]poornima_g_r@yahoo.com, [3]bvhiremath@msrit.edu}

[1,3] Department of Medical Electronics & Engineering, Ramaiah Institute of Technology, Bengaluru
[2]Department of Electronics & Communication Engineering, Sri Venkateshwara College of Engineering, Bengaluru

Abstract. Breast Cancer has been one of the most common reasons for mortality and morbidity among the females around the world especially in developing countries. In this regard, Mammography is a popular screening technique for breast cancer diagnosis so as to label the existence of cancerous cells. The present work encompasses the design and development of a M-ResNet (Modified ResNet) approach so as to classify the breast cancer into benign and malignant conditions with the inclusions for supervised classification models with the training of both upper as well as the lower layers of the designed networks. The efficacy of the developed approach was evaluated using various performance evaluators such as those of sensitivity, specificity, accuracy and F1-Score. Bi-Rads score was used as a basis for the classification process wherein a score of 0-3 correlated to benign and it is non-cancerous nature of tissues whereas malignancy was denoted by a score of 4 and above. InBreast dataset, a publicly available online dataset with 112 breast images were used for the evaluation of the developed paradigm. The present paradigm portrayed an accuracy of 96.43% with Area Under the Curve (AUC) of 95.63%.

Keywords: Breast Cancer, classification, benign, Malignant Bi-rads, M-ResNet.

1 Introduction

Breast cancer is one of the most common pathology seen in females, with the statistics suggesting this to be the leading cause of death in the recent times [1]. Although these exist numerous ways to approach this issue, early detection and diagnosis has been the key in every therapeutic approach pertaining to breast cancer, the most common tool being that of mammography [2].However, mammography does not contribute towards early detection, as the normal as well as abnormal tissues seem similar in the initial phases. Early diagnosis of cancer aids in a better treatment plan. Radiological imaging finds its use in cancer diagnosis with mammography being a cost-effective approach [3].Full Field Digital Mammography (FFDM) provides images of good quality with higher contrast resolutions which are extremely helpful in the diagnosis of cancer in subjects with denser breasts, as seen in women of younger age. Cancerous cells are imaged as masses with a diameter of 5-30nm. At times, non-palpable masses of above 20 nm diameter are not elicited as cancerous cells in mammography. Cranial-caudal (CC) and Medio Lateral Oblique (MLO) are the two most common views used to

predict the anatomical structure of the breasts in mammography [4]. Both these views can identify the cancer if the lesions are visible. MLO is commonly used to define the morphological descriptors such as those of architectural distortion, bilateral asymmetry, micro-calcifications and abnormal lesions, based on which further clinical decisions are made. Manual interpretation of mammograms poses numerous challenges due to which CAD based approaches are incorporated with the advent of technology in FFDM based diagnosis. However, CAD based approaches seem to have a higher False Negative (FN) and False Positive (FP) rates due to image based limitations. Overlapping and surrounding approaches with regard to dense parenchymal tissues have resulted in an accuracy of close to 80% in certain case studies. Non-Palpable lesions are identified with histopathological approaches with biopsy. Accurate assessment of lesions can aid in better diagnosis and help the subjects as well. There have been numerous attempts to improve the accuracy of the detection of lesions in case of breast cancers with non-invasive and technology-oriented approaches, trying to avoid invasive aspects such as those of biopsy with machine learning models. Many statistical approaches with 1st and 2nd order statistics have been helpful in the assessment of various disorders such as those of occupational disorders including those relevant to psychoacoustics as well as visual psychophysics [5].These approaches have largely been subjective and the cancer-based assessment approaches require objective assessment with statistical aspects wherein machine learning models find their applications with a higher accuracy in the predictive results pertaining to breast cancer.

2 Background

Several researchers have been focusing in the diagnosis as well as the assessment of breast cancer. In this regard, X Sun considered the image pixels to be an important feature with region growing approach so as to segment the tumor in the mammogram [6]. N Saidin employed Water shed algorithm to obtain a coarser segmentation of the breast tumor with the assessment of the edge of the image based on the integration of the regions with similar gray scale value [7]. M H Yap proposed a novel approach so as to assess the suspicious masses in mammogram images with local thresholding and adaptive global segmentation [8]. M H Yap incorporated various deep learning models such as those of transfer learning method, Patch-based Le-Net, a U-Net with the pre-trained FCN-Alex.Net for the assessment of lesions in breast ultrasound images [9]. M Roberts differentiated malignant and benign tissues in breast images using Bayesian network with CAD approach [10]. Z wang used Extreme Learning Machine (ELM) based approach to categorize the breast tumors and compared the same with the Support Vector Machine (SVM) classifiers [11]. Y Qui used Convolution Neural Networks (CNN) to predict the risks of breast cancer [12]. W Sun introduced Deep Neural Network (DNN) approach to estimate the breast cancers at a 420-time series [13]. Z Jiao incorporated a deep feature-based framework for the assessment of mass of the breast with the CNN and decision-making approaches [14]. J Arevalo used CNN to abstract the breast tumor and then, categorize the same into malignant and benign, with the aid of an automated mammogram analysis using deep learning for the risk evaluations [15]. G Carneiro introduced a novel image retrieval approach with Zernike Moments (ZMs) so as to extract the features affecting the efficiency of CAD based breast cancer recognition [16]. Y Kumar developed the CAD based breast cancer assessment approach incorporating the extraction of significant features and then, to conclude about the nature of the image [17]. Costa, Moura and Kim have

introduced several CAD system-oriented textures-based features so as to ascertain the tumor, to detect the Region of Interest (ROI) and to extract the lesions in breast images. The aspects such as those of Gabor wavelets, Singular Value Decomposition (SVD), Gray-level-co-occurrence matrices (GLCM), Local-binary-pattern-histogram (LBPH), Gray-level-run-length matrices (GLRLM), Gaussian derivatives and discrete wavelets were incorporated in their approaches [18-20]. Zhang, Wei and Vadivel have demonstrated the use of CADx systems based on the shape-oriented features [21-23]. Homer and Sickles used the shape-based features to extract the ROIs and then the coordinates of the same were used as markers. The mathematical descriptors were used to categorize the areas as malignant or benign. The likelihood of malignancy was correlated to the shape of the mass under consideration. While malignant masses were found to be speculated, well-circumscribed, ill-circumscribed, round, micro lobulated and oval, the benign portions were homogenous with well-defined boundaries [24,25]

With the afore mentioned literature, available with regard to the assessment of breast cancers, the present work incorporated a novel approach based on modified ResNet (M-ResNet) algorithm due to the fact that this is based on Bi-Rads score. Deep learning models have been incorporated as well. The standard dataset available online (Source: InBreast dataset) was used to test the developed approach for sensitivity, specificity and accuracy. Details of the algorithm developed as well as the results obtained have been provided in the succeeding sections

3 Materials & methodology

This section provides a complete overview of the novel approaches incorporated in the present work, with the development of M-ResNet model using a dell inspiron laptop with 1 TB HDD, 16 GB RAM and 2GB Nvidia graphics cord.

Basis

Consider two domains (source domain and the destination domain). Equation 1 indicates the source domain and equation 2 indicates the destination domain.

$$\mathbb{E}_t = \left\{ \left(\mathbb{Z}_j^T, \mathbb{Y}_j^T \right) \right\}_i^{o_t} \tag{1}$$

$$\mathbb{E}_u = \left\{ \mathbb{Y}_k^u \right\}_{k=1}^{\mathbb{O}_u} \tag{2}$$

Moreover Equation 1 and equation 2 are characterized through the probability distribution which can be denoted as $\mathbb{Q}$ and $\mathbb{R}$ respectively. Here a neural network can be constructed so as to be used for the cross-domain as well as for the design of a classifier.

$$\mathbb{V} = \omega(\mathbb{Y}) \tag{3}$$

The equation 3 denotes the classifier that can further reduce the risk, similarly the equation 4 is for risk

$$A_v(\omega) = Pr_{(\mathbb{Y},\mathbb{Z})\sim r}[\theta(\mathbb{Y}) \neq \mathbb{Z}] \tag{4}$$

Hence,

$$\mathbb{E}_a = \left\{ \left(\mathbb{Z}_j^b, \mathbb{Y}_j^b \right) \right\} \tag{5}$$

Hence, the modified ResNet can be developed for feature assessment and learning.

Model
Consider a CNN architecture, which has the five layers of convolution and three layers of FC (Fully Connected). Here each layer of FC, learns the non-linear mapping. Further learning is given through the equation 6.

$$\mathbb{I}_j^m = fun^m\left(\mathbb{C}^m + \mathbb{X}^m \mathbb{I}_j^{m-1}\right) \tag{6}$$

In equation 6, h^ℓ is the lth hidden layer of given point x. b^ℓ and W^ℓ are bias and weight of the given layer, f^ℓ indicates the activation layer. Further, the Rectifier units given in equation 2 are considered in case of hidden layer and in case of output layer. The softmax equation is given in equation 8

$$func^m(\mathbb{Y}) = max(0, \mathbb{Y}) \tag{7}$$

$$fun^m(\mathbb{Y}) = e^{\mathbb{Y}}\left(\sum_{k=1}^{|\mathbb{Y}|} e^{\mathbb{Y}}k\right)^{-1} \tag{8}$$

Furthermore, set of parameters used in the architecture is considered and equation 9 presents the same.

$$\Theta = \{\mathbb{C}^m, \mathbb{X}^\ell\}_{m=1}^m \tag{9}$$

Moreover, the learning of the architecture can be represented through the equation 10.

$$\min_{\Theta} \frac{1}{o_b} \sum_{i=1}^{o_b} k\left(\mathbb{Z}_j^b, \theta\left(\mathbb{Y}_j^b\right)\right) \tag{10}$$

In equation 10, k presents the loss function, $\theta\left(\mathbb{Y}_j^b\right)$ presents the probability of assigning of architecture from $\left(\mathbb{Y}_j^b\right)$ to label$\mathbb{Z}_j^b$.

In any standard architecture of convolution neural network, the deep features transits from general to the particular through the least existed layer of network. Later the architecture is fine-tuned by adding the regularized parameter. Hence, equation 6 of learning feature is added with regularization parameter and depicted in the equation 11.

$$\min_{\Theta}(o_b)^{-1} \sum_{j=1}^{o_b} \mathbb{Z}_j^b + \lambda \sum_{m=m_1}^{m_2} e_l^2(\mathbb{E}_u^m, \mathbb{E}_t^m,) \tag{11}$$

λ is penalty function, m_1 and m_2 are the layer index and these layers are selected based on the size of dataset and number of fine tune parameter, $e_l^2(\mathbb{E}_u^m, \mathbb{E}_t^m,)$ is the learning feature between the domain on the given layer.

Selection of the optimal kernel
With the available information being limited, selection of an optimal kernel becomes quintessential for learning. Let $\mathbb{I}_l$ be the reproduced kernel with the Kernel$\mathbb{I}$, further mean embedding is calculated through $\mathbb{Q}$ in $\mathbb{I}_l$ such that it should satisfy the particular equation given below.

$$\mathbb{F}_{\mathbb{Y} \sim \mathbb{Q}} func(\mathbb{Y}\ ) = \left(\vartheta_l(\mathbb{Q}), func(\mathbb{Y}\ )\right)_{\mathbb{I}_l} \tag{12}$$
$$\forall func \in \mathbb{I}_l$$

$E_l^2(\mathbb{Q}, \mathbb{R})$ Is learning model and Distribution between $\mathbb{Q}$ and $\mathbb{R}$ is defined as the reproduced kernel space distance between the Mean Embedding.

$$E_l^2(\mathbb{Q}, \mathbb{R}) \triangleq \left\| \mathbb{F}_{\mathbb{Q}}[\sigma(\mathbb{Y}^T)] - \mathbb{F}_{\mathbb{Q}}[\sigma(\mathbb{Y}^T)] \right\|_{\mathbb{I}_l}^2 \tag{13}$$

Later $\mathbb{Q} = \mathbb{R}$, if and If and only if and only if $E_l^2(\mathbb{Q}, \mathbb{R})$ is equal to zero, hence activation map associated with characteristic kernel is presented in the below equation.

$$\sigma, l(\mathbb{Y}^3, \mathbb{Y}^u) = \left\langle \sigma\left(\sigma(\mathbb{Y}^u, \mathbb{Y}^t) \right) \right\rangle \tag{14}$$

Moreover, the characteristics kernel associated with the activation map is depicted through below equation

$$l \triangleq \left\{ l = \sum_{v=1}^{n} \varepsilon_v l_v : \sum_{n=1}^{n} \varepsilon_v = 1, \right\} \tag{15}$$

$$\varepsilon_v \geq 0, \forall v$$

ε_v Is the COC (Constraints on Coefficient), In order to guarantee the characteristics kernel β_u is imposed so that the multi- kernel l is derived. Furthermore, it is observed that mean map of $\mathbb{Q}$ and $\mathbb{R}$ fails to provide the least test error; hence, multi-kernel is used for the ideal kernel selection

Learning feature algorithm

In this section,a two-step mechanism is developed to learn the origin feature, the first being the parameter learning of the architecture followed by the kernel parameter learning. In order to train the parameter, the kernel trick is used (i.e. from equation 1) and the computed kernel function depicted in equation 16.

$$E_l^2(\mathbb{Q}, \mathbb{R}) = \mathbb{F}_{y^s y'} {}_s l(\mathbb{Y}^t, \mathbb{Y}'^{\,t}) + \mathbb{F}_{y^s y'} {}_s l(\mathbb{Y}^u, \mathbb{Y}'^{\,u}) - 2\mathbb{F}_{y^s y'} {}_s l(\mathbb{Y}^t, \mathbb{Y}'^{\,t}) \tag{16}$$

In the above equations, $\mathbb{Y}^t$ and $\mathbb{Y}'^{\,t}$ is approximately equal to $\mathbb{Q}$, similarly $\mathbb{Y}^u$ and $\mathbb{Y}'^{\,u}$ is approximately equal to q and l belongs to $\mathbb{L}$, however further computation takes long time and causes the obstacle thus to avoid the we use the bias function estimatorthat formulatesthe linear complexity.

$$E_l^2(\mathbb{Q}, \mathbb{R}) = 2(o_t)^{-1} \sum_{j=1}^{o_t/2} \mathbb{h}l\left(\mathbb{A}_j \right) \tag{17}$$

In the above equation quad tuple is used denoted by $\mathbb{A}_j$ and it is given in below equation.

$$\mathbb{A}_j \triangleq \left(\mathbb{Y}_{2j=1}^u, \mathbb{Y}_{2j}^u, \mathbb{Y}_{2j=1}^t, \mathbb{Y}_{2j}^t \right) \tag{18}$$

Further multiple kernel function k is evaluated onthe given tuple $\mathbf{z}_i$ using the below equations.

$$\mathbb{h}l\left(\mathbb{A}_j \right) \triangleq l\left(\mathbb{Y}_{2j=1}^t, \mathbb{Y}_{2j}^t \right) + l\left(\mathbb{Y}_{2j=1}^u \mathbb{Y}_{2j}^u \right) - l\left(\mathbb{Y}_{2j=1}^u \mathbb{Y}_{2j}^u \right) - l\left(\mathbb{Y}_{2j}^t, \mathbb{Y}_{2j=1}^u \right) \tag{19}$$

The above scenario computes the expected independent variable given in equation 16. Further while trainingthe gradient of the objective is considered with respect to the data points $\mathbb{x}_i$, and the gradient of neural network learning $\frac{\partial k(\mathbb{A}_j)}{\partial \theta^m}$ is computed as in equation 20.

$$\mathbb{A}_j^m = \left(\mathfrak{I}_{2j-1}^{um}, \mathfrak{I}_{2j}^{um}, \mathfrak{I}_{2j-1}^{tm}, \mathfrak{I}_{2j}^{tm} \right) \tag{20}$$

$k\left(\mathbb{A}_j \right)$ is the labeled example for the quad tuple

$$k(\mathbb{A}_j) = \sum_{j'} \ \left\{ \mathbb{Y}_{j'}^b, \mathbb{Z}_{j'}^b \right\}, k\left(\theta\left(\mathbb{Y}_{j'}^b \right), \mathbb{Z}_{j'}^b \right) \tag{21}$$

Further we compute the gradient of the objective function w.r.t the given lth layer parameter Θ^m.

$$\nabla_{\Theta^m} = \lambda \frac{\partial_{\mathbb{h} l}\left(\mathbb{A}_j^m\right)}{\partial \Theta^m} + \frac{\partial J(\mathbb{A}_j)}{\partial \Theta^\ell} \tag{22}$$

Now, consider the kernel k as the linear combination of given Gaussian kernels m is given in the below equation.

$$\left\{ l_v\left(\mathbb{Y}_j, \mathbb{Y}_k\right) = e^{-\left\| \mathbb{Y}_{j,} - \mathbb{Y}_k \right\|^2 / \gamma_v} \right\} \tag{23}$$

In the above equation chain rule is employed to obtain the gradient $\dfrac{\partial_{\mathbb{h} l}\left(\mathbb{A}_j^m\right)}{\partial \Theta^m}$

$$\frac{\partial k\left(\mathfrak{J}_{2j-1,}^{tm}, \mathfrak{J}_{2j}^{tm}\right)}{\partial \mathbb{X}^m} = -\sum_{v=1}^{n}\left(\mathbb{I}\left[, \mathfrak{J}_{2j}^{um\ t(m-1)}{}_{2j-1} \right] \right.$$
$$\left. - \mathbb{I}\left[\mathfrak{J}_{2j}^{um\ u(m-1)}{}_{2j} \right] \right)^{\mathrm{U}} X \frac{2\beta_v}{\gamma_v} l_v\left(\mathfrak{J}_{2j-1,}^{tm}, \mathfrak{J}_{2j}^{um}\right) \times \left(\mathfrak{J}_{2j-1}^{tm} - \mathfrak{J}_{2j}^{um}\right) \tag{24}$$

The above equation computes the gradient of the layer , here $\mathbb{I}$ is indicator.

Learning optimal kernel parameter

In this sub-section a model is developed to learn the kernel parameter in efficient way, further this is carried by reducing the type 2 error and reducing the test power, later these two are integrated to observe the desired result and this optimization is given in the below equation.

$$\max_{l \in \mathbb{L}} E_l^2(\mathbb{Q}, \mathbb{R}) \, (\mathbb{E}_t^m, \mathbb{E}_u^m)\sigma_l^{-2} \tag{25}$$

In the above equation, σ indicates the estimated variance and it is computed through the equation 18, where E=$(E_1, E_2, \ldots\ldots, E_n)^U$, further the estimated variance is computed through the below equation.

$$\sigma_l^2 = \mathbb{F}_A h_l^2(\mathbb{A}) - [\mathbb{F}_A h^2(\mathbb{A})]^2 \tag{26}$$

Further, the covariance is computed using the equation 28, whereas equation 27 presents the covariance

$$\mathbb{R} = \text{cov}(h_l) \in \mathbb{S}^{n \times n} \tag{27}$$

$$\mathbb{R}_{vv'} = 4\,(o_t)^{-1} \sum_{k=1}^{o_t/4} h_{l_v}^{\Delta}\left(\overline{\mathbb{A}}_j\right) h_{l_{v'}}^{\Delta}\left(\overline{\mathbb{A}}_j\right) \tag{28}$$

$$\overline{\mathbb{A}}_j \triangleq \left(\mathbb{A}_{2j-1}, \mathbb{A}_{2j}\right) \; and \, h_{l_{v'}}^{\Delta}\left(\overline{\mathbb{A}}_j\right) \triangleq h_{l_{v'}}\left(\mathbb{A}_{2j-1}\right) h_{l_{v'}}(\mathbb{A}_{2j}) \tag{29}$$

Once the covariance is formulatedequation28 is reduced to equation 30 for the optimization problem

$$\min_{\mathbb{E}^{\mathrm{U}}\beta = 1, \beta \geq 0} \beta^{\mathrm{U}} (\mathbb{R} + \varepsilon \mathbb{J})\beta \tag{30}$$

Initially, the regularizer of $\varepsilon = 10^{-3}$is set for ideal problem formulation, later by minimizing the false negative an intermediate optimization is adopted that can update both.

Residual Block Formulation

The ResNet possesses the lower convergence if overfitting is avoided and residual connection helps in accelerating the deep layer convergence. Moreover, huge number of layers are added to maximize the performance and avoid the overfitting. Hence,the residual block is formed which is given in the below equation. Here $\mathbb{I}_m$ is input block and $\mathbb{I}_{m+1}$ is the output block, w_l is parameter, $f(.)$ indicates mapping function

$$\mathbb{I}_{m+1} = \mathbb{V}(func(\mathbb{I}_m, \mathbb{X}_m) + \mathbb{I}_m) \qquad (31)$$

Moreover if $func(\mathbb{I}_m, \mathbb{X}_m)and\mathbb{I}_m$ are not it then $\mathbb{B}_m$ is used for dimension matching and the above equation are written as:

$$\mathbb{I}_{m+1} = Relu(func(\mathbb{I}_m, \mathbb{X}_m) + \mathbb{B}_m \mathbb{I}_m) \qquad (32)$$

These blocks have the convolution layer and contains the neurons with bias and learnable weights. The weights update while training

$$Z_k^j = \tau \left(\sum_{n=1}^{N} \mathbb{C}_k^j + \mathbb{X}_{kn}^j * Z_n^{j-1} \right) \qquad (33)$$

$\mathbb{X}_{kn}^j$ is feature map weight in the given layer, $\mathbb{C}_k^j$ is jth bias in the i-th layer, further$\tau(.)$ is non-linear activation function of convolution layer. $\tau(\mathbb{Y}) = \max(\mathbb{Y}, 0)$ is the activation function.Further the residual block is optimized to reduce the cost along with its dimensions and it is presented in the below equation.

$$\mathbb{I}_m = \sum_{i=1}^{M-1} func\left(\mathbb{I}_j, \mathbb{X}_m\right) + \mathbb{I}_m \qquad (34)$$

L Indicates the RU (Residual unit) and computed as the sum of the mapping and unit. Further we derive the back-propagation from the above equation and depicted in the below equation. Since the optimization in ResNet mainly depends on the back-propagation algorithm.

$$\frac{\partial \theta}{\partial \mathbb{I}_j} = \frac{\partial \mathbb{I}_j}{\partial \mathbb{I}_j}(1 + \frac{\partial}{\partial \mathbb{I}_j} \sum_{i-1}^{L-1} func(h_i, w_i)) \qquad (35)$$

Hidden layerupdation (Iterative manner)
The above model possesses the unstable gradients, hence to avoid that we train the hidden layer discriminatively, further this is achieved through feature feedback in the given hidden layer, the weight is updated.

$$\varphi(\mathbb{Y}; \mathbb{Z}) = \sum_{\mathbb{Z}_j \in \mathbb{Y}} -logp\left(\mathbb{U}_j | \mathbb{y}_j; \mathbb{X}\right) \qquad 36$$

In the above equation, training dataset is indicated by X, T_i indicates the target label of input image $\mathbb{y}_j$. Furthermore, to train the data we use $\mathbb{X}_e$ from $\mathbb{X}_e = (\mathbb{x}_1, \mathbb{x}_2, \mathbb{x}_3, \mathbb{x}_4 \ldots\ldots \mathbb{x}_e)$ which represents the weight for the dth layer feature. Further we compute the auxiliary loss through the below equation.

$$\varphi_e(\mathbb{Y}; \mathbb{X}_e; \widetilde{\mathbb{x}_e}) = \sum_{x_i \in X} -log\mathbb{Q}\left(\mathbb{U}_j | x_i; \mathbb{X}_e; \widetilde{\mathbb{x}_e}\right) \qquad 37$$

Further, the algorithm is reduced as follows, through the backpropagation.

$\varphi = \varphi(\mathbb{Y}; \mathbb{X}\ ) + \displaystyle\sum_{E \in \mathbb{E}} \eta_E \varphi_E\ (\mathbb{Y}; \mathbb{X}_e; \widetilde{\mathbb{x}_e}) + \lambda \left(\displaystyle\sum_{E \in \mathbb{E}} \|\widetilde{\mathbb{x}_e}\|^2 + \|\mathbb{X}\|^2 \right)$	38

In the above equation η_e acts as the balancing weight for φ_e since the weight might decay while training, set of hidden layers are denoted by $\mathbb{E}$. Meanwhile the above equation has three consecutive part, first part comprises the output result, second part is for the intermediate hidden layer result. λ indicates the trade-off parameter.

Datasets

In order to evaluate the M-ResNet, the InBreast dataset, (publicly available) is used. This online dataset has 112 breast images. These images are cropped into the pixels size of 256X256, and contains the ROI (Region of Interest). Moreover, according to the BI-RADS(Breast Imaging Reporting and Data system), there are 75 mass with the BI-RADs, the value of 4,5 and 6 which indicate that it is malignant whereas 37 masses are value of 2 and 3 which are classified as benign. Further, there are two different view namely CC and MLO.

Image Classification
Benign

Ground_Truth	Original _Image	Affected area

Ground_Truth	Original _Image	Affected area
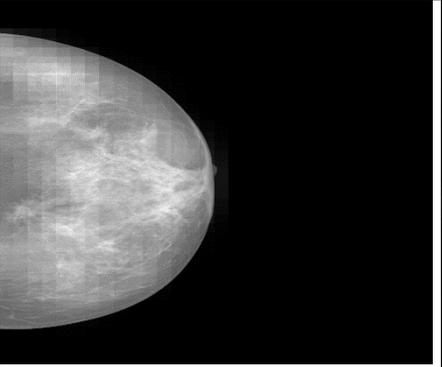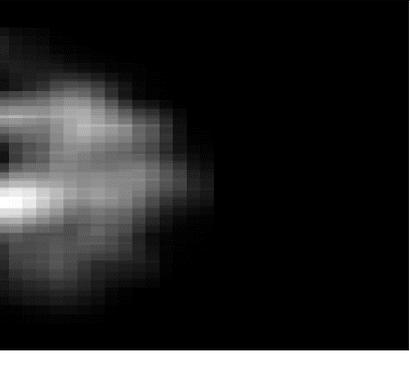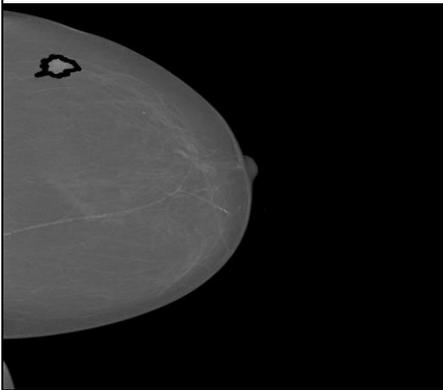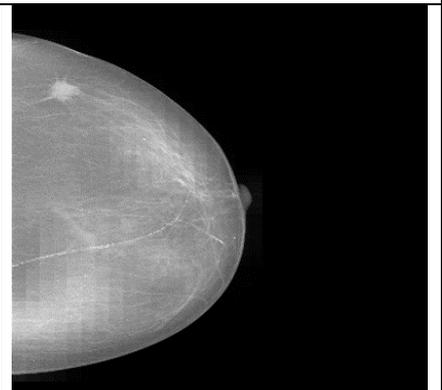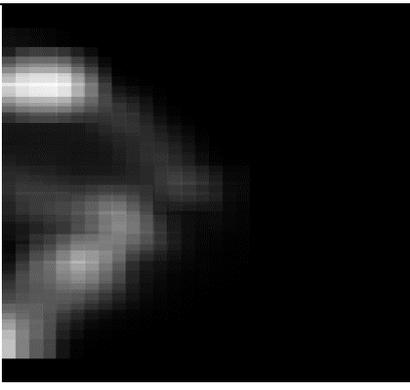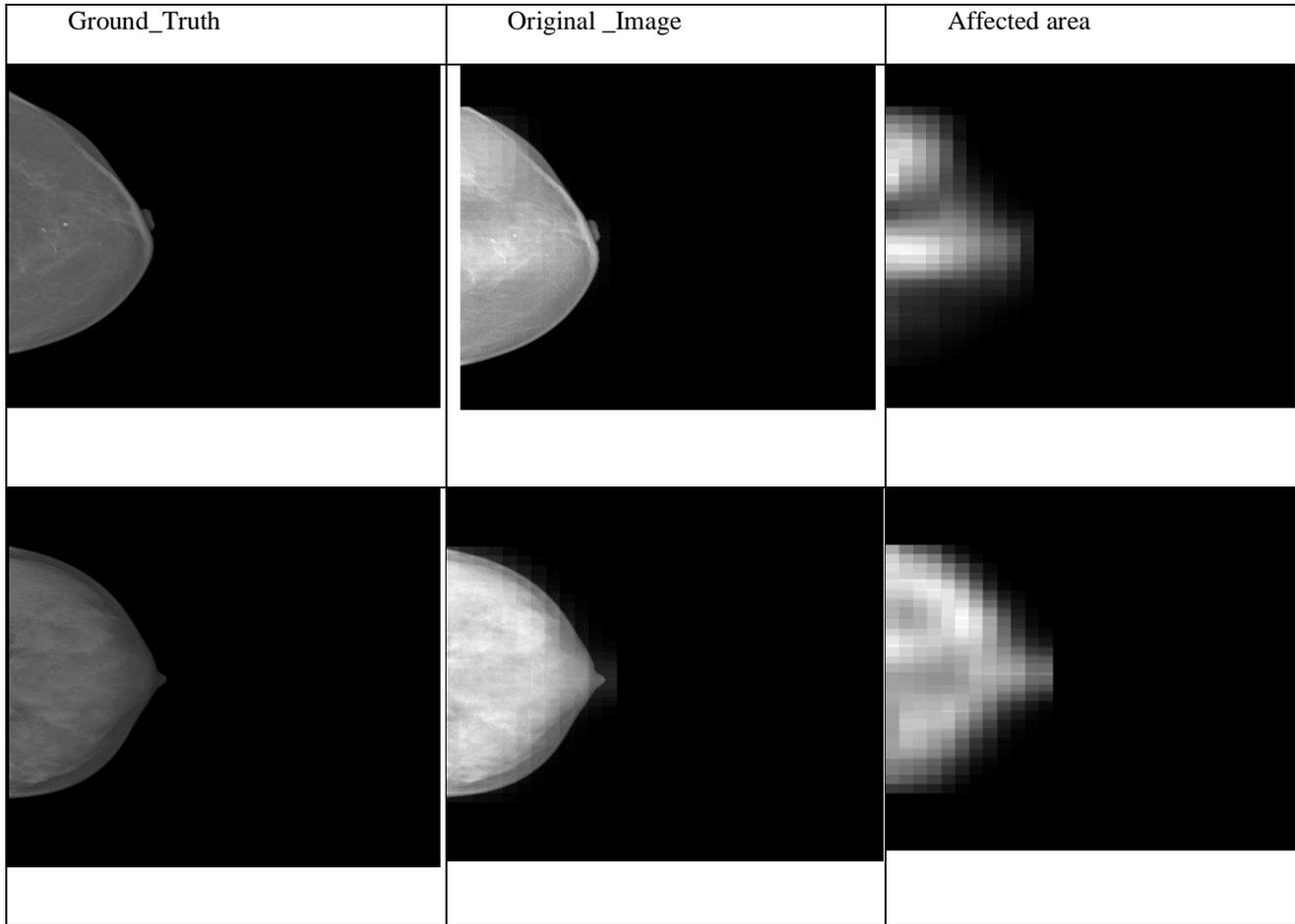
Comparative Analysis

The classification performance of proposed M-ResNet is evaluated by considering the five important measuring metrics namely Accuracy, Sensitivity and specificity and all three metrics are compared with various existing technique and it is depicted in table 1.

Table 1: Performance of classifiers

Methodology	Accuracy	Precision	Specificity	Recall(Sensitivity)	F1-SCore
MS-U-Net[26]	80.88	88.89	85.19	78.05	83.12
MS-FCN-8s[27]	86.76	90.00	85.19	87.80	88.89

MS-U-SegNet[28]	88.24	86.67	77.78	95.12	90.70
MS-SegNet[29]	88.24	94.59	92.59	85.37	89.74
MS-ResCU-Net[30]	94.12	93.02	88.89	97.56	95.24
M-ResNet(proposed)	96.43	95.24	87.50	98.99	97.56

Accuracy

In general, accuracy is referred as the measured value to the standard value. In table 1, the first column shows the different methodologies, second column presents the accuracy. Higher value of the model suggest that the model has been a better classifier. Here it is observed that MS-FCN-8S, MS-SegNet, MS-U-Net, MS-U-segNet achieved the accuracy of 86.76, 88.24, 80.88, 88.24, 80.88, 88.24 and 94.12 respectively whereas in comparison to the same, the proposed model i.e. M-ResNet achievedan accuracy of 96.43 (all these values are observed in percentage).
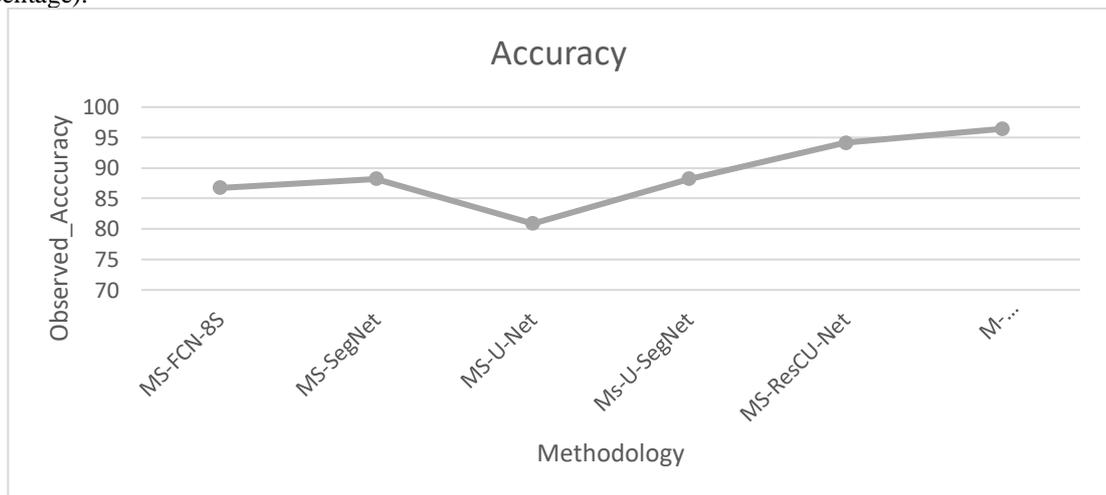
Sensitivity

Sensitivity is the statistical measure of binary classification test performance, it measures the actual proportion identified correctly, and the value should be higher. The third column of table 1 shows the sensitivity of different methodology, in here MS-FCN-8S, MS-SegNet, MS-U-Net,MS-U-segNet achieves the value of 87.80, 85.37, 78.05, and 95.12 respectively whereas proposed model M-ResNet achieves the value of 98.79.

Sensitivity
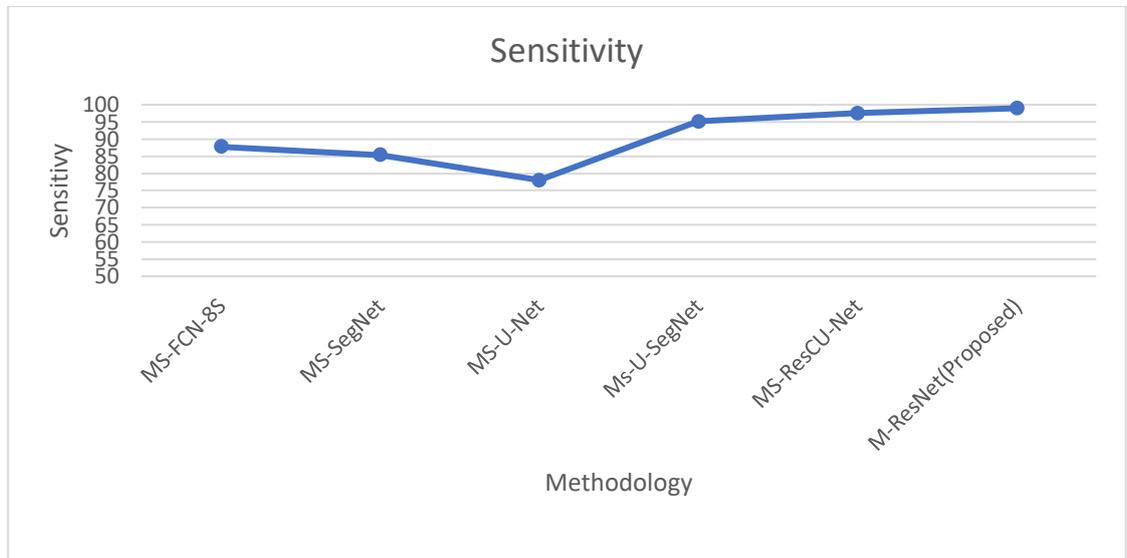
Specificity

Specificity is the measure of actual negative, which is identified correctly, in here MS-FCN-8S; MS-U-Net, MS-U-segNet achieves the value of 85.19, 85.19 and 77.78 respectively. Proposed model achieves the 87.50; here it is observed that existing model achieves the higher value of 88.89. Moreover, MS-SegNet achieves the highest of all i.e. 92.59.

Specificity
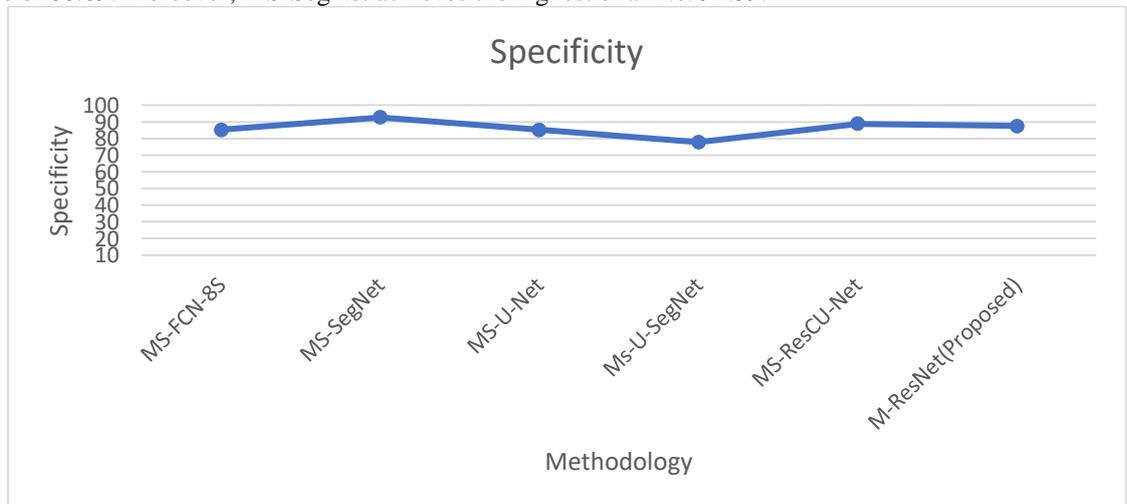
Precision

Precision isconsidered one of the measurement metrics, whichmeasures the closeness to each other, higher precision indicates the better model. In table 1 we observe that MS-U-Net, MS-FCN-8s, MS-U-SegNet, MS-U-SegNet and MS-ResCU-Net achieves the precision value of 88.89, 90.00, 86.67, 94.59 and 93.02 respectively whereas our model i.e. M-ResNet achieves higher than all of 95.24.

F1-Score

F1-score also known as F-measure or F-score is the measure of test's accuracy; it conveys the balance between recall and precision. It is considered to be one of the important metrics, higher indicates. In table 1, MS-U-Net, MS-FCN-8s, MS-U-SegNet, MS-U-SegNet and MS-ResCU-Net achieves the precision value of 83.12, 88.89, 90.70, 89.74 and 95.24 respectively whereas our model i.e. M-ResNet achieves higher than all of 97.56.

AUC-ROC curve

AUC-ROC curve is one of the performance measurement metrics for the classification problem and plotted at the different threshold; ROC is the probability curve whereas AUC presents degree of separabilityFurthermore,it indicates the how much capable model is to distinguish between the class. Higher AUCindicates better model prediction, in our case higher AUC indicates that our model is better in distinguishing between benign and malignant i.e. non-cancerous and cancerous.
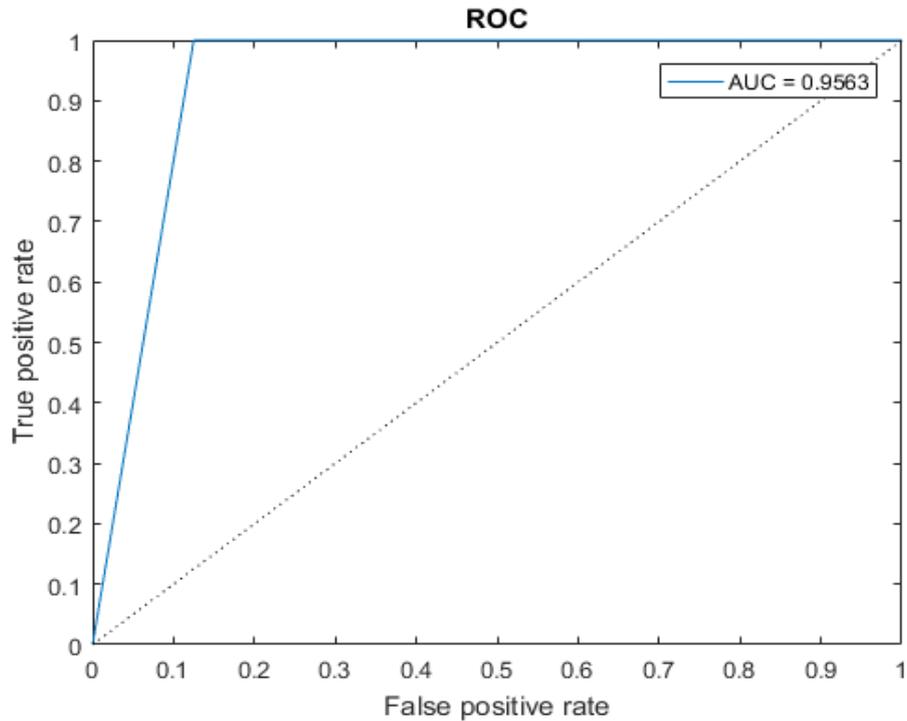
Conclusion

This research presents the novel M-ResNet for mammography-based cancer diagnosis with novel classification approach based on the Bi-Rads score. From the datasets available, five important measuring metrics namely Accuracy, precision, specificity, Recall and F1-score

have been used to quantify the present approach. The results depicted that the novel proposed method outperforms the existing methodology by achieving an accuracy of 96.43, 95.24, 87.50, 98.99 and 97.56. Classification technique plays an essential part in cancer detection as based on the bi-Rads score it decides whether patient is suffering from cancer or not. Hence to adopt in real time environment it has to be designed very carefully by considering several circumstances. The present work portrays a higher accuracy and the model perform on lower side in terms of specificity and hence, further optimization can be carried out with several constraints with regard to the performance of the model.

References

[1] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. S. Meester, A. B. M. PhD, and A. J. D. PhD, "Colorectal cancer statistics, 2017," Ca A Cancer Journal for Clinicians, vol. 67, no. 3, pp. 177–193, 2017.

[2] J. B. Harford, "Breast-cancer early detection in low-income and middle-income countries: do what you can versus one size fits all," Lancet Oncology, vol. 12, no. 3, pp. 306–312, 2011.

[3] Xuejun Sun, Wei Qian, Dansheng Song and A. C. Robert, "Ipsilateral multi-view CAD system for mass detection in digital mammography," Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001), Kauai, HI, USA, 2001, pp. 19-26.

[4] Sanjay H S, Prithvi B S, Nikhil M N, "Auditory Temporal Resolution based Psychophysical Evaluation of Healthy Individuals Exposed to Occupational Noise and Solvents", WSEAS Transactions on Acoustic & Music, Vol 5(1), 20-30, (2018)

[5] G. Carneiro, J. Nascimento and A. P. Bradley, "Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning," in IEEE Transactions on Medical Imaging, vol. 36, no. 11, pp. 2355-2365, Nov. 2017.

[6] X. Sun, W. Qian, and D. Song, "Ipsilateral-mammogram computer-aided detection of breast cancer," Computerized Medical Imaging and Graphics the Official Journal of the Computerized Medical Imaging Society, vol. 28, no. 3, pp. 151–158, 2004.

[7] N. Saidin, U. K. Ngah, H. Sakim, and N. S. Ding, Density based breast segmentation for mammograms using graph cut and seed based region-growing techniques. IEEE Computer Society, 2010.

[8] S. Xu, H. Liu, and E. Song, "Marker-controlled watershed for lesion segmentation in mammograms," Journal of Digital Imaging, vol. 24, no. 5, pp. 754–763, 2011.

[9] K. Hu, X. Gao, and F. Li, "Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms," IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 2, pp. 462– 472, 2011.

[10] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," IEEE J Biomed Health Inform, vol. 22, no. 4, pp. 1218–1226, 2017.

[11] K. C. Jr, L. M. Roberts, K. A. Shaffer, and P. Haddawy, "Construction of a bayesian network for mammographic diagnosis of breast cancer," Computers in Biology and Medicine, vol. 27, no. 1, pp. 19–29, 1997.

[12] Z. Wang, G. Yu, Y. Kang, Y. Zhao, and Q. Qu, "Breast tumor detection in digital mammography based on extreme learning machine," Neurocomputing, vol. 128, no. 5, pp. 175–184, 2014.

[13] Y. Qiu, Y. Wang, S. Yan, M. Tan, S. Cheng, H. Liu, and B. Zheng, "An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology," in Medical Imaging 2016: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2016.

[14] W. Sun, T. L. Tseng, B. Zheng, and W. Qian, A Preliminary Study on Breast Cancer Risk Analysis Using Deep Neural Network. Springer International Publishing, 2016.

[15] 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.

[16] J. Arevalo, M. A. G. Lopez, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," Computer Methods and Programs in Biomedicine, vol. 127, pp. 248–257, 2016.

[17] 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, no. 11, pp. 2355–2365, 2017.

[18] Y. Kumar, A. Aggarwal, S. Tiwari, and K. Singh, "An efficient and robust approach for biomedical image retrieval using zernike moments," Biomedical Signal Processing and Control, vol. 39, pp. 459–473, 2018.

[19] Costa, D.D., Campos, L.F., Barros, A.K.: 'Classification of breast tissue in mammograms using efficient coding', Biomed. Eng. Online, 2011, 10, (1), p. 55

[20] Moura, D.C., Guevara Lopez, M.A.: 'An evaluation of image descriptors combined with clinical data for breast cancer diagnosis', Int. J. Comput. Assist. Radiol. Surg., 2013, 8, (4), pp. 561–574

[21] Kim, D.H., Lee, S.H., Ro, Y.M.: 'Mass type-specific sparse representation for mass classification in computer-aided detection on mammograms', BioMed. Eng. Online, 2013, 12, (Suppl 1), p. S3, http://www.biomedical-engineeringonline.com/content/12/S1/S3.

[22] Zhang, Y., Tomuro, N., Furst, J., et al.: 'Building an ensemble system for diagnosing masses in mammograms', Int. J. Comput. Assist. Radiol. Surg., 2012, 7, (2), pp. 323–329.

[23] Wei, C.-H., Chen, S.Y., Liu, X.: 'Mammogram retrieval on similar mass lesions', Comput. Methods Programs Biomed., 2012, 106, (3), pp. 234–248.

[24] Vadivel, A., Surendiran, B.: 'A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories', Comput. Biol. Med., 2013, 43, (4), pp. 259–267.

[25] Homer, M.J.: 'Mammographic interpretation: a practical approach' (McGraw-Hill, Boston, MA, 1997, 2nd edn.), pp. 1–6.

[26] D'Orsi, C.J., Sickles, E.A., Mendelson, E.B., et al.: 'ACR BI-RADS atlas, breast imaging reporting and data system' (American College of Radiology, Reston, VA, 2013 edn.)

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

[28] 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.

[29] 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.

[30] P. Kumar, P. Nagar, C. Arora, and A. Gupta, "U-segnet: Fully convolutional neural network issue segmentation tool,"inICIP. IEEE, 2018.

[31] SanaUllah Khan, Naveed Islam, Zahoor Jan, IkramUd Din, Joel J. P. C Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, Pattern Recognition Letters, Volume 125, 2019, Pages 1-6.