



# Fully convolutional network for automated detection and diagnosis of mammographic masses

Sujata Kulkarni<sup>1</sup> · Rinku Rabidas<sup>1</sup>

Received: 14 September 2021 / Revised: 19 January 2022 / Accepted: 5 February 2023 /  
Published online: 9 May 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Breast cancer, though rare in male, is very frequent in female and has high mortality rate which can be reduced if detected and diagnosed at the early stage. Thus, in this paper, deep learning architecture based on U-Net is proposed for the detection of breast masses and its characterization as benign or malignant. The evaluation of the proposed architecture in detection is carried out on two benchmark datasets— INbreast and DDSM and achieved a true positive rate of 99.64% at 0.25 false positives per image for INbreast dataset while the same for DDSM are 97.36% and 0.38 FPs/I, respectively. For mass characterization, an accuracy of 97.39% with an AUC of 0.97 is obtained for INbreast while the same for DDSM are 96.81%, and 0.96, respectively. The measured results are further compared with the state-of-the-art techniques where the introduced scheme takes an edge over others.

**Keywords** Deep learning · U-Net · Mammogram · Mass detection · Classification

## 1 Introduction

Breast cancer is one of the leading causes of deaths among women globally. As per World Health Organization (WHO), breast cancer registers 2.09 million cases with 6,27000 deaths globally in 2019 [41]. Among different types of cancers, approximately 13% of total cases are diagnosed with breast cancer globally which clearly depicts its fatal nature. Detection and diagnosis of breast cancer at the early stage can reduce the mortality rate by manifolds [16]. Screening mammograms is preferred technique for detection of abnormalities [33, 43] but at the same time, the assessment of mammograms is a challenging task due to variations in anomalies, low contrast images, and less number of skilled radiologists which causes fatigue due to continuous examinations of large number of mammograms. To overcome this,

---

Rinku Rabidas contributed equally to this work.

✉ Sujata Kulkarni  
sujata.kulkarni@aus.ac.in

Rinku Rabidas  
rabidas.rinku@gmail.com

<sup>1</sup> Department of Electronics & Communication Engineering, Assam University, Silchar, 788010, Assam, India

Computer aided detection/diagnosis (CADe/x) systems are evolved which acts as second evaluator and improved the overall screening efficiency significantly [8].

Micro-calcification, architectural distortion, bilateral asymmetry, and masses are the early symptoms of breast cancer but among all, detection and diagnosis of masses is a challenging task due to its variations in shape, margin and size. A mass with round or oval shape and low density is generally benign while ill-defined shape, spiculated margin with high density is categorized as malignant. To address this problem, many methodologies are reported and treats the problem of detection and diagnosis separately, but fails to guarantee complete success. Several researchers have introduced CAD systems, both traditional or deep learning based, to improve the performance in the detection and diagnosis of mammographic masses [17, 18, 25, 31]. In traditional methods, Chakraborty et al. proposed radial region growing method based on thresholding for mass detection and achieved a sensitivity of 90.0% at 1.4 false positives per image (FPs/I) and an AUC of 0.87 with an accuracy ( $A_{cc}$ ) of 77.89% for the diagnosis of malignant masses when evaluated on 1257 mammograms of DDSM database [11]. Casti et al. introduced combination of Gaussian curvature and gabor filter responses for detection and contour independent features for diagnosis and reported a sensitivity of 0.80 at 5.21 FPs/I for detection and an AUC of 0.61 when evaluated on 1662 DDSM images [10]. Laishram and Rabidas proposed a hybrid model of fast and robust fuzzy c-means clustering (FRFCM) for the detection of masses on mini-MIAS dataset and achieves sensitivity of 96.6% with 2.29 FPs/I while the same obtained an improvement of 74% in FPs/I after performing false positive reduction [24]. A comparison of traditional CAD system and CNN networks is conducted by Kooi et al. on a private dataset where CNN outperforms the traditional CAD systems and obtained an AUC of 0.87 [23]. Arevalo et al. also compared the features extracted via CNN architecture against other hand crafted measures like Histogram of Oriented Gradient (HOG) and Histogram Gradient Divergence (HGD) methods [7] and obtained an AUC value of 0.86 on Breast Cancer Digital Repository (BCDR) dataset. Sethy et al. performed a survey for the methods useful for the detection and diagnosis of breast cancer and also developed a robust and cost effective approach using the HOG feature and SVM to detect and diagnose the breast cancer and achieved an accuracy of 99.64% [40]. From the above literature, it can be observed that the majority of the CAD systems with conventional machine learning methods mainly used hand crafted features which is immensely dependent on the human experts. This restricts the complete automation of CAD systems.

Deep learning models have shown promising performance in various biomedical applications such as diagnosis of breast cancer, coronavirus [39], arrhythmia [20] and myocardial infarction [19] etc. In case of breast cancer, Dhungel et al. combines deep learning strategy with random forest classifier for identifying the mass cases and achieved a true positive rate (TPR) of 0.96 at 1.2 FPs/I on INbreast when tested on 115 images while the same for DDSM-BCRP dataset is 0.75 at 4.8 FPs/I [13] using 79 images which is very limited size of dataset. Patch based CNN network is proposed by Agarwal et al. for mass detection and obtained a TPR of 0.98 at 1.67 FPs/I on CBIS-DDSM dataset [2] while the author improves the architecture by proposing Faster region based CNN (Faster R-CNN) for mass detection and reported a TPR of 0.93 at 0.78 FPs/I on FFDM dataset and 0.95 at 1.14 on INbreast dataset [1]. The Faster R-CNN is further modified by Akselrod-Ballin et al. where an accuracy of 72% and 78% is obtained for detection and diagnosis, respectively [3] when examined on a private dataset. Transfer learning approach is utilized by Ribli et al. where model based on Faster R-CNN is trained on DDSM dataset and tested on INbreast dataset with an accuracy of 90% with 0.3 FPs/I [36]. However, Faster R-CNN involves high computations while defining region boxes before classification. In [4], You

only Look Once (YOLO) approach is introduced for mass detection, full resolution convolutional network (FrCN) for segmentation, and CNN network for benign-malignant mass classification with  $A_{cc}$  of 98.96%, 92.97%, and 95.64%, respectively. Similar approach of YOLO based CAD model for detection and classification of masses proposed by Al-masani et al. reported an  $A_{cc}$  of 99.7% and 97%, respectively using DDSM dataset [5]. In YOLO based approaches, YOLO algorithm struggles to identify small objects within the image which is a matter of concern if the abnormality is of smaller size. Hassan et al. [21] proposed classification of breast masses using transfer learning approach where pre-trained AlexNet and GoogleNet models are utilized and achieved an accuracy of 100% for CBIS-DDSM and INbreast datasets with AlexNet model while 98.46% and 92.5% for GoogleNet model when tested on a combination of two datasets having 1920 (augmented) region of interest (ROI) patches. Sarkar et al. developed deep hierarchical mass prediction network for mass detection and utilized wavelet features using CNN for classification [38]. Modified VGG16 network with its fully connected layer replaced by a convolutional layer is used and achieved a sensitivity of 94% and 1.13 FPs/I on 350 mammograms when tested on DDSM dataset and 96% at 0.96 for INbreast dataset with 41 images. The classification accuracy over DDSM dataset with 226 mammograms is 98.05% and for INbreast with 43 images is 98.14%. Mahmood et al. proposed deep learning based ConvNet for classification of masses [28]. The method consists of preprocessing, data augmentation, feature extraction and classification steps and reported classification accuracy of 97.05% when tested on a 2160 (augmented) images which are a combination of mammograms acquired from mini-MIAS as well as private dataset. Alanazi et al. developed three different CNN model with variations in number of convolutional layers (i.e. 2, 6, and 5) for diagnosis of breast cancer [6]. Out of these architectures, model with 5 convolutional layers achieved highest accuracy of 87% when tested on about 275,000 image patches. Qiu et al. developed CAD system for the classification of breast mass into benign and malignant [34]. The network consists of 3 pairs of convolutional and max pooling layers as a feature extractor, one hidden layer and one logistic regression layer are utilized as multiple layer perceptron (MLP). The overall AUC reported is 0.79 when tested on 560 images from full-field digital mammography (FFDM).

From the above discussion, it is observed that the performance of CAD tools can achieve high sensitivity, but at the cost of a relatively higher false-positive rate. Most of the previous works tries to address the detection and diagnosis problems separately hence limited works is reported to address the two issues by an integrated CAD system. It is believed that deep learning methods can obtain better performance in detection and diagnosis of masses, as deep learning has ability to automatically learn features and avoiding complication of the feature designing and selection procedure. Adding to it, only a few works based on deep learning have been presented for mass detection as well as classification of masses from mammograms which motivate to explore deep learning as the underlying framework. Therefore, this paper introduces an integrated deep learning architecture based on U-Net for the detection and diagnosis of mammographic masses as benign or malignant. The main feature of U-Net architecture is that it is independent of dimensions of an input image as there is no dense layer in the detection mode. It has capability to work exclusive of any preprocessing techniques like noise removal, image enhancement and suppression of pectoral muscle or pre-trained weights. The up-sampling and down-sampling operations were performed to overcome the resolution loss of feature maps due to the multiple max-pooling and transposed convolutional layers. The higher sensitivity and lower false positives (FPs) makes

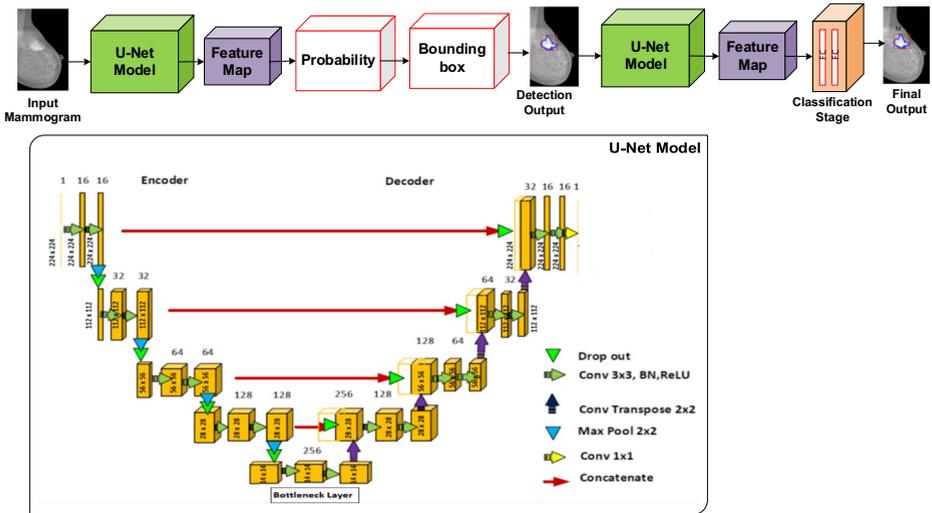
the introduced approach suitable for integration into clinical practice along with skilled radiologists as a second examiner. The key contributions of the work can be summarized as:

- Integrated framework based on deep learning is proposed for automatic detection and characterization of mammographic masses.
- U-Net architecture is introduced for mass detection which is independent of the dimension of the input mammograms.
- With the best of our knowledge, the efficiency of U-Net architecture in classification is investigated for the first time.
- The performance is evaluated on two benchmark databases— INbreast and DDSM dataset where it outperforms the other techniques in the state-of-the-arts with high sensitivity and low FPs.

The paper is organized as follows. Section 2 discusses methodology for the detection and classification of masses. Section 3 provides the experimental set up and an overview of original datasets employed for performance evaluation of the proposed approach. Section 4 represents the performance analysis of proposed model through five-fold cross validation. Section 5 presents analysis of the results obtained against other models and Section 6 concludes the paper.

## 2 Methodology

This paper presents an integrated deep learning (DL) architecture for the automated detection of mammographic masses followed by the characterization of identified masses as benign or malignant. The DL architecture incorporates U-Net structure for both localization and classification of mass lesions. The schematic diagram of proposed approach is illustrated in Fig. 1. The input mammogram is first applied to the U-Net architecture which consists of encoder, decoder and bottleneck part. The detail description of proposed U-Net model is discussed in Section 2.2. The model produces a feature map. The extracted



**Fig. 1** Schematic diagram of the proposed U-Net based deep learning network to detect and classify breast cancer masses

features are utilized for the accurate detection and location of mass lesion from the image. To reflect the probability of predicting a mass or non mass, an activation function, sigmoid, is used in the last layer of U-Net model. From the predicted mask, a bounding box is created. Only the detected mass which is having a minimum of 50% overlapping between the ground truth and the predicted boxes are then applied to next stage for characterization. The features map again generated by the proposed U-Net model followed by two fully connected (FC) layers which uses a classifier in the output layer. The classifier is a softmax activation function with two nodes to distinguish between benign or malignant lesions.

## 2.1 Data preprocessing

Although the introduced approach can take full size mammograms, but to avoid the computational complexity, which incurred due to the processing of large size mammograms, all mammographic images are resized to  $224 \times 224$  followed by normalization. To overcome the requirement of large number of samples for training the deep learning architectures which is a key issue in case of biomedical application, data augmentation is employed as done in [32, 42] for both the datasets. Data augmentation is a process of increasing the number of images and diversity by applying different transformation to the original ones. In the present work, the original mammograms are augmented four times ( $\times 4$ ) by applying flip in up-down directions with rotation by an angle of  $90^\circ$ . Similarly for eight times ( $\times 8$ ), in addition to the operations performed for  $\times 4$ , augmentation is done by applying flip in left-right directions with rotation by an angle of  $90^\circ$ . The same with other details are provide in Table 1.

## 2.2 U-Net based detection

The architecture of U-Net was first proposed by Ronneberger et al. for segmentation in biomedical images [37]. U-Net is an end-to-end neural network architecture typically bears a ‘U’ shape. It is also a Fully-convolutional network (FCN) which is independent of the size of the input and the architecture of the same is demonstrated in Fig. 2.

Broadly, the U-Net structure can be divided into encoder and decoder part. The encoder stage comprises of pooling layers which performs downsampling operation which reduces the dimensions of each image retaining the salient information while the decoder does the opposite of it i.e. upsampling and enable precise localization using transposed convolutions. The feature maps, in encoder, after each downsampling step is combined with the corresponding feature map after upsampling in the decoder network using concatenation operation. The red and purple arrows shown in Fig. 2, indicate concatenation of two feature maps. The concept of U-Net in this sense is that while upsampling in the network, it

**Table 1** Detailed description about images used for experimental purpose from DDSM and INbreast dataset

Data set	Original	$\times 4$ Augmented	$\times 8$ Augmented
	mammogram	Flip_U_D with Rot $90^\circ$	Flip_U_D+L_R with Rot $90^\circ$
INbreast	Benign	$37 \times 4 = 148$	$37 \times 4 + 37 \times 4 = 296$
	Malignant	$69 \times 4 = 276$	$69 \times 4 + 69 \times 4 = 552$
DDSM	Benign	$300 \times 4 = 1200$	$300 \times 4 + 300 \times 4 = 2400$
	Malignant	$700 \times 4 = 2800$	$700 \times 4 + 700 \times 4 = 5600$
		Total	Total
		424	848
		4000	8000

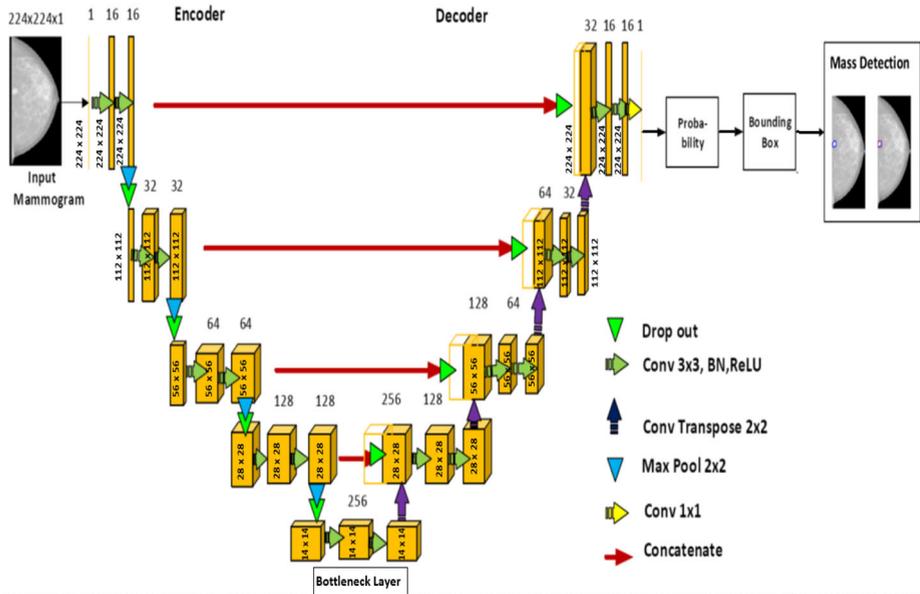


Fig. 2 Proposed U-Net architecture for automatic detection of masses

is also concatenating the higher resolution feature maps from the encoder network in order to learn better representations with following convolutions. Since upsampling is a sparse operation, a good prior from earlier stages is good to represent the localization. The proposed DL architecture incorporates U-Net for both mass detection as well as classification with a few modifications in the respective stages. For mass detection, the full mammograms are applied as input to the encoder of proposed U-Net architecture as it uses single convolutional neural network to the entire image. At the end of the decoder, 1 × 1 convolutional layer is appended to obtain the final feature map. To reflect the probability of predicting a mass or non mass, an activation function, sigmoid, is used on the last layer. From the predicted mask, a bounding box with the confidence score is calculated which expresses the probability of identified region as mass multiplied with the percentage of the intersection over union (IoU) between the ground truth and the predicted boxes. More details about the encoder and decoder are presented in the upcoming subsections.

### 2.2.1 Encoder section

Encoder, also called as the contraction path, is used to extract the features from the image. Encoder starts with an input image of size  $h_{inx} \times h_{iny} \times n_c$  where  $h_{inx}$  and  $h_{iny}$  are spatial dimensions of the image with  $n_c$  as channel dimension. In the present work, the values of  $h_{inx} \times h_{iny} \times n_c$  are set as  $224 \times 224 \times 1$ , respectively. A 3 × 3 convolution is combined with the batch normalization and Rectified linear unit (ReLU) as an activation function. The rectified feature map is further operated with 2 × 2 max pooling function which is useful for extracting the extreme features such as edges. The operators of convolution, ReLU, and max pooling is applied iteratively and the feature map obtained at each iteration is concatenated to the decoder module as shown in Fig. 2. This concatenation operation helps to avoid loss of

significant information. Moreover, a drop out of 0.5 is also added to the network of encoder to avoid overfitting problems.

### 2.2.2 Decoder section

Decoder, also called as the expansion path, is used to provide precise localization using transposed convolution. In contrast to the encoder, it does upsampling with convolutional block of transposed convolution. The up-sampling stage allows network to transfer context information from the low spatial resolution layer to the higher spatial resolution layer. The depth of the image is decreases gradually with increase in the dimensions of the image. The upsampling stage retrieves the information in addition to the measures provides by the concatenation of its counterpart from encoder. The last layer employs a  $1 \times 1$  convolution to map each 16-component feature vector to the desired number of classes. The proposed architecture for automatic detection of mammographic masses using U-Net is illustrated in Fig. 2. The number of encoder and decoder blocks is kept equal while the number of filters increase and decrease in encoder and decoder parts, respectively. The output of a decoder is also of size  $h_{out} \times h_{in} \times n$  and the relationship between  $h_{in}$  and  $h_{out}$  can be expressed as:

$$h_{out} = \frac{h_{in} + 2p - n}{s} + 1 \tag{1}$$

where,  $h_{in}$  and  $h_{out}$  represents number of input and output features, respectively with  $n$  as convolutional kernel size of  $3 \times 3$ ,  $p$  indicates convolutional padding of same while the convolutional stride size of 2 is represented by  $s$ .

### 2.3 U-Net based classification

The characterization of correctly identified mass lesions are conducted via same U-Net structure with slight modifications in the architecture. The correctly detected mass regions are cropped automatically and fed as an input to the encoder part of U-Net. The encoder path consists of two consecutive  $3 \times 3$  convolutional layers, batch normalization, and ReLU as an

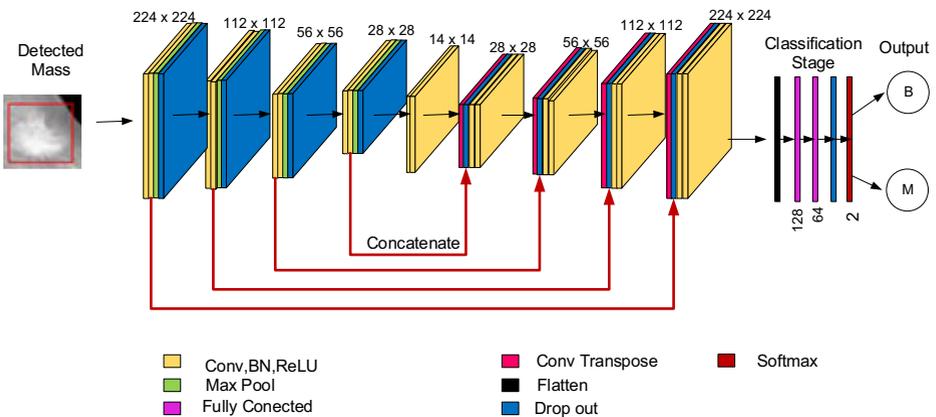


Fig. 3 Proposed U-Net architecture for characterization of masses

activation function followed by max pooling operation with a stride of 2. So, in encoder path, the feature maps get doubled at each step and becomes 256 from 16. The decoder part, also called as upsampling path, consists of 2x2 convolutional transpose layer and concatenation layer which add features from equivalent encoder path as shown in Fig. 3. This is followed by two consecutive 3x3 convolutional layers with batch normalization and ReLU as an activation function. In decoder path, the feature maps go down at each step from 256 to 16. Drop out of 0.5 is also added at every step in encoder and decoder to avoid overfitting of the model. At the end of the decoder path, flatten, drop out of 0.5 and two fully connected layers with 128 and 64 nodes are used followed by the softmax layer which is used to predict the masses as benign or malignant. The introduced network for diagnosis is represented in Fig. 3.

### 3 Experimental set up

The proposed deep learning architecture for the localization and characterization of mammographic masses is examined on a workstation with Intel(R) Xenon(R) processor clocked at a speed of 3.60 GHz with 64GB RAM and NVIDIA Quadro RTX 4000 GPU. The implementation is done using Python, OpenCv, and Keras, where CUDA, cuDNN library developed by NVIDIA using Tensor flow. The evaluation is conducted on the benchmark datasets— DDSM and INbreast with data augmentation as discussed in Section 2.1. The augmented dataset is randomly splitted in the ratio of 60:20:20 for training, testing, and validation purposes, respectively. The 5-fold cross validation is performed for all detection deep learning models: Faster R-CNN [35], RetinaNet [27], and the proposed U-Net. For training all these models, Adam optimizer with binary cross entropy loss function is used where the number of epochs is set to 100. In the detection stage, the probable region of mass may be decided on different values of threshold but in the present study, the threshold value is set to 0.3 as lower than this value will result in higher FPs while the values greater than this will have high missed cases as observed from various experiments. However, the final call of a mass detected (True positive) is done by considering intersection over union (IoU) criteria [9]. Intersection over Union (IoU) is an assessment metric which is used to measure the accuracy of detection algorithm on a dataset. IoU can be defined as:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

Where A is the bounding box drawn on ground truth mask, B is the bounding box drawn on predicted mask from model. Bounding box indicates a rectangular region formed by the coordinates of its top left and the bottom right point of the contour of the ground truth as provided along with the dataset and the identified mass lesions. So, IoU is simply the ratio of area of intersection with respect to area of union among ground truth mask and predicted mask. The aim of model would be to keep improving the prediction so that the IoU between the two boxes becomes equal to 1. AUC ROC is a curve [29] which shows the performance of a classification model at all classification thresholds and free-response operating characteristic (FROC) curve shows the accuracy of detected mass which is plotted between true positive rate versus false positive per image. F1-score sums up the predictive performance of a model by combining two competing matrices— precision and recall. Precision is one indicator of a model's performance which refers to the number of true positives divided

by the total number of positive predictions. The Recall is the measure of model correctly identifying True Positives. The expressions for the same are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$F1 - score = \frac{2 * TP}{2 * TP + FP + FN} \tag{6}$$

where TP indicates true positive;TN is true negative; FP is false positive, and FN is false negative. Adding to these, FPR, MCC and Kappa co-efficient are also measured. FPR (false positive rate) defines how many incorrect positive results occur among all negative samples available during the test, MCC(Mathew’s correlation coefficient) shows the ineffectiveness of the classifier in classifying especially the negative class samples and k (Kappa co-efficient) is a measure of how closely the classified label matched with ground truth [12]. So, the higher the kappa coefficient, the more accurate the classification is. The definitions for all the metric are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \tag{9}$$

$$k = \frac{2 * (TP * TN - FP * FN)}{(TP + FP) * (TN + FP) + (TP + FN) * (TN + FN)} \tag{10}$$

### 3.1 Materials

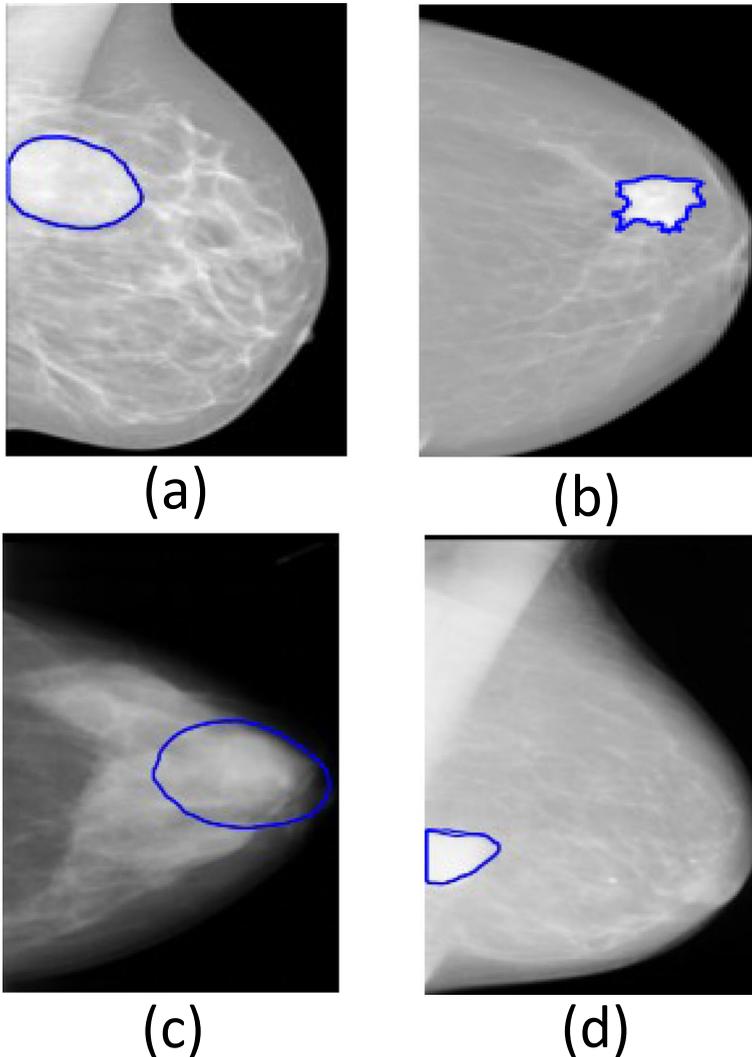
#### 3.1.1 INbreast

The INbreast database is prepared in Centro Hospitalar de S. Joao (CHSJ) hospital, Breast Centre, Porto [30]. The images are acquired by MammoNovation Siemens FFDM having a pixel size of 70 μm , 14 bit resolutions, and image sizes of 3328 × 4084 or 2560 × 3328. Database consists of a total of 115 cases of women with mediolateral oblique (MLO) and craniocaudal (CC) views accounting to a total of 410 mammograms. Out of these, there are 106 mammograms with mass lesions (37 benign and 69 malignant cases) which has been considered in the current work. The annotations about the anomalies, defined by skilled radiologists, are provided along with the database.

#### 3.1.2 DDSM

The Digital Database of Screening Mammography (DDSM) dataset is also a publicly available dataset [22] managed by the University of South Florida. The mammograms are digitized at spatial resolutions of 50 μm/pixel, 43.5 μm/pixel, and 42 μm/pixel with

gray-level resolution of 16 and 12 bits/pixel. DDSM also contains MLO and CC views. The present work employs 1000 mammograms selected randomly out of which 300 cases belongs to benign while the remaining are malignant in nature. The annotations about the abnormalities are also available along with the dataset. The sample mammograms for DDSM as well as INbreast dataset are shown in Fig. 4.



**Fig. 4** [a-b] illustrates the sample mammograms from INbreast dataset while [c-d] represents the images from DDSM dataset. Blue color contour indicates the mass region containing benign cases in [a] and [c] while malignant cases in [b] and [d]

## 4 Results

### 4.1 Mass detection

The proposed U-Net deep learning model is investigated on publicly available database namely INbreast and DDSM and the outcomes for mass detection obtained with five-fold cross validation are provided in Tables 2 and 3 for respective datasets. Tables 2 and 3 present a comparative analysis of the proposed approach in mass identification with and without data augmentation. Initial study without data augmentation obtained a sensitivity of 91.71%, with precision, F-1 score and FPs/I are 85.87%, 85.69% and 0.53, respectively for the INbreast dataset while the same for the DDSM are 89.86%, 87.60%, 88.71% and 0.62 respectively. When the analysis is extended with data augmentation of 4 times, Sensitivity (Sen), precision, F1, and FPs/I for INbreast improved to 93.82%, 90.71%, 92.10% and 0.40, respectively and the same for DDSM dataset are 90.46%, 90.24%, 90.35% and 0.55. The 4x augmentation has shown slight improvement in the performance. Therefore, 8 x augmentations is performed where 99.64%, 99.64%, 99.63% and 0.25 is observed as Sen, precision, F1, and FPs/I respectively for INbreast dataset while the same for DDSM are 97.36%, 98.06%, 97.71% and 0.38, respectively. It can be observed that the performance with 4 times augmentation doesn't improve much as compared to 8 times on both the datasets. But the analysis is limited to 8 times to avoid computational complexity. In case of detection, it is necessary to achieve high value of recall than precision because the recall measures the model's ability to detect positive samples and the same can be observed from Tables 2 and 3 which establishes the efficiency of the proposed model. The same observation at different augmentation is compared in terms of FROC curves are illustrated in Figs. 5 and 6 for the INbreast and DDSM datasets, respectively. The FROC curve represents the tradeoff between the FPs/I and TPR and the curve between the two is expected to cover the maximum region. The more is the area the better is the performance in detection. Ideally,

**Table 2** Mass detection results obtained via proposed approach for 5-fold cross validation using INbreast dataset

	Parameters	1st Fold	2nd Fold	3rd Fold	4th Fold	5th fold	Average
Original	Sen=Rec (%)	80	92.86	92.86	92.86	100	<b>91.71</b>
Data	Precision (%)	80	86.67	92.86	76.47	93.33	<b>85.87</b>
set	F1-score (%)	80	86.19	91.65	76.33	94.28	<b>85.69</b>
	FPs/Image	0.60	0.50	0.50	0.50	0.57	<b>0.53</b>
Data	Sen=Rec (%)	96.36	83.64	94.55	100	94.55	<b>93.82</b>
Augmentation	Precision (%)	85.48	93.88	88.14	94.83	91.23	<b>90.71</b>
(×4)	F1-score (%)	90.60	88.46	91.23	97.35	92.86	<b>92.10</b>
	FPs/Image	0.40	0.39	0.42	0.36	0.42	<b>0.40</b>
Data	Sen=Rec (%)	99.09	100	99.09	100	100	<b>99.64</b>
Augmentation	Precision (%)	99.09	100	100	100	99.09	<b>99.64</b>
(×8)	F1-score (%)	99.09	100	99.54	100	99.54	<b>99.63</b>
	FPs/Image	0.27	0.23	0.28	0.23	0.23	<b>0.25</b>

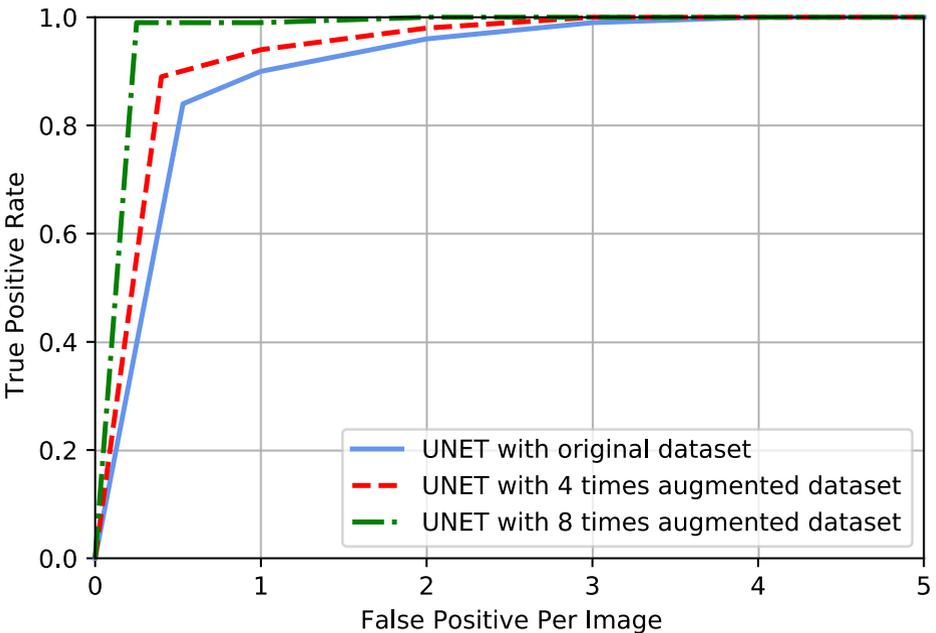
Bold values indicate the average value performance of proposed approach U-Net for mass detection via 5-fold cross validation using INbreast dataset for Original dataset, with 4 times data augmentation and 8 times data augmentation technique

**Table 3** Mass detection results obtained via proposed approach for 5-fold cross validation using DDSM dataset

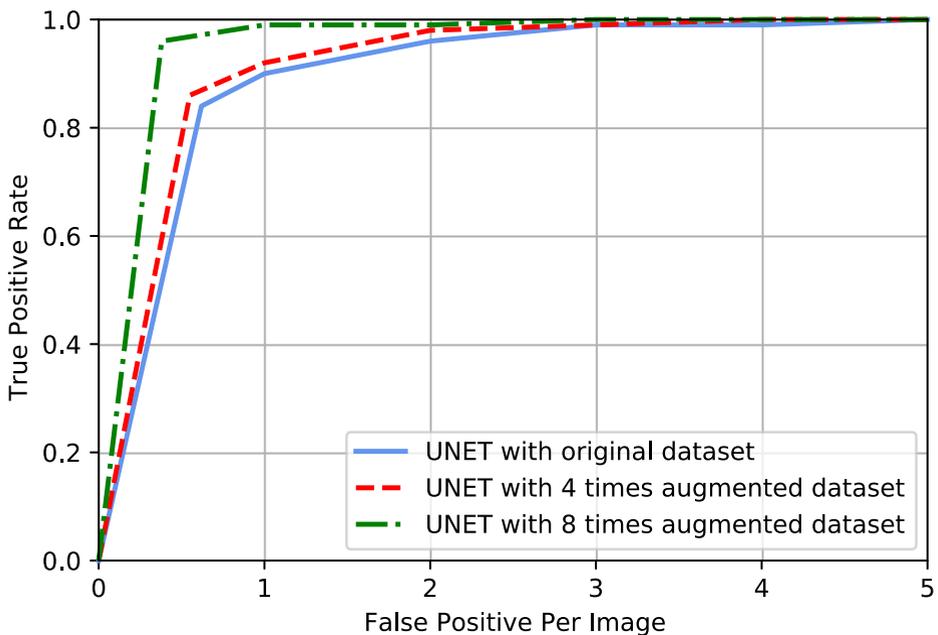
	Parameters	1st Fold	2nd Fold	3rd Fold	4th Fold	5th fold	Average
Original	Sen=Rec (%)	88.57	88.57	89.29	90	92.86	<b>89.86</b>
Data	Precision (%)	86.11	86.71	87.41	88.73	89.04	<b>87.60</b>
set	F1-score (%)	87.32	87.63	88.34	89.36	90.91	<b>88.71</b>
	FPs/Image	0.64	0.64	0.62	0.61	0.57	<b>0.62</b>
Data	Sen=Rec (%)	88.39	91.07	90.54	90.36	91.96	<b>90.46</b>
Augmentation	Precision (%)	89.19	89.63	90.21	91.17	90.99	<b>90.24</b>
(×4)	F1-score (%)	88.79	90.35	90.37	90.76	91.47	<b>90.35</b>
	FPs/Image	0.63	0.54	0.55	0.56	0.49	<b>0.55</b>
Data	Sen=Rec (%)	96.96	97.32	97.41	97.32	97.77	<b>97.36</b>
Augmentation	Precision (%)	97.31	98.2	98.11	98.29	98.38	<b>98.06</b>
(×8)	F1-score (%)	97.14	97.76	97.76	97.8	98.07	<b>97.71</b>
	FPs/Image	0.40	0.38	0.38	0.38	0.36	<b>0.38</b>

Bold values indicate the average value performance of proposed approach U-Net for mass detection via 5-fold cross validation using DDSM dataset for Original dataset, with 4 times data augmentation and 8 times data augmentation technique

the curve along Y-axis should be straight and have less bending towards X-axis that indicates higher TPR with minimum FPs/I which is desired from every system performance. The same can be observed with x8 augmentation having a TPR of 99.64% with 0.25 FPs/I



**Fig. 5** Performance of the introduced mass detection scheme in terms of FROC curve for INbreast dataset with different variations of data augmentation



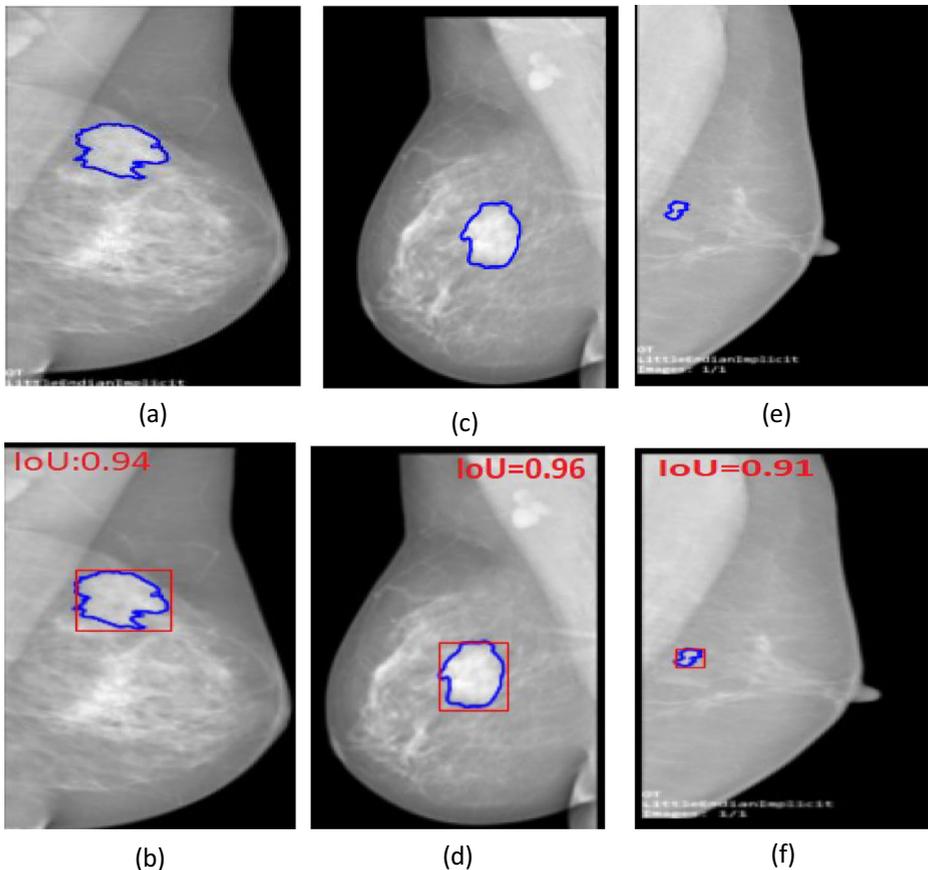
**Fig. 6** Performance of the introduced mass detection scheme in terms of FROC curve for DDSM dataset with different variations of data augmentation

for INbreast dataset and the same for DDSM are 97.36% and 0.38 as shown in Figs. 5 and 6, respectively. The outcome of the proposed model in detection over some sample mammograms is demonstrated in Figs. 7 and 8 for INbreast and DDSM datasets, respectively. Blue contour indicates the ground truth while the rounding box represents the identified mass region. The IoU mentioned over the output mammograms indicates the close fit of the rounding box over the mass lesions. Higher the value of IoU, the better is the case for detection and the same can be observed from Figs. 7 and 8 where IoU value observed is greater than 0.9. The proposed method is unable to detect only 4 masses out of 848 for INbreast dataset and 255 masses out of 8000 for DDSM dataset.

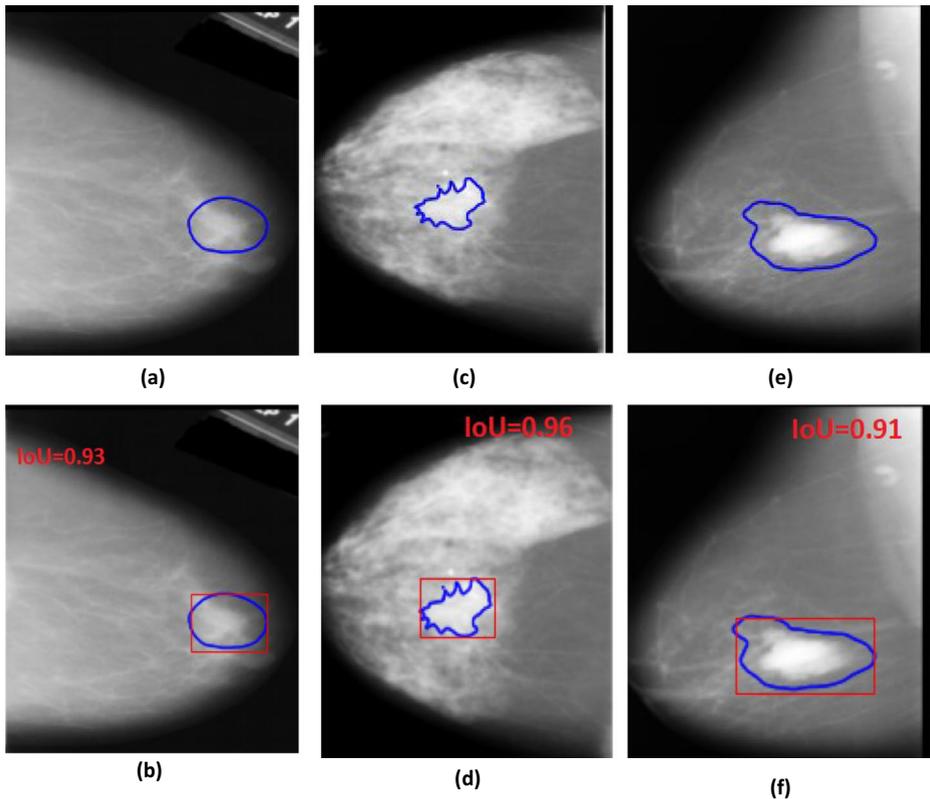
## 4.2 Mass classification

Since, the performance of the dataset augmented 8 times ( $\times 8$ ) gives the optimal tradeoff between the sensitivity and computational complexity hence the correctly identified mass cases for both the INbreast and the DDSM datasets are considered for the diagnosis. The total number correctly identified mass lesions for INbreast is 844 and for DDSM dataset is 7745 which are splitted in the ratio of 60:20:20 for training, validation, and testing respectively with 5-fold cross-validation technique. The obtained results in terms of confusion matrix, sensitivity, specificity, accuracy, AUC, F1-score, FPR, MCC and k for each fold are summarized in Tables 4 and 5, respectively. The proposed system provide better results over all 5-fold cross validations in all measurements as can be seen from Tables 4 and 5. In an average, 97.07% of malignant masses are diagnosed correctly with 97.98% for benign in case of INbreast dataset while the same for the DDSM are 97.67% and 94.69%, respectively. The overall sensitivity, specificity, accuracy, and AUC over INbreast mammograms

are 97.07%, 97.97%, 97.39%, and 0.97, respectively. The same respective parameters when evaluated on the DDSM datasets are 97.66%, 94.68%, 96.81%, and 0.96. In case of uneven class distribution, F1 Score is more critical parameter than accuracy. From Tables 4 and 5, it can be observed that F1 Score is higher for both the datasets than the accuracy i.e. 98% for INbreast dataset and 97.8% for DDSM dataset. A correlation of 94.37% for INbreast and 92.25% for DDSM dataset signifies that the predicted class and the true class are strongly correlated.  $k$  value obtained for INbreast and DDSM dataset are 0.94 and 0.92 shows the classification is almost perfect. The results of classification show the feasibility of the proposed system, minimizing the false positive and negative rates. A ROC curve of the same for both the datasets is depicted in Fig. 9. The area with the curve and the X-axis as the boundaries is called AUC. An AUC score of 0.97 for INbreast and 0.96 for DDSM dataset means that the model is able to successfully distinguish the benign and malignant masses with high sensitivity and low false positives.



**Fig. 7** Mass detection results on images of INbreast dataset with [a],[c] and [e] represents the ground truth whereas [b],[d] and [f] represent the detected mass. Blue contour represents the groundtruth and red box represent the detection of mass. It also shows the IoU value in each case



**Fig. 8** Mass detection results on images of DDSM dataset with [a],[c] and [e] represents the ground truth whereas [b],[d] and [f] represent the detected mass. Blue contour represents the ground truth and red box represent the detection of mass. It also shows the IoU value in each case

### 4.3 Statistical analysis and performance comparison

The analysis is extended further and a comparative study of the performance of the proposed model against some of the competitive schemes, Faster R-CNN and RetinaNet, is performed. The experimental set up, dataset, and performance evaluation parameters are kept same as done in case of proposed model. The performances in detection presented in Table 6 is the average of the all the values obtained from each fold of testing. The proposed U-Net based deep learning model undoubtedly performed better than other methods. Similar observation can be made from Table 7 where classification performance of the proposed U-Net model against Faster R-CNN and RetinaNet is provided. Again, the performance of the proposed approach is better against both the methodologies. Moreover, a various statistical analysis test are conducted such as t-test, Kolmogorov-Smirnov Predictive Accuracy (KSPA) test, Wilcoxon sign rank test [15], Friedman test [44] and Branin, Extended Beale test [26] etc. In this paper t-test is carried out to establish the superiority of the proposed approach. Smaller value of  $p$  indicates more confident about the null hypothesis being false. If  $p$  value is below significance level of 0.05 then we can reject the null. From Table 6,  $p$  values of 0.02, 0.04, and 0.04 for FPs/I, sensitivity, and precision can be observed when compared against RetinaNet while for Faster R-CNN, the values are 0.006, 0.04 and 0.05

**Table 4** Experimental results of proposed model for mass classification in INbreast dataset with 5-fold cross validation

Fold No.	Actual classes	Predicted Classes		Sen (%)	Spe (%)	Acc (%)	AUC	F1-score (%)	FPR	MCC	k
		Benign	Malignant								
1st	Benign	100	0	95.37	100	97.04	0.97	98	0	93.89	0.94
	Malignant	4.63	95.37								
2nd	Benign	98.18	1.82	98.24	98.18	98.22	0.98	99	0.02	95.98	0.96
	Malignant	1.75	98.25								
3rd	Benign	98.15	1.85	97.39	98.14	97.63	0.97	98	0.02	94.64	0.95
	Malignant	2.61	97.39								
4th	Benign	96.77	3.23	98.13	96.77	97.63	0.97	98	0.03	94.91	0.95
	Malignant	1.87	98.13								
5th	Benign	96.77	3.23	96.22	96.77	96.42	0.96	97	0.03	92.41	0.92
	Malignant	3.77	96.23								
Average	Benign	<b>97.98</b>	<b>2.02</b>	<b>97.07</b>	<b>97.97</b>	<b>97.39</b>	<b>0.97</b>	<b>98</b>	<b>0.02</b>	<b>94.37</b>	<b>0.94</b>
	Malignant	<b>2.93</b>	<b>97.07</b>								

Bold values indicate the average value of experimental results using 5-fold cross validation for proposed models in mass classification in INbreast dataset

which signifies that the proposed model is statistically significant. Similar analysis is conducted in case of classification and results are mentioned in Table 7 where, again, it can be observed that for both Faster R-CNN and RetinaNet, model probability of no difference is very small, so the null hypothesis can be rejected. Figure 10 shows the mean value obtained

**Table 5** Experimental results using 5-fold cross validation for proposed model in mass classification of DDSM dataset

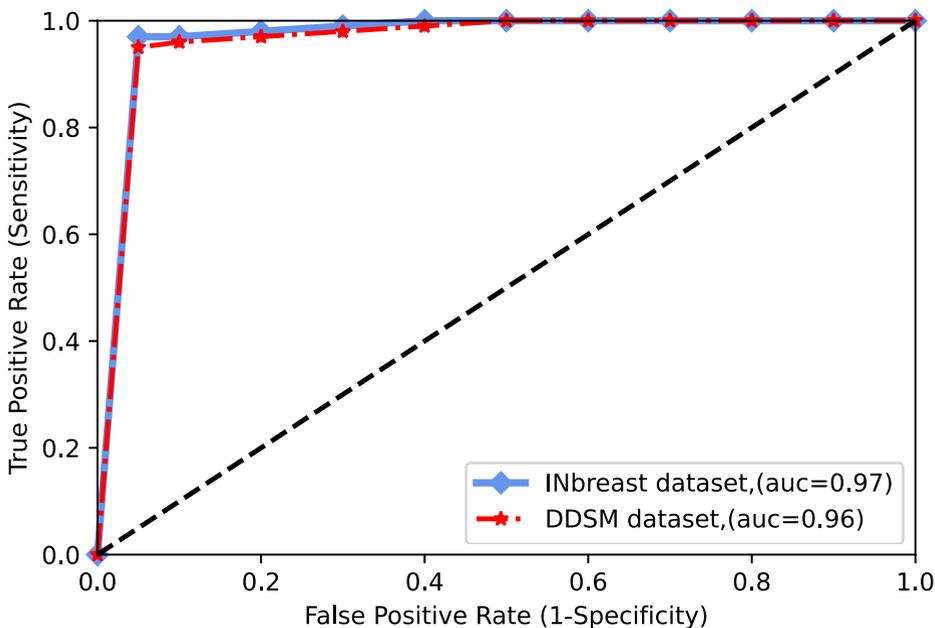
Fold No.	Actual classes	Predicted Classes		Sen (%)	Spe (%)	Acc (%)	AUC	F1-Score (%)	FPR	MCC	k
		Benign	Malignant								
1st	Benign	82.30	17.70	92.01	82.29	89.28	0.87	93	0.18	73.70	0.74
	Malignant	7.99	92.01								
2nd	Benign	97.59	2.41	98.53	97.59	98.25	0.98	99	0.02	95.83	0.96
	Malignant	1.47	98.53								
3rd	Benign	97.41	2.59	99.35	97.4	98.77	0.98	99	0.03	97.07	0.97
	Malignant	0.64	99.36								
4th	Benign	97.64	2.36	99.07	97.63	98.64	0.98	99	0.02	96.78	0.97
	Malignant	0.92	99.08								
5th	Benign	98.51	1.49	99.35	98.51	99.09	0.98	99	0.01	97.86	0.98
	Malignant	0.65	99.35								
Average	Benign	<b>94.69</b>	<b>5.31</b>	<b>97.66</b>	<b>94.68</b>	<b>96.81</b>	<b>0.96</b>	<b>97.8</b>	<b>0.05</b>	<b>92.25</b>	<b>0.92</b>
	Malignant	<b>2.33</b>	<b>97.67</b>								

Bold values indicate the average value of experimental results using 5-fold cross validation for proposed models in mass classification in DDSM dataset

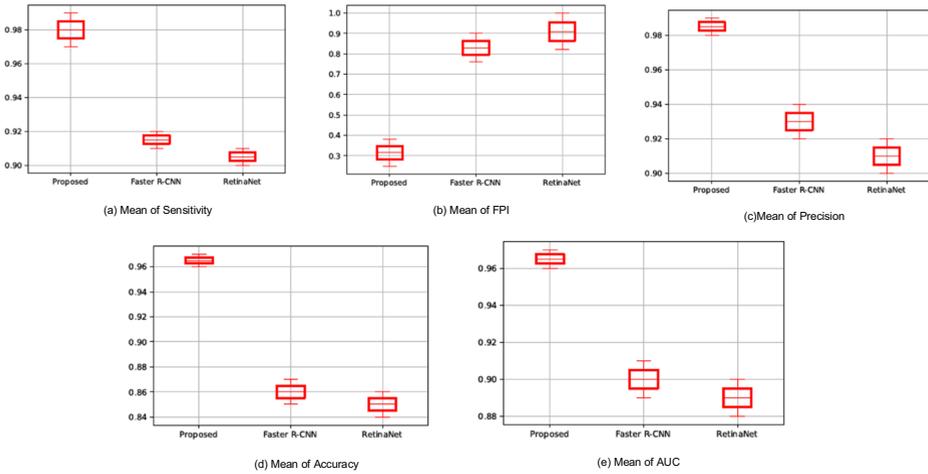
**Table 6** Comparison of detection performance of proposed system with other methods on the test set

Dataset	Method	Detection Parameter		
		Sensitivity	FPI	Precision
INbreast	Faster R-CNN	0.92	0.76	0.94
	Proposed	0.99	0.25	0.99
DDSM	Faster R-CNN	0.91	0.90	0.92
	Proposed	0.97	0.38	0.98
	p value	0.04	0.006	0.05
INbreast	RetinaNet	0.91	0.82	0.92
	Proposed	0.99	0.25	0.99
DDSM	RetinaNet	0.9	1	0.9
	Proposed	0.97	0.38	0.98
	p value	0.04	0.02	0.04

during testing with respect to both the competitive schemes where it can be observed that the mean value for each parameter of the proposed approach is better with respect to the other method.



**Fig. 9** ROC curve showing the performance of the proposed architecture in mass classification for the DDSM and INbreast dataset



**Fig. 10** Box plot showing a comparison of mean value of all parameters observed from the proposed approach against RetinaNet and Faster R-CNN model

### 5 Discussion

In this paper, a new automatic deep learning based U-Net architecture is designed for detection and classification of mammographic masses. The architecture is investigated on two publicly available datasets— INbreast and DDSM. The analysis includes the incorporation of data augmentation to increase the number of samples to train the models more efficiently. This resulted in enhanced performance of the proposed architecture. Not only in detection stage, the first time investigated U-Net architecture in classification has also demonstrated promising performance and achieved an accuracy of 97.39% with an AUC and F1-score of 0.97 and 98%, respectively for the INbreast dataset. For the DDSM dataset, accuracy of 96.81% is measured with AUC of 0.96 and F1-score of 97.8%. The best results are compared

**Table 7** Comparison of classification performance of proposed system with other methods on the test set

Dataset	Method	Classification parameter	
		Accuracy	AUC
INbreast	Faster R-CNN	0.87	0.91
	Proposed	0.97	0.97
DDSM	Faster R-CNN	0.85	0.89
	Proposed	0.96	0.96
	p value	0.03	0.04
INbreast	RetinaNet	0.86	0.9
	Proposed	0.97	0.97
DDSM	RetinaNet	0.84	0.88
	Proposed	0.96	0.96
	p value	0.02	0.04

**Table 8** Comparison of proposed system with some of the existing methods for DDSM dataset

Reference	Detection Parameter	Classification	Parameter
	TPR at FPI	Acc(%)	AUC
Dhungel et al. [13]	0.75 at 4.8	–	–
Dhungel et al. [14]	0.95 at 5	0.91	0.76
Casti et al. [10]	0.96 at 1.83	–	–
Chakraborty et al. [11]	0.9 at 1.4	83.30%	0.92
Agarwal et al. [2]	–	84.16%	0.93
Proposed method	<b>0.97 at 0.38</b>	<b>96.81%</b>	<b>0.96</b>

Bold values indicate the proposed approach performs better than some of the competing schemes for DDSM dataset

with the results reported in the state-of-the-art for detection and diagnosis of masses as mentioned in Tables 8 and 9, respectively. Moreover, the performance of the proposed approach is further compared with the other competing schemes— Faster R-CNN and RetinaNet where the efficiency of these methodologies are measures keeping the same experimental set up, dataset, and parameters as in the case of proposed approach and the outcomes are provided in Tables 6 and 7 for detection and diagnosis of mass cases, respectively over both the datasets. It can be observed from Tables 6 and 7 that the proposed approach takes an edge over methods. Also, the statistical analysis conducted establishes the supremacy of the proposed approach. Although the efficiency of the proposed framework is satisfactory, but still in high density regions where the areas in and around the mass cases are difficult to differentiate and in case of tiny lesions, more emphasis should be given. This may be overcome by incorporating data augmentation by higher folds.

## 6 Conclusion

In this paper, a deep learning based integrated system for automatic detection and diagnosis of mammographic masses is proposed. The deep learning network incorporates U-Net

**Table 9** Comparison of proposed approach with some of the existing methods for INbreast dataset

Reference	Detection Parameter	Classification	Parameter
	TPR at FPI	Acc(%)	AUC
Dhungel et al. [13]	0.96 at 1.2	–	–
Agarwal et al. [2]	0.98 at 1.67	–	–
Ribli et al. [36]	0.9 at 0.3	–	0.95
Al.Antari et al. [4]	0.98	95.64%	0.94
Agarwal et al. [1]	0.95 at 1.14	–	–
Proposed method	<b>0.99 at 0.25</b>	<b>97.39%</b>	<b>0.97</b>

Bold values indicate the proposed approach performs better than some of the competing schemes for INbreast dataset

architecture for the detection while the same is examined for the first time in classification. The evaluation is conducted on two different standard databases– INbreast and DDSM. The assessment includes data augmentation and studies its significance in model training and efficiency of the system. With increase in the number of the samples for training and testing due to augmentation, the performance has also improved which outperforms some of the recent competitive schemes in the state-of-the-art. Although the performance is satisfactory, but there is still a scope of improvement by the minimizing the FPs/I and improving the sensitivity especially in case of DDSM database. Hence, Future work may incorporate deeper architecture to detect the missed cases and also can be investigated to perform multiclass classification of normal, benign, and malignant masses.

## Declarations

**Conflict of Interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Agarwal R, Díaz O, Yap MH, Lladó X, Martí R (2020) Deep learning for mass detection in full field digital mammograms. *Comput Biol Med* 121:103774. <https://doi.org/10.1016/j.compbiomed.2020.103774>
2. Agarwal R, Diaz O, Lladó X, Yap MH, Martí R (2019) Automatic mass detection in mammograms using deep convolutional neural networks. *J Med Imag* 6(3):1–9. <https://doi.org/10.1117/1.JMI.6.3.031409>
3. Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, Barkan E (2016) A region based convolutional network for tumor detection and classification in breast mammography. In: Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, Papa JP, Nascimento JC, Loog M, Lu Z, Cardoso JS, Cornebise J (eds) *Deep learning and data labeling for medical applications*. Springer International Publishing, Cham, pp 197–205
4. Al-antari MA, Al-masni M, Choi M-T, Han S-M, Kim T-S (2018) A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Informatics*, 117. <https://doi.org/10.1016/j.ijmedinf.2018.06.003>
5. Al-masni MA, Al-antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, Valarezo E, Choi M-T, Han S-M, Kim T-S (2018) Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. *Comput Methods Programs Biomed* 157:85–94. <https://doi.org/10.1016/j.cmpb.2018.01.017>
6. Alanazi SA, Kamruzzaman MM, Islam Sarker MN, Alruwaili M, Alhwaiti Y, Alshammari N, Siddiqi MH (2021) Boosting breast cancer detection using convolutional neural network. *J Healthcare Eng* 2021:5528622
7. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Guevara Lopez MA (2016) Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed* 127:248–257. <https://doi.org/10.1016/j.cmpb.2015.12.014>
8. Astley SM, Gilbert FJ (2004) Computer-aided detection in mammography. *Clin Radiol* 59(5):390–399. <https://doi.org/https://doi.org/10.1016/j.crad.2003.11.017>
9. Behera SK, Rath AK, Sethy PK (2021) Fruits yield estimation using faster r-cnn with miou. *Multimed Tools Applic* 80(12):19043–19056
10. Casti P, Mencattini A, Salmeri M, Ancona A, Mangieri F, Rangayyan RM (2014) Development and validation of a fully automated system for detection and diagnosis of mammographic lesions. In: 2014 36th Annual international conference of the ieee engineering in medicine and biology society, pp 4667–4670. <https://doi.org/10.1109/EMBC.2014.6944665>
11. Chakraborty J, Midya A, Rabidas R (2018) Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. *Expert Syst Appl* 99:168–179. <https://doi.org/10.1016/j.eswa.2018.01.010>
12. Chicco D, Warrens MJ, Jurman G (2021) The matthews correlation coefficient mcc is more informative than cohen’s kappa and brier score in binary classification assessment. *IEEE Access* 9:78368–78381

13. Dhungel N, Carneiro G, Bradley AP (2015) Automated mass detection in mammograms using cascaded deep learning and random forests. In: 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp 1–8, <https://doi.org/10.1109/DICTA.2015.7371234>
14. Dhungel N, Carneiro G, Bradley AP (2017) A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Anal* 37:114–128. <https://doi.org/10.1016/j.media.2017.01.009>
15. Fan G-F, Yu M, Dong S-Q, Yeh Y-H, Hong W-C (2021) Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Utilities Policy* 73:101294. <https://doi.org/10.1016/j.jup.2021.101294>
16. Ferlay J, Soerjomataram I, Dikshit R, Eser S (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int J Cancer* 5:E359–E386
17. Greenspan H, van Ginneken B, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging* 35(5):1153–1159. <https://doi.org/10.1109/TMI.2016.2553401>
18. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R (2018) Deep learning in mammography and breast histology, an overview and future trends. *Med Image Anal* 47:45–67. <https://doi.org/10.1016/j.media.2018.03.006>
19. Hammad M, Alkinani MH, Gupta BB, Abd El-Latif AA (2021) Myocardial infarction detection based on deep neural network on imbalanced data. *Multimedia Syst*
20. Hammad M, Iliyasu AM, Subasi A, Ho ESL, El-Latif AAA (2021) A multitier deep learning model for arrhythmia detection. *IEEE Trans Instrum Meas* 70:1–9
21. Hassan SA, Sayed MS, Abdalla MI, Rashwan MA (2020) Breast cancer masses classification using deep convolutional neural networks and transfer learning. *Multimed Tools Applic* 79(41):30735–30768
22. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer P (2000) The digital database for screening mammography. In: Proceedings of the fourth international workshop on digital mammography. <https://doi.org/10.1007/978-94-011-5318-875>
23. Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N (2017) Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* 35:303–312. <https://doi.org/10.1016/j.media.2016.07.007>
24. Laishram R, Rabidas R (2020) Detection of mammographic masses using frfcm optimized by pso. In: 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp 327–332, <https://doi.org/10.1109/CISP-BMEI51763.2020.9263644>
25. Laishram R, Rabidas R (2021) Wdo optimized detection for mammographic masses and its diagnosis: a unified cad system. *Appl Soft Comput* 110:107620. <https://doi.org/10.1016/j.asoc.2021.107620>
26. Li M-W, Wang Y-T, Geng J, Hong W-C (2021) Chaos cloud quantum bat hybrid optimization algorithm. *Nonlinear Dyn*, 103. <https://doi.org/10.1007/s11071-020-06111-6>
27. Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2018) Focal loss for dense object detection. *arXiv:1708.02002*
28. Mahmood T, Li J, Pei Y, Akhtar F, Jia Y, Khand ZH (2021) Breast mass detection and classification using deep convolutional neural networks for radiologist diagnosis assistance. In: 2021 IEEE 45th annual computers, software, and applications conference, pp 1918–1923
29. Metz CE (2006) Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol* 3(6):413–422. Special Issue: Image Perception
30. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) Inbreast: toward a full-field digital mammographic database. *Acad Radiol* 19(2):236–248. <https://doi.org/10.1016/j.acra.2011.09.014>
31. O’Shea K, Nash R (2015) An introduction to convolutional neural networks
32. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning
33. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, Conant EF, Fajardo LL, Bassett L, D’Orsi C, Jong R, Rebner M (2005) Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 353(17):1773–1783. <https://doi.org/10.1056/NEJMoa052911>. PMID: 16169887
34. Qiu Y, Yan S, Gundreddy RR, Wang Y, Cheng S, Liu H, Zheng B (2017) A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. *J Xray Sci Technol* 25:751–763
35. Ren S, He K, Girshick R, Sun J (2016) Faster r-cnn: towards real-time object detection with region proposal networks. *arXiv:1506.01497*

36. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I (2018) Detecting and classifying lesions in mammograms with deep learning. *Sci Rep* 8(1):4165. <https://doi.org/10.1038/s41598-018-22437-z>
37. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation
38. Sarkar PR, Prabhakar P, Mishra D, Subrahmanyam G (2019) Towards automated breast mass classification using deep learning framework. In: 2019 IEEE international conference on data science and advanced analytics, pp 453–462
39. Sedik A, Hammad M, Abd El-Samie FE, Gupta BB, Abd El-Latif AA (2021) Efficient deep learning approach for augmented detection of coronavirus disease. *Neural Comput Appl*
40. Sethy PK, Pandey C, Khan MR, Behera SK, Vijaykumar K, Panigrahi SS (2021) A cost-effective computer-vision based breast cancer diagnosis. *J Intell Fuzzy Syst* 41:5253–5263
41. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. *CA Cancer J Clin* 69(1):7–34. <https://doi.org/10.3322/caac.21551>
42. Taylor L, Nitschke G (2017) Improving deep learning using generic data augmentation
43. Yang WT, Lai CJ, Whitman GJ, Murphy Jr WA, Dryden MJ, Kushwaha AC, Sahin AA, Johnston D, Dempsey PJ, Shaw CC (2006) Comparison of full-field digital mammography and screen-film mammography for detection and characterization of simulated small masses. *AJR Am J Roentgenol* 187(6):W576–W581. <https://doi.org/10.2214/AJR.05.0126>. 17114508[pmid]
44. Zhang Z, Hong W-C (2021) Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads. *Know-Based Syst* 228:C. <https://doi.org/10.1016/j.knosys.2021.107297>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.