International Journal on Artificial Intelligence Tools
Vol. XX, No. X (2022) 1–25
© World Scientific Publishing Company

An Integrated Framework with Deep learning for Segmentation and Classification of Cancer Disease

Hemanta Kumar Bhuyan *
Department of Information Technology,
Vignan's Foundation for Science, Technology & Research
(Deemed to be University), Guntur, Andhra Pradesh, India
hmb.bhuyan@gmail.com

Vinayakumar Ravi
Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 34754, Saudi Arabia
vravi@pmu.edu.sa

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

Abstract: This paper addresses radiologists' specific diagnosis of cancer disease effectively using integrated framework of deep learning model. Although several existing diagnosis systems have been adopted by a physician, in few cases, it is not so practical to see the infected area from images in the normal eye. Thus, a fully integrated diagnosis framework for disease detection is proposed to find out the infected area from image using deep learning approaches in this paper. In this proposed framework, various components are designed through deep learning approaches such as detection, segmentation, classification etc. based on mass region. The classification technique is used to classify the disease as either benign or malignant. The vital part of this framework is developed by using a full resolution convolutional network (FrCN) that supports different stages of image processing, especially breast cancer disease. Different experimental evaluation is taken to perform on the accuracy, cross-validation tests, and the comparative testing. Since we have taken 4-fold evaluation, the FrCN performs with an average 98.7% Dice index, 97.8% TS/CSI coefficient, 99.1% overall accuracy, and 98.15% MCC. Our experiments demonstrated that the proposed diagnosis system performs on the deep learning approaches at each segmentation stage and classification with good results.

Keywords: Deep learning; segmentation; Classification; Deep neural network; Convolutional neural network.

1. Introduction

Now a days, breast cancer is a crucial form of cancer affecting women with high rates of distressed disease compared to other cancers. Breast Cancer is a leading cause of death

Hemanta Kumar Bhuyan, Department of Information Technology, Vignan's Foundation for Science, Technology & Research, (Deemed to be University), Vadlamudi, Guntur, Andhra Pradesh, India, 522213 hmb.bhuyan@gmail.com
Vinayakumar Ravi, Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 34754, Saudi Arabia, vravi@pmu.edu.sa

worldwide, accounting for nearly 685 000 deaths, and the most common new cases of breast cancer were 2.26 million cases in 2020 as per World Health Organization (WHO) data [1]. The percentage of women's deaths is increased worldwide due to breast cancer. Thus, it needs to minimize the women's death rate using early detection of this disease. Although different techniques are used to detect breast cancer by analyzing mammograms and microcalcifications, different challenges for radiologists are still taken to make automated detection of cancer to identify the patient as either benign or malignant in the diagnosis system. Several researchers have been developing mammography for breast cancer [2,3,4] and thoroughly analyzing the different mammograms [2,5,6]. But they found more false positives (FPs) and false negatives (FNs) after segmentation during the evaluation of different datasets. For which, their performance needs to improve by reducing the number of FPs and FNs. Thus, we aim to reduce the quantity of false data through segmentation using the proposed model. We emphasized on finding the region of interest (ROI) that is more effective during segmentation. We tried to reduce the number of FPs and FNs for different validation tests. Therefore, during the diagnostic process, the validation tests have been considered to avoid anomalies of disease. For these tests, four parameters (i.e., true positive (TP), true negative (TN), false positive (FP) and false-negative (FN)) are taken a crucial role to determine several validation tests, which are explained in the methodology and experimental section. Based on the above parameters, the proposed framework is designed the integrated schemes through mass identification, segmentation, and classification of cancer disease using deep learning approaches. For the above part of the scheme, clinical practices and the expert is needed to verify each process of the diagnostic system to improve the performance. Ground-level masses detect the infected regions based on surrounding tissues. Generally, physicians identify the suspicious masses by the manual or naked eye, but the proposed methods identify the disease more effectively as validation tests.

The proposed integrated system contains the above three components: First part is considered for mass identification using creating a sub-region with ROI. This part is designed with the help of length (W) and width (H) of image and contour of the object in the image, which helps to find the region of cancer infected area of breast images. In the second part, the role of mass segmentation is significant to get a region of the infected area using the region of interest (ROI) by the disease, which needs to be extracted from the rest of the image for more clarification. The U-Net model is used for mass segmentation as well as ROI for our proposed work. It is also considered fully resolution convolutional network (FrCN) to get more benefits with high resolution of features of image maps i.e., pixel to pixel mapping with segmentation. The accuracy for this segmentation is improved by reducing false positive and false negative rates, which is a challenging task. Since the infected area is not fixed and is increased by disease with shapes, sizes, locations, and boundaries, the mass segmentation is analyzed through region growing and active contour [2, 3, 4]. In the third part, classification is designed with the help of the Convolutional Neural Network (CNN) model based on (convolutional layer + ReLU + normalization and maxpolling layer) and fully connected layer for class as benign and Malignant [1,7]. Since our goal is to develop an integrated diagnosis system through mass identification, segmentation, and classification. It focuses on finding out the clarity of images using deep learning approaches explained in section III.

We also use various deep learning approaches as successful deep CNN models for computer vision classification, including InceptionV3, DenseNet121, ResNet50, VGG16, and MobileNetV2 methods for comparative evaluation results. As per comparison models, our proposed model is better performance than the above models for classification.

The contribution of this paper is summarized as follows.

(a) The integrated framework is developed for segmentation and classification in a single system.

(b) Mass segmentation and mass classification is designed using FrCN and CNN, respectively.

(c) The encoder network is designed using a convolutional layer with Rectified Linear Unit (ReLU. It also designed the decoder network using a convolutional layer with a Softmax layer.

(d) Convolutional Neural Network (CNN) based deep learning model is designed internally through (convolution + ReLU normalization and max pooling) and fully connected layer. The softmax layer is used for classification.

(e) Evaluation matrix is considered 20 equations for not only accuracy but also more performance in a different direction of evaluations

(f) For the performance comparison, the additional deep learning model has taken for segmentation with the help of the U-Net model to identify ROI (breast region) and remove the duplicates. It left the rest of the image intact for segmentation and different models, including the InceptionV3, DenseNet121, ResNet50, VGG16, and MobileNetV2 for classification using MIAS, DDSM, and CBIS-DDSM data sets.

The various sections of this paper are represented for the entire processing of the whole work. The history of the proposed system or methods with relevant research work is described in section II. Section III describes the framework of the diagnosis system with a single setting to analyze deep learning approaches for processing images in the paper. The evaluation matrix with different equations is mentioned in section IV. Section V explained experimental settings, whereas different evaluation results with analysis are described in section VI. There is a brief discussion mentioned in section VII, and section VIII concluded this paper.

2. Related work

Since this paper is related to the diagnosis model of disease in the healthcare system in a single framework setting, it needs more relevant approaches to strengthen our proposed model. Although many approaches are developed for image processing, few researchers have taken steps to create a single setting framework based on identification, segmentation, classification, which are described as following sub-sections.

2.1. *Mass detection*

The deep learning-based methodologies can provide a more comprehensive representation of mass characteristics than other conventional methods [2,6]. Breast imaging is typically used in all Computer-aided design (CAD) systems; the technique of mass detection has proven to be effective in detecting mass concentrations of cancer [2]. Manual discrimination methods were most often used in these CAD systems for image

analysis applying CNN model with classifiers. Generally, traditional mass detection was utilized in the CAD systems based on deep CNN model to classify the masses as either benign or malignant [13].

2.2. *Segmentation*

In this part, the mammogram is considered a series of separate portions and scanned across the full breadth of the breast image to locate each portion. Next, it trained an R-CNN model [8] to determine whether the regions were benign or malignant and expand it with histology of medical images. To complete the refinement of the picture search, this model was implemented in two stages, which is done sequentially using layers of deep R-CNN model [9]. Mammography has provided considerable evidence showing an increase in the mass in X-ray images, as evidenced by using the segmentation algorithm. Despite the numerous advantages of simple edge detection [15], these methods possess the primary drawback of requiring accurate topographic or statistical contour data [16]. Based on feature mapping and feature sub-setting in each block, VGG-16 feature maps designed the group feature minimization. At the same time, they can achieve size reduction, removal of duplication, and acceptable computation costs [19]. Thus, the level of detail in the resulting feature map will be proportional to the amount of subsampling and maximum pooling used in the algorithm.

The Fully Convolutional Network (or the FCN) is one of the CNN models used in most segmentation applications. A full-semantic segmentation for pixel-to-pixel semantic segmentation was put forth by [18]. It is also used with a Gaussian mixture classifier (GMC) while this algorithm performs the segmentation task. For DDSM-BCRP and INBreast, the mass-enhancing approach yielded Dice indices resulting 87% and 88% (the DDSM-BCRP [20] and INbreast [14] datasets) respectively. A second is an iterative algorithm that eliminates a sequence of unpredictable or random class assignments [16]. Finally, the Chan-Vese active contour model was also used to increase the segmentation results [11, 12].

Recently, Negi et al., [9] have used the hybrid model Wasserstein Generative Adversarial Network Residual-Dilated-Attention-Gate-UNet (WGAN-RDA-UNET) for segmenting the tumor in Breast Ultrasound images. The GAN model comprises two modules: generator and discriminator. Residual-Dilated-Attention-Gate-UNet (RDAU-NET) is used as the generator, which serves as a segmentation module, and a CNN classifier is employed as the discriminator. It has limitations of results compare to other models. Similarly, Zhuang et al., [10] had proposed a modified U-Net model named GRA U-Net to assist specialists in acceptably ascertaining a tumor in an ultrasound image for nipple segmentation. Bhuyan et al., [44] also designed for segmentation and classification to test COVID-19 disease. Further, the feature analysis of image data and its privacy can be considered [26, 39]. Singh et al., have designed a conditional Generative Adversarial Network (cGAN) which help to make segmentation of a breast tumor within a region of interest (ROI) in a mammogram. This model insight learning to identify the tumor area and form a binary mask that outlines it [48].

2.3. *Classification*

Different models on breast cancer detection, segmentation, and classification have been published [32,33,34]. Generally, classification depends on segmentation. If segmentation is adequately designed, the performance will be good for classification.

Few authors have developed breast cancer detection using deep learning approaches [28,29,31] from single or multi-images or mammography. In [30], authors have developed cloud computing-based breast cancer diagnosis using Extreme Learning Machine-based Diagnostics. Further, few authors have developed the technology based on the analysis of IoT, cybersecurity, and Blockchain [45] and determined the performance of technology as [46]. Although few authors have developed their own technology to determine the health information based on analysis of several case study like [47] with the help of wireless service prototype model. Since our aim to detect the cancer disease, the related work, we have analyzed and found limitations which are given below.

Different authors have developed their models and methods for detecting breast cancer using image processing, we found the limitation of the above model in a single setting with varying performance. For example, they considered common confusion matrices such as Sensitivity, precision, Accuracy, F1-score, etc., and lack of diagnosis components in a single setting framework. Thus, we proposed a single setting framework with all components of the diagnosis model such as mass identification, segmentation, classification, quantification, statistical evaluation, and comparison model. Thus, the quality of images improved by using these systems to enhance human judgment, and make them more legible and understandable to the computer, making them more useful for later processing.

3. Framework for Diagnosis System

3.1. *Traditional Deep Learning Architecture – Convolutional Neural Network (CNN)*

The Convolutional Neural Network (CNN) model is the most popular deep learning architecture due to its deep learning similarities to conventional neural networks. It receives an image as input instead of one extensive network (i.e., Neural Network (NN)) as shown in Fig. 1(a). The CNN comprises a series of layers: a convolutional layer, Rectified linear unit (ReLU) layer, pooling layer, or fully-connected layer in Fig. 1(b).

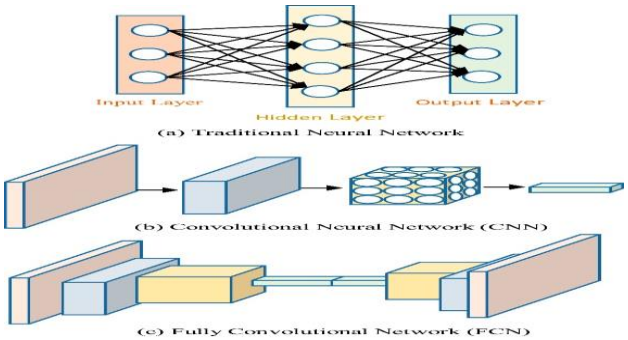


Fig. 1. Traditional deep learning architecture as [35]

The ReLU layer computes the value of the max function and reduce overfitting by reducing the spatial dimensionality of the image to an invariant scale. Because, all

neurons in the preceding layer are connected to those in the next layer. A fully connected layer is also known as the hidden layers of traditional CNN which generate small patches of equal size of input image. In CNN, the center pixel of the patch is called the 'center pixel.' Despite the inefficiency, applying such an approach is effective, as the overlapping features of the sliding patches are not re-used, causing spatial information to be lost in the image. The transposed convolutional layers are used instead of fully connected layers, as shown in Fig. 1(c). The systems and techniques are designed to classify breast cancer patients in the full functionality of a diagnosis system, including three diagnosis categories such as identification, segmentation, and classification, all in one setting. The work processing flow on infected object detection is mentioned in Fig. 2(a) and 2 (b).

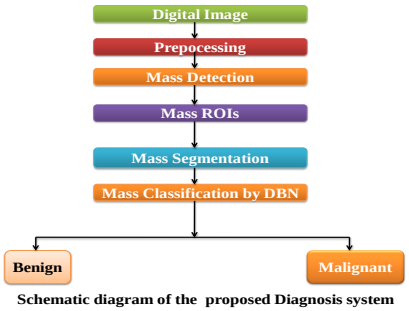


Fig. 2(a). Process of the image to classify cancer disease by proposed diagnosis system

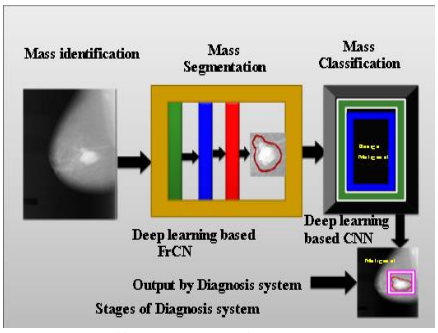


Fig. 2 (b). Different stages of proposed diagnosis system using deep learning to segment and classify breast cancer masses.

In Fig. 2(a), the flow of work is considered for the whole diagnosis system. Initially, the system is considered for taking the digital image (i.e., cancer disease image). It is pre-processed for the whole image for the suspected area of the image where breast cancer can be suspected. After getting the suspected area, the image is ready for mass detection or identification, generating the region of interests (ROIs). The remaining part, like segmentation and classification, is designed in Fig.2 (b).

In the first category, the naked eye obtains an automatic mass identification, which is used to process the image for the next step based on deep learning approaches. In the second category, an FrCN technique-based learning is proposed for mass segmentation. In the third category, we suggest an automated mass classification system that integrates different parts of the CNN model to classify the disease. The design of the proposed diagnosis scheme is shown in Fig. 4.

3.2. *Creation of Sub-Region*

The process for dividing the ROI into non-overlapping sub-regions is considered in this section. The fixed search area to find the masses is x_s, y_s, x_d, and y_d (where x_s & x_d are first and second non-zero abscissa, y_s & y_d are first and second non-zero ordinates). Here, $W = x_d - x_s$ and $H = y_d - y_s$ define the search area's length and width, especially the rectangular area for segmentation. To conduct a rectangular search using a sliding window with $W \geq w$ and $H \geq h$. For segmentation, it performs the following procedures. In the first step, it creates a rectangular pursuing ground as [36]. The method finds results within the rectangular searching area ($W \times H$). Thus, algorithm 1 is developed for ROI as follows.

……
Algorithm 1: Region of Interest identification
……
1 Input: Mammography I
2 Output: Region of interest area R
3 For k = 1 to P do
4 Create the length with first and the last nonzero pixel x_s , x_d .
5 End for
6 For k = 1 to Q do
7 Create the width with first and the last nonzero pixel y_s , y_d .
8 End for
9 Create the area for a rectangle R by the coordinates (x_s, x_d) and (y_s, y_d).
10 Return R.
……

Here, the sliding window (w × h) is transferred with a certain size, negotiating the probing area deprived of the ROI boundary trip. Thus, the ROI generates overlapping sub-regions with the same size (w × h), which assists in selecting subsequent features. For the sub-region, we have taken the fixed size of the sliding window as 48 × 48, and the searching step size is 48. At last, the N non-overlapping sub-regions (s_1,s_2, · · · ,s_N) [36] are divided by ROI in this paper.

3.3. *Data enhancement and transfer learning*

Data augmentation is an increasingly common approach proposed for dealing with this model, where pre-processed data is transferred to a new system. We have rotated the initial mammograms eight times by angle and manipulated their shape using different angles. That yields 880 mammograms from the 110 original mammograms considered all augmented mammograms, with 280 benign and 600 malignant cases used for experiments. There are two different ways to initialize deep learning models: a method known as random initialization and a process known as transfer learning [11,20,21]. Instead of using raw, unlabeled images to fine-tune deep models such as ImageNet [22]. The models are then retrained fine-tuned with an augmented image dataset, such as mammograms [5,6,11,21]. The best use of transfer learning for image processing in the second approach is treating breast cancer diagnosis using diagnosis systems [5,6,11,13]. Based on the sub-region and enhancement of the image, the image is processed for segmentation. Thus, the infected area of the breast image is explained in the next section.

3.4. *FrCN for the mass segmentation*

If the mass is expanded in the mass of infection, the area of infection will be highly dense. Once an image has been centered over a contaminated region, the naked eye can see the image. Then, it is sent to the next stage (mass segmentation) or on its way, which is briefly referred to as expanded. In general, segmentation is a separate technique done to separate a region of interest (ROI) from the context of an image. There are many methods for improving contrast in the image and elevating the masses, also allowing their homogeneity to be more easily distinguishable [9,10,11]. Much initial training has revealed the SegNet and U-Net for pixel-to-pixel segmentation, resulting in higher segmentation accuracy comparing to existing approaches [17]. With this resolution reduction in features, the extent of the feature maps is reduced. On the other hand,

segmentation models using multiple max-pooling and subsampling layers in their encoder networks are produced in these models with their decoder networks. Instead of reducing the feature map sizes in the decoder models, the "up-sampling" and "deconvolution" layers are applied to extend the number of training parameters. The expanded FrCN deep learning model is suggested in this paper, based on this mass segmentation method. Because of this, the encoder's low-level deeper function maps using convolution layers, and the encoder's full resolution is kept intact, as fig. 3 with the formation of a pattern appears in the form FrCN. The proposed deep learning segmentation-based FrCN model is designed to handle all the mathematical processing and needed pixel-to-pixel segmentation. The results in a network that better represents reality [13]. We can handle the segmentation task with full resolution and a competitive computational time with this complete model.

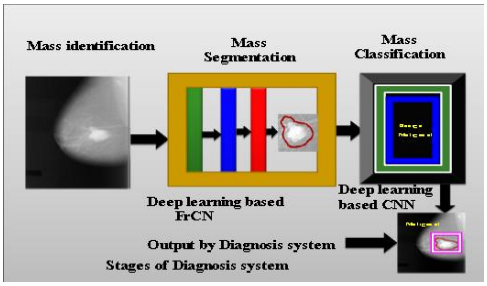


Fig. 3. A full resolution convolutional network (FrCN) model for the mass segmentation stage. (a) Detected ROI on the original mammogram with its ground truth, (b) detected ROI (i.e., input mass) with highlighted ground truth (red), (c) output segmented map of input mass, and (d) segmented output mass.

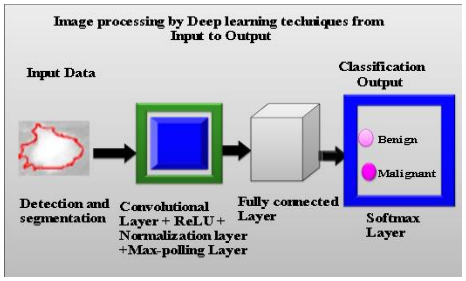


Fig. 4. Convolutional Neural Network (CNN) based deep learning model for the mass classification stage.

3.5. *U-Net Model*

The U-Net model is the most widely applied image segmentation model. We considered a U-Net model suggested as [40,41,42], which was applied to biomedical image segmentation. A wavelet-based process was applied to enhance the frequency of spatial pictures, and the area of interest (ROI) was manually extracted using this technique. Using the segmentation technique, S. Duraisamy et al. [42] extracted the contour of the calcify area from the image. Thus, we considered the modified segmentation of the U-Net model applied to the Breast ROC analysis, where it removed the unwanted regions.

3.6. *Segmentation networks*

This is critical for making the diagnosis of breast cancer early. Segmentation could be seen as a classification task where one classifies each pixel in an image dataset into two separate groups: those with a specified either ROI (breast region) or those without background. Using the modified U-Net model, our mammography images are segmented based on the U-Net approach. U-Net helps to create modified segmentation based on the encoder and decoder networks. This lowermost layer of the U-Net network processes the decoder information. In the decoder section, features at a higher resolution are extracted from the encoder section. After this, a skip occurs, and delicate segment structures are

generated. Instead of rectified linear unit (ReLU) and standardization of batch, leaky ReLU and normalization of the instance are used.

3.7. *CNN for mass classification*

We use a classification step that depends entirely on others, such as segmentation and feature extraction because classification is challenging. Typically, a modern CNN has one or more convolutional layers based on stacked inputs, processed by one or more fully connected (FC) layers to perform well classification. Solving the problem of mass classification, AlexNet (i.e., ConvNet [21]) is used to simplify the classification of the segmented masses. In this model, Deep feature maps are formulated with the addition of three subsequent convolution layers based on convolution + ReLU+ normalization and max-pooling; then, Deep feature reduction is modeled and implemented with max-pooling followed by ReLU. Two layers are taken for fully connected (FC) layers, as shown in Fig. 4.

In deep feature maps, first and second convolution layers, there are (combine) 20 and 64 filters with a linear scale of 5×5 and (combined) second-stage length of 32. The third, fourth, fifth, and sixth convolution layers, also known as third- and fourth-order filters, use 512 filter scales of 256; the sixth uses 1024 filter scales of 1024. As illustrated in Fig. 4, a non-overlapping max-pooling involves a size of 2×2 and is used to expand the sub-sample by a factor 2 from the original input patch size. This model is also designed through [27]. It is less costly and would only be required if the number of functional systems is globally expanded. Then, the second set of 1024 and 4096 nodes is built from the two FC layers. This step applies a probabilistic decision logic to the logistic regression layer, and two Softmax nodes are then used to differentiate between benign and malignant diagnoses. Except for the last layer, the activation function of ReLU (Rectified Linear Unit) is used. Rather than using ordinary sigmoid and tanh activation functions, ReLU activation is used in deep models since it has a high value of the ReLU's activation function. It would take a longer to achieve with excellent training [21].

3.8. *CNN based convolution layer*

In-depth features are extracted from the ROI sub-regions using CNNs in this research as [36]. Figure 4 illustrates the design of a CNN network built from 7 successive layers, with three convolutional layers, three max-pooling layers, and one fully connected layer. CNN's input is a 48 × 48 sub-region image that was previously captured from other stages. First, the 48×48×3 input images are filtered with 12 kernels of size 9×9×3. Next, the 12 resulting images are down sampled to size 40×40×12. The form of a convolution is defined as Eq. (1).

$$Conv^{k}(i,j) = \sum_{u,v} W^{k,l}(u,v).input^{i}(i-u,j-v) + b^{k,l} \qquad (1)$$

Where $W^{k,l}$ contains k^{th} kernel and $b^{k,l}$ represents the bias of k^{th} layer, and (u,v) in length and width. It considered the tanh that is used to define the activation value's range [−1, 1]. Thus, the output function can be defined as Eq. (2).

$$Output^{k}(i,j) = \tanh(Conv^{k}(i,j)) \qquad (2)$$

A max-pooling layer is connected to the first convolution layer's output. For obtaining the output with size 2 × 2 × 6, the next and the third convolution/max-pooling layers are connected. For this clustering analysis, the fully-connected layer has 2 × 2 × 6 = 24 neurons, which are the features.

3.9. *Additional Deep learning models*

Various methods are considered from [38] in this paper for comparative performance. We have taken the most successful deep CNN classification methods in the computer vision field, including InceptionV3, DenseNet121, ResNet50, VGG16, and MobileNetV2 models. It must include Certain parameters on the mammography dataset to begin the fine-tuning process as following methods.

(i) InceptionV3: The ImageNet is used to train the InceptionV3 model. With this design, the computing ability of the network is increased. The rate is increased to $10^{-4,}$ and the number of iterations is also increased to 10^6 in our proposed work.

(ii)DenseNet: Dense CNN (DenseNet121) is a dense feed forward neural network with full connectivity. Every layer in DenseNet121 has a different function map assigned. The next layer is fed from the previous layer's character map. Also, with DenseNet121, fewer parameters are required. Also, the number of iterations and the iteration rate are increased to 10^5 and reduced to 10^{-2}. The epoch number is modified to 80.

(iii)ResNet-50: This is a pre-trained model that uses the ImageNet. The ResNet50 skips a layer or uses more complex routing to handle gradient vanishing difficulty. The most significant benefit of ResNet50 is its ability to be optimized quickly. Instead of the traditional gradient descent, the ResNet50 model uses ReLu, a different kind of nonlinear activation function. Forward and backward propagation methods are utilized with the ResNet50. The iteration count and the iteration rate are set to 10^4 and 10^{-3}, and the epoch count is 130.

(iv)VGG16: The AlexNet has its architecture updated to VGG16, which has an additional layer. When more layers are included in the model, the generalization is proportional to the number of layers. While the benefits of VGG16 are in the use of only 3 x 3 convolutional filters, the drawback is that it can have difficulty handling more complex structures. For re-training the VGG16 model, 10^5 iterations and 10^{-4} are required, while the number of epochs is 80.

(v)MobileNetV2: The MobileNetV2 has two blocks: The residual block has a stride of 1, while the downsizing block has a stride of 2. Each block has three layers. ReLU is applied to the first layer, and depth-wise convolution is used for the second layer. Next, a 3rd-layer convolution with a single 1 x 1 nonlinearity. Each iteration is set to a value of 10^7 with a maximum number of iterations, and a rate of 10^{-5}, with a minimum number of the epochs 160.

4. EVALUATION MATRIX

When deep learning applies to image processing, precisely the solution for the image classification problem, a confusion matrix is considered to obtain the accuracy of the object from images or videos. Moreover, it is assumed to be a two-dimensional contingency table based on "actual" and "predicted" and its variables. An expanded confusion matrix is created by defining terms and tallying their values in two columns: predictors that identify real issue causes, where the issue is detected as real and false. The

signal identifies the real issue and those that are misdiagnosed as false positives. The classification figures give the comparison an idea of how they compare to the class and expand on a few measurements, such as calculating correct scores on standard tests (accuracy). The above approach to validation testing is done by applying various equations to actual parameters, such as TP, TN, FP and FN.

In our model, the proposed diagnosis method is processed by two stages with precise analysis, starting with segmentation assessment. To determine the overall harmonic mean, we have used an unbalanced dataset, which means we used a dataset with data points (because of our emphasis on precision/sensitivity for Dice similarity coefficient or F1-score) to work with Matthew's correlation coefficient (MCC). For expanding on the preceding statement, it is stated that a score of 1 indicates the best precision and sensitivity; otherwise, 0 indicates worst measurements. The sensitivity, precision, overall accuracy, and Dice coefficient are all employed in the test of the proposed segmentation system to see if the number of graphs belongs to the different folds. The parameters for all of these metrics are defined as follows, according to the following description. The following equations for the evaluation matrix is considered as [37]. Although many authors are developed their model using few equations from the following equations. Still, we tried to use all the following equations for the proposed model to make more precise and well performance. The reader can refer to [37] for more clarification about the following equations.

Sensitivity/Recall/True Positive Rate (TPR)

$$= \frac{TP}{P} = \frac{TP}{(TP+FN)} = \text{1-FNR} \quad (3)$$

Specificity/Selectivity/True Negative Rate (TNR)

$$= \frac{TN}{N} = \frac{TN}{(TN+FP)} = \text{1-FPR} \quad (4)$$

Precision/Positive Predictive Value (PPV)

$$= \frac{TP}{(TP+FP)} = \text{1-FDR} \quad (5)$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{(TN+FN)} = \text{1-FOR} \quad (6)$$

Miss Rate/False Negative Rate (FNR)

$$= \frac{FN}{P} = \frac{FN}{(FN+TP)} = \text{1-TPR} \quad (7)$$

Fall-out/False Positive Rate (FPR)

$$= \frac{FP}{N} = \frac{FP}{(FP+TN)} = \text{1-TNR} \quad (8)$$

$$\text{False Discovery Rate (FDR)} = \frac{FP}{(FP+TP)} = \text{1-PPV} \quad (9)$$

$$\text{False Omission Rate (FOR)} = \frac{FN}{(FN+TN)} = \text{1 –NPV} \quad (10)$$

$$\text{Prevalence Threshold (PT)} = \frac{\sqrt{(TPR(-TNR+1)}+TNR-1}{((TPR+TNR-1))} \quad (11)$$

Threat score (TS)/Critical Success Index (CSI)/TS

$$= \frac{TP}{(TP+FN+FP)} \quad (12)$$

$$\text{Accuracy (ACC)} = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

Balanced Accuracy (BA)= $\frac{TPR+TNR}{2}$ (14)

F1 Score (Dice) = $2*\frac{PPV*TPR}{PPV+TPR} = \frac{2TP}{2TP+FP+FN}$ (15)

Matthews correlation coefficient (MCC)

$$= \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (16)$$

Fowlkess-Mallows index (FM)

$$= \sqrt{\frac{TP}{TP+FP}} * \frac{TP}{TP+FN} = \sqrt{PPV * TPR} \quad (17)$$

Informedness or Bookmarker informedness (BM)

= TPR +TNR – 1 (18)

Markedness (MK)/ Delta P=MK = PPV +NPV – 1 (19)

Positive likelihood ratio (LR+) = $\frac{TPR}{FPR}$ (20)

Negative likelihood ratio (LR-) = $\frac{FNR}{TNR}$ (21)

Diagnostic odds ratio (DOR) = $\frac{LR+}{LR-}$ (22)

IoU = $\frac{|y \cap y'|}{|y \cup y'|}$ (23)

DC = $\frac{2.|y \cap y'|}{|y|+|y'|}$ (24)

Each equation is provided by TP, TN, FP, and FN, which are represented per pixel to represent as their quantity. For deriving all of these parameters, the confusion matrix is used. Building a perfect segmentation, we must ensure high sensitivity and precision. We must, therefore, accurately detect all the masses and tissues in the surrounding region. Ground-truth regions are used as an approximation of their corresponding predicted regions. The Dice (F1-Score) and Jaccard (number of FP pixels in expected regions) are considered the main constituents to determine ground truth regions for the test. In the event of problems with the ground-truth data, the MCC is a valuable method for determining how well the segmented mass pixels correlate with it. Another benefit of providing ROC curves with AUC and a tradeoff between sensitivity and precision is to evaluate the segmentation process. Segmentation training expands its taxonomy by increasing image classification, which uses a new range of classification, or sensitivity, by boosting the quality curve with AUC, by developing ROC and F1-scores rather than developing a single-pixel ROC curve for each segment [1,2,5,11].

5. EXPERIMENTAL SETTINGS

5.1. *Dataset*

For experiment, the proposed methods for diagnosis is applied on the breast dataset's from INBreast database [14], particularly for accuracy assessments in addition to the methodology and simulations. It contains 400 radiograms/mammograms (all together, benign and malignant) and images of both the MLO and the imaging screens or CC of 110 women patients as malignant and non-malignant mammograms. Thus, for experimentation, two kinds of data sets are selected: 96 for test and 400 for training. The

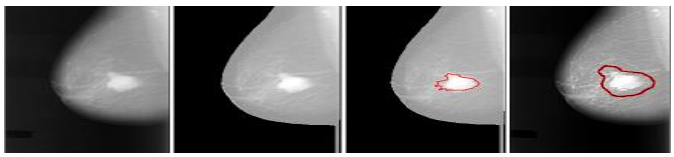
performance is done on the mass from the mammograms. Another on the sum of the masses of the scanning results using the data, there were 110 in total using Breast Imaging Reporting, and Data System (BI-RADS) provided by the loses disease which illustrates the efficiency of the system (BI-RADS). Further, we considered the number of breast images included in the MIAS, DDSM, and CBIS-DDSM datasets and utilize the approaches suggested in [44], which help in comparison evaluation as per the proposed model. We considered 496, 534, 302, and 300 for INBreast, DDSM, MIAS, and CBIS-DDSM datasets, respectively for our experiments. In all experiments, we considered 400 for testing data set and 96 for training data set.

5.2. *Hardware and software supported tools*

All tests are conducted on a Personal Computer (PC) with software and hardware specifications such as (a) An Intel Core i7 CPU with 16 GB of RAM, (b) Python 3.0.7 on a Ubuntu 16.04 OS for implementation of diagnosis system. The evaluation of deep segmentation was carried out using Theano [23] and Keras [24] deep learning libraries, the model that integrates different types inputs or multiple inputs is generated, while the models for identifying and classifying input data are implemented with Tensorflow [25].

5.3. *Experimental settings for identification of breast cancer image*

A step-by-step analysis of the system efficiency uses the INBreast database [18]. This experiment considers 4-fold cross-validation to conduct multiple tests for each stage to ensure each mammogram gets checked in the same way, with the training, validation, and test datasets created by stratified partitioning to ensure no bias error. In all experiments, two kinds of data sets are considered for both benign and malignant as 400 for testing data set and 96 for training data set.



(a) (b) (c) (d)

Fig. 5. Mass detection using YOLO technique on the test images of INbreast dataset. (a) and (b) show the detected ROIs (i.e., masses) for benign cases, while (c), and (d) for malignant cases. Detected ROI is superimposed on the original images: benign, malignant (red), and ground truth (red).

Sometimes, an unbalanced training dataset creates bias errors for all segmentation and classification deep learning models during the training phase. So, it utilizes the following conditions. First, the training set is shuffled through a mini-batch to ensure each image is used only once as defined in [18]. Another noteworthy aspect of these studies is that they utilized a loss function (weighted cross-entropy) to determine the parameters when a deep model is under training. The above data illustrates that the two views of mammograms (i.e., MLO and CC) in the INbreaset dataset are found with dissimilar image sizes [14]. In addition, all images are resized to 256×256. Once all images have been normalized to the range of [0,1], they all get the same image property. In this analysis, we determine image having any suspicious infection or not. The YOLO techniques of the image is detected as shown in Fig.5. If any subspinous area is found, then the image is processed for testing through segmentation, which is explained in the next section.

5.4. *Experimental settings for mass segmentation via FrCN*

When it determines the object's mass with a segmentation process, only those masses have been correctly measured for use. At the same time, all those that have been wrongly reported are now gone. Training all these complicated neural network models with a learning rate acceleration strategy allows all of them to be trained correctly. This will continue concurrently with the parameter expansion process. Several iterations of mini-batches are also being applied to pick out the models' parameters using training and validation datasets through a cross-entropy loss function. When it's possible, dropped down after the first convolution layer and the Dropout feature is activated, then images appear to have 0.5 less dense (or be expanded) in Fig. 3

5.5. *Experimental settings for mass classifications via CNN*

Different segmented masses are taken to resize to 50×50 by using a bicubic interpolation [9,10,11,13,16]. The masses are all sent to the CNN in the final image, each ending up in the classification stage. For a simple comparison, we use the segmentation without the diagnosis systems, and then for the real-world use, we employ the medical diagnosis systems. Going through a CNN is especially important in the latter, which causes the masses to be automatically classified directly and sent to the subsequent lower-expanded segments to reduce bias. In the above case, the mass fragments are divided into two halves and then placed into the CNN's classification stages. The classification of the framework is consistent, ranging from the parameters to the design to the classifier itself. In a batch training system, the weight decay regulates the learning rate in proportion to the amount of data that has been passed, with the weight decay for batches of data passing being 0.5. However, the number of epochs is set to 100, the same as the scale of the mini-batches are 100 epochs. A 0.4% undetected dropout rate on fully connected layers is both useful for accelerating training but also prevents overfitting [9,18].

6. Results And Its Analysis

The above setting evaluates the considered data for mass segmentation, mass classification, quantifying evaluation, and comparing results as follows. The classification and segmentation accuracy of breast cancer images are explained based on end-to-end training models experimented in this paper. Based on the validation matrix, different equations are considered for the above evaluation. Initially, it considered Table 1 to obtain true and false quantities of benign and malignant based on four-fold cross-validation.

Table 1. Quantity of true and false of benign and Malignant over 4-fold cross-validation via deep learning approaches of the breast caner dataset.

Fold Test	Benign		Malignant		Total		Accuracy
	True	False	True	False	True	False	
1st fold	163	2	233	2	396	4	
	98.78%	1.22%	99.15%	0.85%	99.0%	1.00%	99.0
2nd fold	163	2	234	1	397	3	
	98.80	1.2	99.60	0.4	99.25	0.75	99.2
3rd fold	164	1	233	2	397	3	
	99.4	0.6	99.15%	0.85%	99.25	0.75	99.2
4th fold	164	1	232	3	396	4	

99.4	0.6	98.72	1.28	99.0	1.0	99.0	

6.1 Mass segmentation results

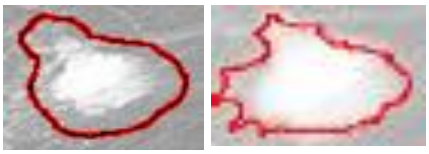
During the segmentation, falsely identified breast cancer masses are removed from each test fold in the segmentation process. The several outputs of mass segmentation of the proposed FrCN are specified in Table 2. The findings of these testing are quantified from the degree of analysis on the same set of mass. The complete mass segmentation process is performed at a single resolution of the original image. All measurements are taken for each pixel in the segmented maps (i.e., input ROI).

To put it another way, Table 2 shows that FrCN performs with an average 98.7% Dice index, 97.8% TS/CSI coefficient, 99.1% overall accuracy, and 98.15% MCC.

Table 2. FrCN based segmentation on 4-fold test data set

1st Fold	TPR	TNR	PPV	NPV	FNR	FPR	FDR	FOR	PT	TS/CSI
	0.988	0.991	0.988	0.991	0.012	0.008	0.012	0.012	0.98	0.976
	ACC	BA	F1	MCC	FM	BM	MK	LR+	LR-	DOR
	0.99	0.99	0.988	0.979	0.988	0.979	1.979	116.08	0.01	9515.1
2nd Fold	TPR	TNR	PPV	NPV	FNR	FPR	FDR	FOR	PT	TS/CSI
	0.993	0.991	0.987	0.995	0.006	0.008	0.012	0.004	0.09	0.981
	ACC	BA	F1	MCC	FM	BM	MK	LR+	LR-	DOR
	0.992	0.992	0.990	0.984	0.989	0.984	0.982	124.12	0.01	20687.5
3rd Fold	TPR	TNR	PPV	NPV	FNR	FPR	FDR	FOR	PT	TS/CSI
	0.987	0.995	0.993	0.991	0.012	0.004	0.006	0.008	0.06	0.982
	ACC	BA	F1	MCC	FM	BM	MK	LR+	LR-	DOR
	0.992	0.991	0.990	0.984	0.989	0.982	0.984	246.75	0.01	20562.5
4th Fold	TPR	TNR	PPV	NPV	FNR	FPR	FDR	FOR	PT	TS/CSI
	0.982	0.995	0.993	0.987	0.017	0.004	0.006	0.012	0.06	0.976
	ACC	BA	F1	MCC	FM	BM	MK	LR+	LR-	DOR
	0.99	0.988	0.98	0.979	0.987	0.977	0.98	245.5	0.017	14441.17

Furthermore, FrCN is better in 2nd and 3rd fold compare to other fold demonstrated by the ACC value obtained when all of the test samples were included.



(a) (b)

Fig. 6. (a) fully convolutional network (FCN) (b) full resolution convolutional network (FrCN) of the segmentation performance with contours-based ground truth (red) with FrCN.

The ACC of FrCN is 99.2 percent in 2nd and 3rdfold, which means that the efficiency of FrCN is 0.2 higher than the lower fold. The accuracy is better if the quantity of false is less in Benign. Based on this complement, the Dice index has considered mass accuracy for each image which is compared shown in Fig 6. In Fig 6, the contour-based ground truth is taken to identify the infected area by cancer. Here, it is easily identifying the difference between FCN and FrCN by ground truth of image or region of interest (ROI).

6.2 Mass classification results

As defined in the previous segmentation stages, all segmented masses reach the classification stage sequentially, one mass after another. In terms of different output, classification evaluation includes various metrics such as TPR, TNR, overall accuracy, F1-Score, MCC, etc. At the same time, a diagnosis with segmentation in the test data set has higher accuracy, with TPR of 98.8%, TNR of 99.1%, the overall accuracy of 99.0%, their MCC and F1-score by 97.9% and 98.8%, respectively from Table 2. Meanwhile, the performance of mass identification in table 2 represents the product of test folds performed with the diagnosis systems (i.e., the systems in the mentioned scenario), including segmentation. Thus, 98.78% of benign and 99.15% of malignant cases are correctly categorized, while 1.22% of benign and 0.85% of malignant cases are negatively classified in the first fold of Table 1. Similarly, other information is mentioned in Table 1.

From Table 2, several experimental outputs are obtained based on four-fold tests with several true and false related evaluations as per different equations (i.e., Eq. 3-24) mentioned in section IV. Thus, several performances are also shown in Fig. 7 differently based on four-fold testing data sets.

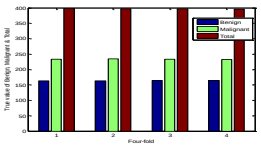


Fig.7 (a). True value of Benign, Malignant and Total (in quantity in number) vs Four fold

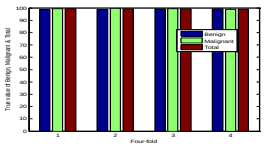


Fig 7(b). True value of Benign, Malignant and Total (in %) vs Four fold

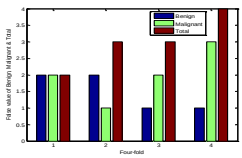


Fig. 7(c). False value of Benign, Malignant and Total (in quantity in number) vs Four-fold

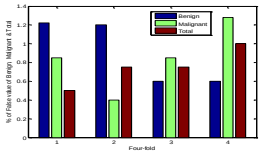


Fig. 7(d). False value of Benign, Malignant and Total (in quantity in number) vs Four-fold

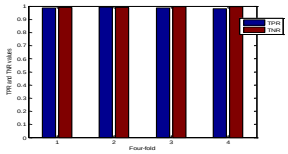


Fig. 7(e). TPR and TNR values on four-fold testing

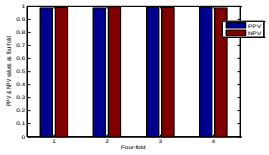


Fig. 7(f). PPV and NPV values on four-fold testing

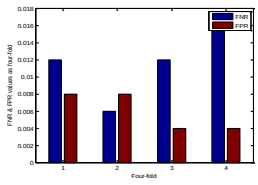


Fig. 7(g). FNR and FPR values as four-fold testing

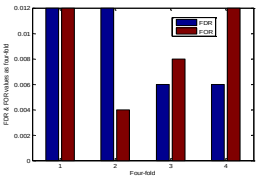


Fig. 7(h). FOR and FDR values as four-fold testing

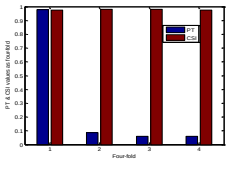


Fig. 7(i). PT and CSI values as four fold testing

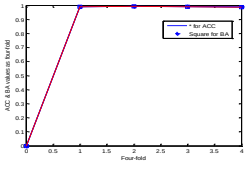


Fig. 7(j): ACC and BA values as four-fold testing

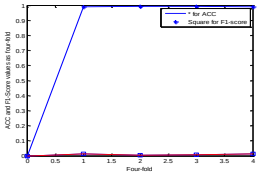


Fig. 7(k): ACC and F1 Score values as four-fold testing

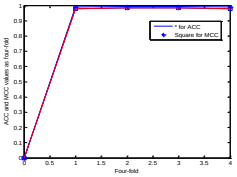


Fig. 7(l): ACC and MCC values as four-fold testing

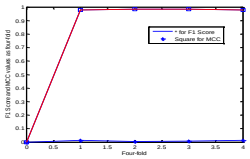


Fig. 7(m): F1-Score and MCC values as four-fold testing

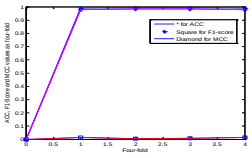


Fig. 7(n): ACC, F1-Score and MCC values as four-fold testing

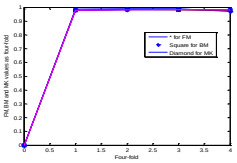


Fig. 7(0): FM, BM and MK values as four-fold testing

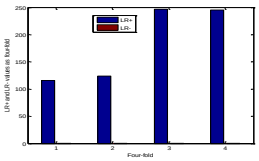


Fig. 7(p): LR+ and LR- values as four-fold testing

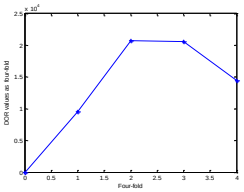


Fig. 7(q): DOR values as four-fold testing

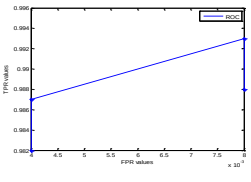


Fig. 7(r). ROC makes on FPR vs TPR

The several parts of Fig. 7 are explained as follows with different observations. The quantitative measurement of true and false values of benign and malignant is shown in Fig. 7 (a)-(d). At the same time, the outstanding performance is determined among relative evaluation or pair of evaluation shown in Fig. 7(c)-7(r) such as (TPR, TNR), (PPV, NPV), (FNR, FPR), (FOR, FDR), (PT, CSI), (ACC, BA), (ACC, F 1-Score), (ACC, MCC), (F 1-Score, MCC), (ACC, F 1-Score, MCC), (FM, BM, MK), (LR+, LR-), DOR, ROC. From this pair of evaluations, it is observed that most of these pairs are close to each other as their performance, whereas few are different due to false values are varied.

If false data is more available in the training data set, the segmentation diagnosis system is adversely impacted by the negative predictive value. As a result, false positive rates (the proportion of inaccurate diagnoses) will be increased as per several classifiers. So, in the long run, with the diagnosis system's segmentation, all evaluations as per the confusion matrix provided well outputs in four-fold cross-validations as in Tables 1 and 2 on the test data set. When reviewing the classification results, it was found that the proposed diagnosis method was better in the tested data set to compare to the training data set due to maximizing the number of false positives and negatives in the training data set as in Table 3.

Table 3. True, false, positive, and negative as per classifiers on the training data set

Classifiers	Benign		Malignant		Total	
	True	False	True	False	True	False
REPTree	26	23	22	25	48	48
Random Tree	26	23	27	20	53	43
Random Forest	27	22	23	24	50	46
J48	22	27	18	29	40	56
Hoeffding Tree	42	7	35	12	77	19
Decision Stump	20	29	10	37	30	66

As per the training data set from Table 3, different classifiers are evaluated and got several outputs accordingly, which are shown in the following figures (i.e., Fig 8(a)-8(g)).

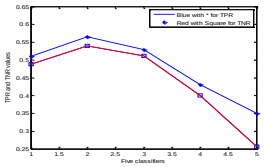


Fig 8(a). TPR and TNR values on five classifiers

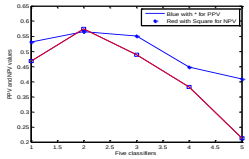


Fig 8(b) . TPR and TNR values on five classifiers

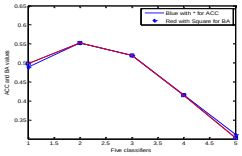


Fig 8(c). ACC and BA values on five classifiers

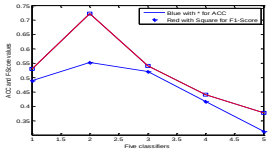


Fig 8(d) : ACC and F1-Score values on five classifiers

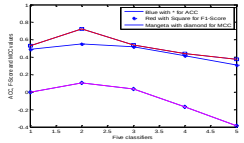


Fig 8(e). ACC, F1-Score and MCC values on five classifiers

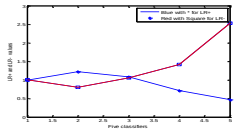


Fig 8(f). LR+ and LR- values on five classifiers

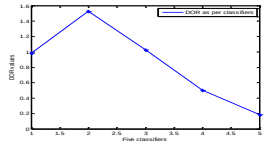


Fig 8(g). DOR values on five classifiers

Compare to training data set with test data, test data is obtained sound results. Although evaluation values are mentioned in several parts of Fig.7 and Fig. 8 as per the tested data set and training data set, it considered only less pair of comparison from fig 8 than Fig 7. Because, maximum false values are available in the training data set to compare to the test data set.

6.3 Quantified results on training data set

Although we have explained different parts of the diagnosis system on both tested and training data sets, the quantification results are still needed for further clarity. Only training data sets are considered for quantification performance since it is less performance than tested data set. The statistical evaluation is considered for quantification performance which is mentioned in table 4.

Table 4: Statistical Evaluation as per classifiers

	REPTree	RandomTree	RandomForest	J48	HoeffdingTree	DecisionStump
CCI	53.125 %	47.9167 %	53.125 %	53.125 %	56.25 %	59.375 %
ICI	46.875 %	52.0833 %	46.875 %	46.875 %	43.75 %	40.625 %
MAR	0.4913	0.5208	0.492	0.4879	0.4609	0.4693
RMSE	0.527	0.7217	0.5227	0.5403	0.5657	0.5021
CC (0.95 level)	97.9167 %	47.9167	100 %	92.7083 %	95.8333 %	98.9583 %
MRRS(0.95 level)	98.9583 %	50 %	100 %	92.1875 %	92.7083 %	99.4792 %

In table 4, the abbreviation as Correctly Classified Instances – CCI, Incorrectly Classified Instances (ICI), Mean absolute error (MAE), Root mean squared error (RMSE), Coverage of cases (CC), Mean rel. region size (MRRS)

From Table 4, it is mentioned correct and incorrect classified instances, several errors, coverage cases and mean region size, which are evaluated by different classifiers. Distinguished errors among classifiers are mentioned in Fig. 9. The error range from different classifiers are also shown in Fig. 9 based on mean absolute error values, and root mean square error values, where mean absolute error performs better than root mean square error.

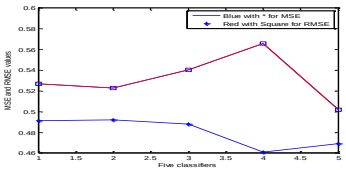


Fig 9: MSE and RMSE values on five classifiers

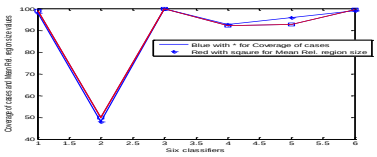


Fig 10: Coverage of cases and Mean rel. region size values on six classifiers

In Fig. 10, it is considered only coverage of cases and mean relative region size due to coverage of cases and mean relative region size are close to each other because they are very sensitive to each other. Thus, coverage of cases and the mean relative region size of the training data set is very effective compared to other quantified evaluations. This quantification performance is also identified the good accuracy of the infected area of breast cancer disease. The coverage of cases and mean relative region size values are so effective that the infected area's contour can be easily identified.

6.4 Comparison performance as additional Deep Learning with proposed model

Based on the end-to-end training models implemented in this paper, screening mammograms have been classified and segmented with a high degree of accuracy. The additional number of breast images are included in the MIAS, DDSM, and CBIS-DDSM databases, which are assessed and the usefulness of the techniques suggested in [43]. Our newly used databases include Cranio Caudal (CC) vision and Mediolateral Oblique (MLO) images of two breasts for each patient. One of the two categories, either Benign or Malignant describes the cluster of cases. Both MLO and CC views of each breast are included in the cases of 534, 302, and 300 for DDSM, MIAS, and CBIS-DDSM datasets, respectively. Therefore, the DDSM and MIAS augmented images are respectively a 2136

and a 1208 image. MIAS, DDSM, and CBIS-DDSM databases include certain non-tumor mammograms, so classification criteria are based on the availability of those mammograms. The Dice Coefficients (DC), accuracy, sensitivity/recall, precision, F1-Score, Area under ROC Curve (AUC), and computational time are our evaluation metrics under the ROC curve. We considered our proposed same equations for the above datasets for comparative evaluations with our proposed model. An accurate prediction gives an accurate result as in Eq. (13). Precision is the percentage of positive confirmed malignant cases corresponding to the ground truth as stated in Eq. (5). Sensitivity is the true positive rate, also known as the positive predictive value, or the fraction of true malignant cases that are found to be malignant, as shown in Eq. (3). F1 score is the harmonic mean of precision and sensitivity, and it incorporates these characteristics in a more generalized manner. The equation illustrates how the sample sets are similar as well as diverse (15). AUC analyses the entire two-dimensional region under the entire ROC curve and calculates an overall output over all possible classification threshold settings. In this instance, all these measurements are specified as Eq. 3 to Eq. 24.

Our proposed model and successive deep CNN methods for classification in the image processing domain include different methods such as InceptionV3, DenseNet121, ResNet50, VGG16, and MobileNetV2 for comparative evaluation that is shown in Table 5 and 6.

Table 5. Performance Comparison between the proposed model and additional Deep Learning model for segmentation with classification Results

Model	Accuracy %	AUC %	Sensitivity %	Precision %	F1-Score %
DDSM Database					
InceptionV3	96.45	96.87	96.96	96.86	95.88
DenseNet121	95.89	95.44	95.67	95.36	95.63
ResNet50	94.87	94.24	95.98	95.69	94.49
VGG16	93.87	93.65	93.88	93.88	93.79
MobileNetV2	92.52	92.87	92.98	91.99	92.87
Proposed FrCN	**99.0**	**97.87**	**98.80**	**98.80**	**98.80**
MIAS Database					
InceptionV3	94.32	94.59	94.88	94.32	94.52
DenseNet121	92.89	92.88	92.99	92.98	92.88
ResNet50	91.57	91.89	91.99	91.51	91.08
VGG16	90.54	90.95	90.88	90.88	90.99
MobileNetV2	89.99	88.89	89.98	89.87	89.87
Proposed FrCN	**99.20**	**97.54**	**99.30**	**98.70**	**99.21**
CBIS-DDSM Database					
InceptionV3	93.21	93.59	93.32	93.63	93.29
DenseNet121	92.96	92.64	91.89	92.21	91.88
ResNet50	91.27	91.81	91.22	91.32	90.99
VGG16	89.96	89.89	89.98	89.98	89.89
MobileNetV2	88.95	88.59	87.79	88.98	87.99
Proposed FrCN	**99.20**	**95.,87**	**98.70**	**99.30**	**99.45**

Table 6. Performance Comparison between proposed model and additional Deep Learning model for classification Results

Model	Accuracy%	AUC %	Sensitivity %	Precision %	F1-Score %
DDSM Database					
InceptionV3	88.87	87.99	88.74	88.59	87.99
DenseNet121	85.97	84.99	85.89	85.84	85.85
ResNet50	84.54	83.99	83.94	84.12	84.21
VGG16	82.81	81.99	82.74	82.43	82.34
MobileNetV2	80.97	80.89	81.74	80.88	80.99
Proposed CNN model	**99.12**	**98.87**	**99.89**	**99.85**	**98.89**

MIAS Database					
InceptionV3	86.77	85.89	86.84	85.48	85.88
DenseNet121	83.87	82.88	83.77	83.74	83.75
ResNet50	82.65	81.79	81.96	82.46	82.65
VGG16	80.98	80.87	80.84	80.83	80.84
MobileNetV2	79.97	79.54	79.84	80.21	80.11
Proposed CNN model	**99.54**	**98.85**	**99.82**	**98.99**	**98.89**
CBIS-DDSM Database					
InceptionV3	84.21	84.87	84.87	83.99	83.96
DenseNet121	82.47	82.65	82.57	81.99	82.34
ResNet50	81.65	80.89	80.99	81.32	81.22
VGG16	80.98	80.77	81.89	81.99	81.84
MobileNetV2	79.82	78.99	79.87	79.99	79.89
Proposed CNN model	**99.89**	**98.89**	**99.91**	**99.96**	**98.86**

Here, the false positive (FP) is the malignant non-lesion pixel segmented as a lesion pixel with the incorrect diagnosis. The false negative (FN) is the benign lesion pixel segmented as a non-lesion pixel. The TP in this case, is the malignant database sample that corresponds to the correct diagnosis of cancer. At the same time, the TN is the benign database sample corresponding to the correct cancer diagnosis.

Table 7. Comparative performance of IoU and DC in %

Quantitative comparison between two data sets for classification				
References	No. of mammograms databases	Name of Dataset	DC %	IoU %
[38]	1804	DDSM	91.89	92.99
Proposed Model	400	INBreast	98.54	99.12
Performance Comparison using MLO and CC view				
[38]	1804	DDSM	94.79	94.89
Proposed Model	400	INBreast	98.99	99.28

We also use the Intersect over Union (IoU) to quantify the percentage overlap between the target mask and our prediction of whether the patient will be malignant or benign with dignity as in Table 7. When IoU increases, system performance also improves. The ground truth mask is denoted by y, and the generated probability map is given by y´ using neural network. To conclude, DC is also considered as the loss function.

7. Discussion

This section discussed the proposed integrated diagnosis model based on different stages of deep learning approaches for identification, segmentation and classification. It appears that the proposed framework integrates diagnosis challenges when scanning masses since these tend to be located within dense tissues, including fat tissue of the breast. The FrCN techniques are used for mass segmentation, which utilizes multiple processing layers and is proposed to address the mass segmentation challenge and obtain a better classification with a diagnosis system. While it is true that segmentation does yield more precise and representative shape features that benefit classification. In fact, segmentation results in better classification results due to improvements in the mass regions.

Moreover, the segmentation of the mass improves the classification rates of the proposed diagnosis system. Since each pixel produces an individual sample during training, which significantly identified the amount of training samples, each pixel can represent a different example during training. When the false rate is reducing, the segmentation will

perform well as Table 1 and 2. Classification performance is also better accordingly, as shown different figures from Fig 7. In Fig 7, we found the number of pairing performances that are very close to each other such as (TPR, TNR), (PPV, NPV), (FNR, FPR), (FOR, FDR), (PT, CSI), (ACC, BA), (ACC, F 1-Score), (ACC, MCC), (F 1-Score, MCC) etc. The diagnosis classification by CNN produces even better results. The benefit of the deep learning FrCN model is the improved ability to segment tissues apart from the main mass. The segments are then categorized into mass-exclusion tissue (with less false positives and negatives) and non-mass-exclusion tissue (with more false positives and negatives). More false positives and negatives are affected on the accuracy or other items of confusion matrix during evaluation as per Table 3. Thus, less performance is found in fig 8. Although other methods have been taken for comparison performance, our model is well than others as Table 5. 6 and 7. From statistical evaluation, we got mean relative region size and coverage of cases are close to each other as Fig 10. The proposed deep learning model of CNN helped enhance the diagnosis system's efficiency by using the deeper features found in the model. Using this description, accurate mass detection and segmentation are crucial to increasing a diagnosis system's feasibility and reliability. Finally, a distinction has been made between our proposed diagnosis system using our methods and another similar system.

7.1. *Implications for Technology Management*

Breast cancer disease are increasing in women's health and has been tried to control as per health management strategy. Since last two decades, this disease affects healthcare technological system due to lack of integrated cancer testing system. The existing testing system is based on time consuming testing for individual parts of testing such as detection, segmentation and classification. It needs strong technological skill to design the integrated framework for detection of cancer disease and can manage the different cancer testing issues.

In this paper, the application of deep learning is utilized to process the integrated framework of proposed model and technological manages the healthcare system to avoid human errors and other significant limitations. Our proposed model detects the disease as benign or malignant from image when image is processed from different stages of our framework which is our aim of the proposed model. Particularly, the role of CNN model (i.e., convolutional, fully connected and SoftMax layer) is important to build this model. Our solution impact on technologically management of this model when we built the different layer using deep learning model for integrated framework. Thus, proposed framework technically manages large number of cancer image detection to evade of human confusion on disease and can also support healthcare management system.

8. Conclusions

In this paper, an integrated diagnosis model is developed with deep learning approaches that are demonstrated to identify, segment, and classify masses from breast cancer disease images in a single setting framework. Making a functional diagnosis system that segments and predicts the form of mass as benign or malignant was challenging on our proposed model with advances in deep learning approaches. The diagnosis system can produce a good performance on classification based on segmentation capability as per a model that employs a FrCN. In this framework, the U-Net model is used with the pixel-to-pixel mass segmentation, which was vital in lowering the number of false positives and negatives. The encoder and decoder network are designed with RELU and Softmax layer on convolution layer to perform well segmentation. This is, in turn, to help boost the

overall efficiency of the proposed diagnosis system. CNN based classification performance are exhibited through the demonstration of segmentation. For better performance, it evaluated the number of equations of confusion matrix with fourfold evaluation. Each fold provides better results such as 99.0% accuracy, 98.8 % sensitivity, 98.8% precision, 98.8% F1-score, etc. Thus, our integrated diagnosis model will help to detect the breast cancer disease and analyze the patient information in healthcare system. Future work can develop a new system for clinical applications in radiology using the feature selection approach for image data.

REFERENCES

1. Breast cancer - WHO | World Health Organization, 26-Mar-2021. https://www.who.int/news-room/fact-sheets/detail/breast-cancer
2. M.A. Al-antari, M.A. Al-masni, S.U. Park, J.H. Park, M.K. Metwally, Y.M. Kadah, S.M. Han, T.-S. Kim, An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network, *J. Med. Biol. Eng*. Vol. 38 Issue no. 3, Pages: 443–456, 2017. http://dx.doi.org/10.1007/s40846-017-0321-6.
3. Y. Wang, D. Tao, X. Gao, X. Li, B. Wang, Mammographic mass segmentation: embedding multiple features in vector-valued level set in ambiguous regions, *pattern recognition*. 44 (no. 9) (2011) pp. 1903–1915.
4. S. Lee, C. Lo, C. Wang, P. Chung, C. Chang, C. Yang, P. Hsu, A computer-aided design mammography screening system for detection and classification of microcalcifications, *Int. J. Med. Inf*. 60 (no. 1) (2000) pp. 29–57.
5. M. Al-masni, M. Al-antari, j. Park, G. Gi, T. Kim, P. Rivera, E. Valarezo, S.-M. Han,T.-s. Kim, Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network, *39th Annual InternationalConference of the IEEE Engineering in Medicine and Biology Society (EMBC'17),* Jeju Island, South Korea, 2017, pp. 1230–1236.
6. M.A. Al-masni, M. Al-antari, J.-m.P. Park, G. Gi, T.-Y.K. Kim, P. Rivera, E. Valarezo,M.-T. Choi, S.-M. Han, T.-S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, *Comput. Methods Prog. Biomed*. 157 (2018) pp. 85–94.
7. M.A. Al-antari, M.A. Al-masni, Y.M. Kadah, Hybrid model of computer-aided breast cancer diagnosis from digital mammograms, *J. Sci. Eng*. 04 (no. 2) (2017), pp. 114–126.
8. N. Dhungel, G. Carneiro, A.P. Bradley, Automated mass detection in mammograms using cascaded deep learning and random forests, *International Conference on digital Image Computing: Techniques and Applications (DICTA),* Australia, 2015. DOI: 10.1109/DICTA.2015.7371234
9. A. Negi, A. N. Joseph Raj, R. Nersisson, Z. Zhuang, M Murugappan, RDA-UNET-WGAN: an accurate breast ultrasound lesion segmentation using Wasserstein generative adversarial networks, *Arabian Journal of Science and Engineering*, 45, Issue 8, 2020, pp. 6399–6410.
10. Z. Zhuang, A. N. J. Raj, A. Jain, N. Ruban, S. Chaurasia, N. Li1, M. Lakshmanan, and M. Murugappan, Nipple segmentation and localization using modified u-net on breast ultrasound images, *Journal of Medical Imaging and Health Informatics*, Vol. 9, 2019, pp. 1827–1837.
11. G. Carneiro, J. Nascimento, A.P. Bradley, Automated analysis of unregistered multi-view mammograms with deep learning, *IEEE Trans. Med. Imaging*, 36 (no. 11)(2017) pp. 2355–2365.
12. L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, *IEEE Trans. Med. Imaging* 36(no. 4) (2017), pp. 994–1004.
13. Z. Jiao, X. Gao, Y. Wang, J. Li, A deep feature based framework for breast masses classification, *Neurocomputing* 197 (no. C) (2016) pp. 221–231.
14. I. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. Cardoso, J. Cardoso, INbreast: toward a full-field digital mammographic database, *Acad. Radiol*. 19 (no. 2) (2012), pp.236–248.
15. J.S. Cardoso, I. Domingues, H.P. Oliveira, Closed the shortest path in the original coordinates with an application to breast cancer, *Int. J. Pattern Recognit. Artif.Intell*. 29 (no. 1) (2015) 2.
16. N. Dhungel, G. Carneiro, A.P. Bradley, Deep learning and structured prediction for the segmentation of mass in mammograms, *International Conference on MedicalImage Computing and Computer-Assisted Intervention*, (2015), pp. 605–612.
17. E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (no. 4) (2017), pp. 640–651.

18. V. Badrinarayanan, A. Kendall and R. Cipoll, SegNet: A Deep ConvolutionalEncoder-Decoder Architecture for Image Segmentation, *in arXiv preprintarXiv*:1511.00561, 2016.
19. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-ScaleImage Recognition," *in arXiv preprintarXiv*:1409.1556, 2014.
20. R. Llobet, J. Perez-Cortes, A. Toselli, A. Juan, Computer-aided detection of prostate cancer, *Int. J. Med. Inf.* 76 (no. 7) (2007) , pp. 547–556.
21. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, 25th International Conference on Neural information processing Systems, USA, 2012, pp. 1097–1105.
22. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, *IEEE Conference on Computer Vision and PatternRecognition*, (2016).
23. L. lab, Theano, [Online] Available: University of Montreal, 2017 (Accessed 10,2017), http://deeplearning.net/software/theano/tutorial/.
24. F. Chollet, Keras: The Python Deep Learning Library, [Online]. Available: MIT,2017 (Accessed 10, 2017), https://keras.io/.
25. Google Brain Team, TensorFlow, 9 11 2017. [Online]. Available: (2017) (Accessed10, 2017), www.tensorflow.org.
26. H. K. Bhuyan, N. K. Kamila, S. K. Pani, Individual privacy in data mining using fuzzy optimization, *Engineering Optimization*, 2021, pp. 1-19, (Early Access).
27. M. A. Al-antari, M. A. Al-masni, M. Choi, S. M. Han, 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*, Volume 117, 2018, pp. 44-54.
28. T. Mahmood, J. Li, Y. Pei , F Akhtar, A. Imran, and K. Ur Rehman, A Brief Survey on Breast Cancer DiagnosticWith Deep Learning Schemes UsingMulti-Image Modalities, *IEEE Access*, Volume 8, 2020, pp. 165779-165809,.
29. T. Saba, Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges, *Journal of Infection and Public Health*, Vol.13, 2020, pp. 1274–1289,
30. V. Lahoura, H. Singh, A. Aggarwal, B. Sharma, M A. Mohammed , R. Damaševiˇcius, S. Kadry and K. Cengiz, Cloud computing-based Framework for Breast Cancer Diagnosis Using Extreme Learning Machine, *Diagnostics*, 11, 241, Feb. 2021, pp. 1-19,.
31. Y. Jiménez-Gaona, M. José Rodríguez-Álvarez and V. Lakshminarayanan, Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review, *Appl. Sci*. 10, 8298, Nov 2020, pp. 1-28.
32. I. Dankwa-Mullan, M. Rivo, M. Sepulveda, Y. Park, J. Snowdon, and K. Rhee, ``Transforming diabetes care through artificial intelligence: The future is here," *Population Health Manage.*, vol. 22, no. 3, Jun. 2019, pp. 229-242.
33. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.*, vol. 42, no. 9, Dec. 2017, pp. 60-88.
34. S. Duraisamy and S. Emperumal, Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier, *IET Comput. Vis.*, vol. 11, no. 8, Dec. 2017, pp. 656-662.
35. H. Intisar Rizwan I, N. Jeremiah, Deep learning approaches to biomedical image segmentation, *Informatics in Medicine Unlocked*, vol 18, 2020, pp. 1-12.
36. Z. Wang, Mo Li , H. Wang , H. Jiang , Y. Yao , H. Zhang , and J. Xin, Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features, IEEE Access, Vol -7, 2019, pp. 105146- 105158 .
37. (Online) https://en.wikipedia.org/wiki/Confusion_matrix 2021.
38. S. Wessam M., Aly Moustafa H., Deep learning in mammography images segmentation and classification: Automated CNN approach, *Alexandria Engineering Journal*, Vol-60, 2021, pp. 4701-4709.
39. H. K. Bhuyan, C Chakraborty, S. K. Pani, V. K. Ravi, Feature and Subfeature Selection for Classification Using Correlation Coefficient and Fuzzy Model, *IEEE Transactions on Engineering Management*, 2021, pp. 1-15, (Early Access).
40. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical Image Computing and ComputerAssisted Intervention, 9351, Springer, Cham*, 2015, pp. 234–241.

41. N. Alam, A. Oliver, E.R. Denton, R. Zwiggelaar, Automatic segmentation of microcalcification clusters, *Annual Conference on Medical Image Understanding and Analysis, 894, Springer, Cham*, 2018, pp. 251–261.
42. S. Duraisamy, S. Emperumal, Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier, *IET Computing*. Vis., 11 (8) (2017), pp. 656–662.
43. [https://wiki.cancerimagingarchive.net/display/Public/CBISDDSM. Accessed 1 June 2019.
44. HK Bhuyan, C Chakraborty, Y Shelke, SK Pani, COVID-19 diagnosis system by deep learning approaches, Expert Systems, pp. 1-18, 2021 (Early accessed).
45. T. Daim, K. K. Lai, H. Yalcinc, F. Alsoubie, V. Kumar, Forecasting technological positioning through technology knowledge redundancy: Patent citation analysis of IoT, cybersecurity, and Blockchain, *Technological Forecasting and Social Change*, Volume 161, 120329, December 2020.
46. T. R.Anderson, T. U. Daim, Francois F.Lavoie, Measuring the efficiency of university technology transfer, *Technovation*, Volume 27, Issue 5, May 2007,Pages 306-318.
47. T. U. Daim, N. Basoglu &U. Topacan, Adoption of health information technologies: the case of a wireless monitor for diabetes and obesity patients, *Technology Analysis & Strategic Management*, Volume 25, Issue 8, , 2013, Pages 923-938.
48. V. K. Singh, H. A.Rashwan, S. Romani, F. Akram, N.Pandey, Md. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig, J. Torrents-Barrena, Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network, *Expert Systems with Applications*, Volume 139, January 2020, page: 1-15, Article No.112855.