

Engineering Research Express



PAPER

An integrated computer-aided diagnosis BCanD model for detection, segmentation and classification of breast cancer

Mehar Arfi^{1,*} , Subhash Chandra Yadav¹ and Suman Lata Tripathi²

¹ Department of Computer Science & Engineering, Central University of Jharkhand, Ranchi, India

² School of Electronics and Electrical Engineering, Lovely Professional University, Jalandhar, India

* Author to whom any correspondence should be addressed.

E-mail: mehar@cuja.ac.in, dr.scyadav@cuja.ac.in and tri.suman78@gmail.com

Keywords: digital mammograms, breast mass, detection, segmentation, classification, benign, malignant

Abstract

A fully integrated Computer-Aided Diagnosis (CAD) system involves the integration of detection, segmentation, and classification, which makes it very useful for medical applications, particularly while dealing with the detection of breast mass and its classification into malignant and benign. The carried-out research work is intended to propose a Breast Cancer Detection (BCanD) model that is an integrated CAD system, where the system is capable enough for mass detection, its segmentation, and for the classification using mammograms. The proposed integrated system utilizes deep learning based YOLO model to detect the abnormality (mass) in the mammogram, where U-net is used for segmentation of the mass, as it has the capability to produce pixel level segmentation map, and at last stage that is the classification stage deep CNN is used for the classification. The proposed system is evaluated on open-source MIAS database. For the performance evaluation of the proposed BCanD, a three-fold cross-validation test was utilized. The mass detection accuracy of the BCanD is 98.99%, MCC is 97.96%, and F1-score is 98.87%. The model is evaluated with and without automated mass segmentation to study the impact of segmentation on the suggested CAD system. The best results was observed with the segmentation with the overall accuracy of 94.20%, F1-score (Dice) of 93.60%, MCC of 88.33%, and Jaccard of 88.08%. The proposed BCanD model surpasses the latest existing deep learning-based methodologies like fuzzy classifier, CNNI-BCC etc Hence, the proposed CAD system can be implemented and used by radiologists for all the stages from detection to diagnosis of breast mass.

1. Introduction

Breast cancer represents a highly aggressive and rapidly proliferating type of tumor that predominantly affects women, though it is exceedingly rare in men [1], which is also one of the deadliest cancer after lung cancer, causing a very high death rate in women [2]. As per the report of National Institute of Cancer Prevention and Research (NICPR) of India [3], for every new two women diagnosed with breast cancer, only one of them survives. Last five year's studies reveals that early diagnosis of breast cancer results in increment in survival rate to 90% and above [4]. But, late diagnosis can make it even worse and it can propagate to lymph nodes and other body organs and thus dropping the survival rate by 27% [5]. Therefore, the detection of breast cancer at an early stage is very crucial for saving women's lives. The structure of the breast is complex and sometimes it gets quite difficult to identify local lesions. For the diagnosis of breast cancer medical screenings are required. During the COVID period we all have witnessed how overworked the doctors and other healthcare providers were. So, it is necessary to reduce their workloads. Reducing the time consumed in the screening process and increasing the accuracy of the computer-aided diagnosis (CAD) techniques could help in reducing the workload. Recent development in technology has resulted in innovations in CAD. One of the most effective, less harmful, cost-effective, and reliable sources for the breast cancer detection at earliest stage is mammography [6]. Digital

mammography separates the image acquisition, presentation, and storage processes, allowing each to be optimized. An electronic detector absorbs radiation delivered through the breast, and its reaction is consistent over a broad range of intensities [7]. The captured information can be shown using image processing techniques that provides arbitrary adjustments of brightness and contrast of image, without further exposing the patient to radiation [8]. Mammography can reveal abnormalities in the breast such as calcifications, tumors, cysts, and other indications [9]. The abnormalities like mass, density, nodule or any distortion could may or may not be cancerous [10]. For verifying the mass is cancerous or not, biopsy is needed. Radiologists are required to evaluate mammography pictures in order to diagnose breast cancer. The opinion of radiologists may get inconsistent and suffer from human error. Therefore, a reliable model with minimal human interventions for breast cancer detection is needed.

In recent years, numerous researchers have dedicated efforts to detect breast cancer through mammography images, developing a variety of algorithms aimed at enhancing the precision of classification. As per the data of American Cancer Society [11] out of eight, one of the mammograms miss the detection of breast cancer. Hence, the false negative leads to false sense of security and it may result in late detection and sometimes death. Therefore, this study aims to design and propose a model that will be capable of detecting, segmenting and classifying breast masses using mammogram images with high accuracy and least human interference. The major focus of this study is to propose an integrated computer-aided. Breast Cancer Detection model, which is capable of detection of breast mass, segmentation of mass and its classification into benign (non-cancerous) and malignant (cancerous).

Deep learning-based techniques have demonstrated exceptional performance in the medical image analysis [12–14]. Latest studies have shown that mass detection is quite challenging, especially in case of breast tissues [12, 15, 16]. In this research work, an integrated deep learning based fully automatic CAD system is proposed for the detection, segmentation and classification of mass into benign and malignant. The proposed model is implemented using Mammographic Image Analysis Society (MIAS) dataset, which is digitized mammogram screening films [17], for countries like India, and other South-Asian countries, to date mammogram films are used on a large scale. Therefore, for the proposed integrated CAD model, MIAS dataset is used. The paper is organized as follows: first the mass detection stage is discussed, secondly mass segmentation stage, and then mass classification stage. For each stage the performance is compared with the existing models. The proposed integrated CAD model is validated and compared with the latest existing models on the publically available MIAS dataset.

2. Related work

The ability of a Computer-Aided Diagnosis (CAD) system to better diagnose breast cancer can be increased by using the more general, deep learning features that can better describe the features of mass [15, 16, 18]. The identification of masses is a crucial step in computer-aided diagnosis (CAD) systems [19] used for the detection of breast cancer [15, 20, 21]. There is further work to be done to adequately overcome this challenging problem [16, 20, 21]. Typically, CAD systems use manual mass identification and deep CNN for classifying masses into cancerous (malignant) or non-cancerous (benign) [22, 23]. In various systems, to bring about the core features, the manually identified data are then put into CNN [18, 22, 24]. Other CAD systems have employed a multi-layered CNN to extract high-level deep features [25–28]. A specialized CNN architecture was proposed [29] for synovial fluid detection in human knee joint where the concept of transfer learning was used. The experiment was conducted using MR images and resulted in high accuracy. Subsequently, a classifier is employed to differentiate benign from malignant tissues based on the combination of these features [23]. The majority of these CAD systems demonstrated better classification accuracy compared to typical machine learning techniques that rely on manually generated features [12, 30]. Nevertheless, the task of automatically detecting masses continues to pose a major challenge. Various researchers have focused on the automated detection of breast abnormalities [12, 16, 18, 31, 32]. In a study [33] a specialized DCNN model was designed for multi-class skin cancer classification using ISIC-17, ISIC-18, and ISIC-19 dataset. Preprocessing and data augmentation such as flip, rotation and translation have been applied to solve the problem of overfitting and increase the size of dataset. The proposed DCNN achieved an overall accuracy of 92.83% for ISIC-17 and similar for ISIC-18, and 90% for ISIC-19. Currently, there are only a few of deep learning studies that propose automated methods for detecting masses in computer-aided diagnosis (CAD) systems [16, 34, 35]. The initial findings of mass detection are demonstrated by employing the deep learning based YOLO with the DDSM dataset [36]. YOLO exhibited better performance in detecting compared to other previously reported detection techniques Dhungel *et al* [16, 34]. A Computer-Aided Diagnosis (CAD), featuring a deep belief network (DBN), was developed to examine abnormal regions in mammogram images [15]. For the purpose of mass detection, this study employed adaptive thresholding and morphological processes, with an average detection accuracy of 86% [15]. A novel

deep learning model, known as R-CNN, was proposed for the automated detection of breast cancer masses [20]. The model divided the mammogram into multiple patches and trained R-CNN to categorize identified regions as cancerous and non-cancerous. A different automated technique employed a series of deep learning models for mass detection. This strategy consisted of four main stages, using a multiscale deep belief network (mDBN) and a Gaussian mixture classifier (GMC) to identify areas of concern. To further lower the false positive rate, a two-stage cascade method with a conditional random forest (CRF) classifier was applied to the regions remaining after a two-level cascade of R-CNN had been utilized to decrease false positives. For post-processing, the method combined the areas that passed through previous filters using connected component analysis (CCA). Additionally, an enhancement algorithm was introduced to refine mass detection accuracy, comprising Bayesian optimization and a deep R-CNN-based classifier structure, aimed at improving the identified regions' scale and precision. A novel BTS-ST model combining U-net and Swim Transformer was introduced for the segmentation and classification of breast cancer [37]. CBIS-DDSM and MRI-based dataset were used for the performance validation. R2U-Net stands for Recurrent Residual Convolutional Neural Network and it is an extension of U-Net [38]. In several studies, R2U-Net has outperformed U-Net in skin cancer segmentation, lung segmentation, brain tumor segmentation, and many others [38, 39]. However, the application of R2U-net is limited to large-scale datasets, also the residual network can be data-hungry, hence for smaller datasets, U-Net is still a better option. BRAU-Net++ a U-net based architecture, has also performed well in the segmentation of medical images such as skin lesions, and colonoscopies [40]. Hossain *et al* proposed an automated method that utilizes Fuzzy C-means clustering and U-net for segmentation of micro-calcification region [41]. The proposed method performed well with the DDSM dataset with an average accuracy of 98.2%. For the segmentation of skin lesions, a cascade knowledge diffusion network (CKDNet) was for segmentation and classification, and the performance of the network was tested on the ISIC2017 skin lesion dataset [42]. In recent years, DL techniques based on CNN have gained attention due to the automatic segmentation of medical images [26, 41]. Some of the popularly used segmentation techniques based on DL are Mask R-CNN, U-net, and PSP Net [26, 40].

In recent times, there has been a notable upsurge in the realm of research that is devoted to the partitioning of the pectoral muscle through the utilization of computer algorithms. Shen *et al* [43] introduced an algorithm that is designed to automate the segmentation of the pectoral muscle, employing a combination of selection algorithms based on morphology and a genetic algorithm. Through validation on the MIAS database, the authors discovered that the algorithm exhibits a false positive (FP) rate of 2.03% and a false negative (FN) rate of 6.90%. Another significant contribution by Dhungel *et al* [34] involved the development of an algorithm that is designed to segment the boundary between the breast region and breast parenchyma, utilizing an adaptive gamma correction method. This particular approach demonstrated an impressive accuracy of 98.45% and a Jaccard similarity index of 92.79%. Ramesh *et al* [44] proposed a deep learning architecture using MIAS dataset with dice and Jaccard coefficient scores of 82.15% and 89.11% respectively, but the data augmentation was not done, hence the problem of overfitting was not taken into consideration. Recent advancements in hardware technology have paved the way for the application of deep learning techniques, which have proven to be highly valuable across a wide range of domains. Convolutional Neural Networks (CNNs), which represent a specific class of deep learning techniques, have particularly excelled in various imaging applications [25, 45, 46]. When it comes to medical imaging, a multitude of studies that have implemented CNNs have consistently reported superior performance in comparison to traditional image processing techniques [16, 47, 48]. This prevailing trend signifies the escalating influence of deep learning, specifically CNNs, in enhancing the accuracy and efficiency of segmentation in medical image analysis, encompassing the delineation of the pectoral muscle and breast regions.

3. Methodology

A fully automatic and integrated into a single framework CAD system is proposed, and it is named as 'BCanD', using the MIAS dataset for detection, segmentation and classification of masses into benign and malignant. The two terms CAD & BCanD are used in this study interchangeably and represents the same model which is proposed in this study where the model consists of three components: (i) Mass Detection, (ii) Mass Segmentation, and (iii) Mass Classification. Figure 1 illustrates the design of the proposed Computer-Aided Diagnosis (CAD) system, which employs digital mammograms for the identification, segmentation, and classification of breast cancer masses.

3.1. Dataset

The study utilizes MIAS (Mammographic Image Analysis Society) [17] dataset (link: <http://peipa.essex.ac.uk/pix/mias/>) of digital mammography is used for the training, testing and evaluation of proposed CAD system. The MIAS dataset [17] are digital mammography image and was created in 1994. The dataset comprised 322

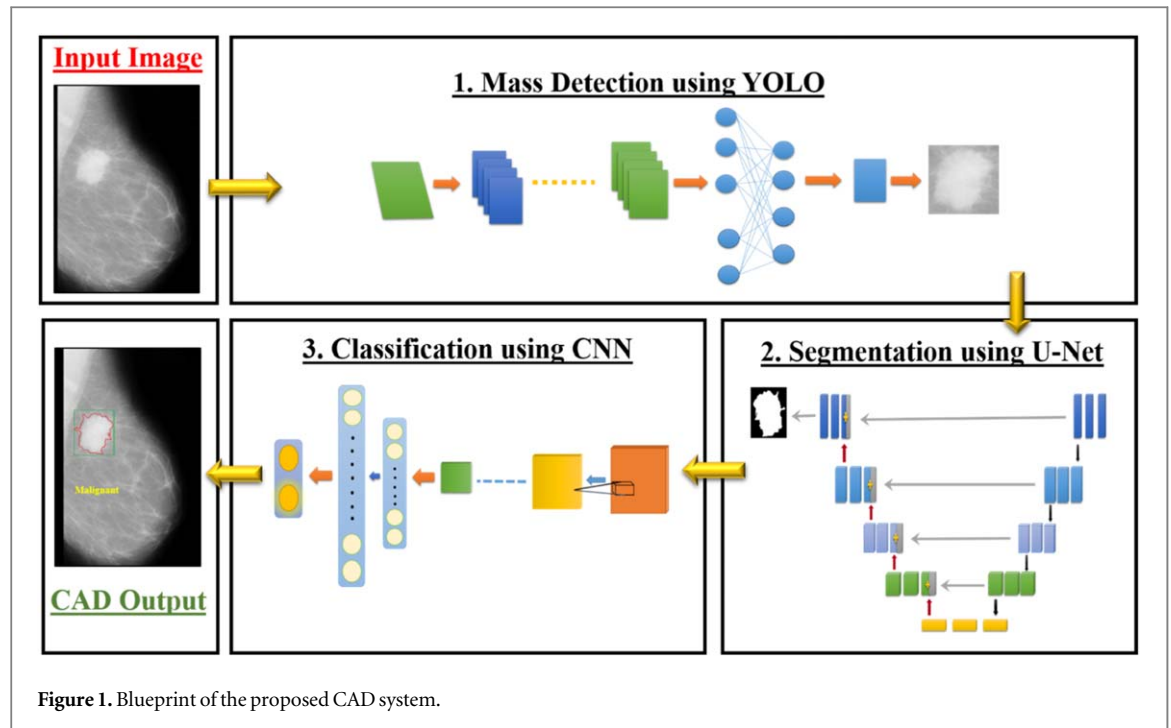


Figure 1. Blueprint of the proposed CAD system.

mammography images, each with dimensions of 1024×1024 . The images are sourced from the United Kingdom National Breast Screening Program (UKNBSP). The resolution of these dataset images are low. The format of the mammography image is in '.pgm' which stands for Pragmatic General Multicast. The dataset has a total of 322 mammography images, with 207 classified as normal and 115 as abnormal. There are two types of abnormal images: where the total number of benign images are 63, and similarly malignant images are 52. For each abnormal image the ground-truth annotations are present. Therefore, a total number of 115 abnormal cases are considered in this work. Some images from the MIAS dataset are shown in figure 2 where each image is marked with the letters 'M' and 'B'. The 'M' denotes malignant and 'B' denotes benign tumor.

3.2. Data augmentation and transfer learning

The MIAS dataset which is one of the smallest breast cancer datasets available, but it is widely used for evaluation of the model, therefore it is used in this study. But, for the training of deep learning-based models large data size is required. The smaller size data is the major problem with the medical dataset. There are two solutions available to cope up with the small size of the dataset. First method is augmentation of data. The term 'augment' means to enlarge, augmentation is the process of enlarging. Data augmentation is one of the commonly used method to increase the size of dataset and avoid overfitting process [15, 33]. In this study, an augmentation ratio of 1:8 is applied, indicating that each image undergoes augmentation eight times through rotation at an angle of $\Delta\Theta = 45$ degrees (i.e., rotation of image eight times at each of $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$ and 315°). After augmentation there are a total of 920 images (115 multiplied by 8). Post augmentation the number of benign cases are 504 and malignant cases are 416. Further, for the model initialization, transfer learning. It involves initially training the deep learning models on large annotated datasets like ImageNet. Subsequently, these models are refined and fine-tuned with augmented annotated datasets, for example, mammograms. In this study transfer learning is used for the parameter initialization of the deep learning models. Many other researchers have used transfer learning for the analysis of breast cancer using CAD system as it makes knowledge transfer between source and target domains possible [29, 49].

3.3. Mass detection using YOLO

The initial and crucial phase in the proposed Computer-Aided Diagnosis (CAD) system is to identify the breast mass. The detection of mass is not an easy task as the shape, size varies a lot and the SNR (Signal to Noise Ratio) of the masses is quite low from its surrounding breast tissue. This study utilizes a deep learning model known as YOLO (stands for You Only Look Once) for the mass detection. YOLO is an object detection algorithm that uses CNN technique for the detection of objects. Earlier studies have suggested that it is significant for the detection purpose [36, 50]. It is capable of detecting the suspicious regions of masses from the mammogram [36, 51]. A study on DDSM and INBreast dataset has shown promising results in the detection of breast mass and bounding box generation [36, 52]. There is a lack of literature available with MIAS dataset as it is digitized film based

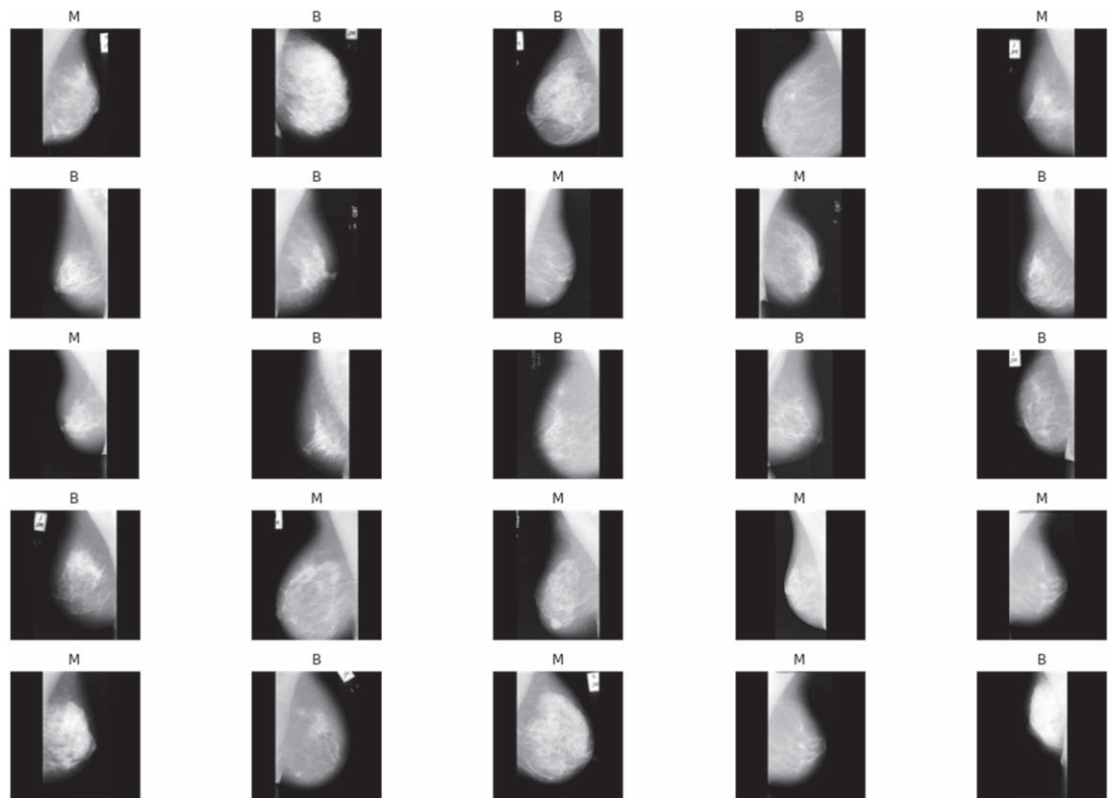


Figure 2. Example of the mammogram images from Mini-MIAS dataset.

dataset. Due to the better performance of YOLOv4 in mass detection as compared to all the previous versions it has been used in this study [51]. It requires very less memory and its running time is very low in contrast to other deep learning models. Originally the images are of 1024×1024 , so all the images are resized into 448×448 as done in earlier work [36, 52]. Here, the criteria for selection of correctly detected mass is the intersection over union ($\text{IOU}_{\text{Ground Truth}}^{\text{Predicted}}$) between the detected (predicted) and annotated (ground truth) bounding boxes is at least 50% [15, 16, 18, 36, 52, 53]. Also, the false detections are manually removed, as observed in the previous studies [12, 15, 52]. This removal is necessary as there is a lack of ground truth (G T) for the falsely detected masses, and due to that, it cannot be considered for performance evaluation [15, 52].

Figure 3 displays the results of applying YOLO for mass detection on the MIAS dataset's test images. The identified ROI (blue) and ground truth (red) images are placed over the original images shown in figure 3.

3.4. Mass segmentation using U-Net

Once the masses have been detected by the CAD system's previous detection stage, they enter directly into the segmentation stage. The Contrast-Limited Adaptive Histogram Equalization CLAHE is utilized for preprocessing all recognized masses. The CLAHE algorithm enhances visual contrast by dividing the image into multiple parts and applying histogram equalization to each region individually [12, 16, 34]. This approach improves picture contrast and contrast between masses and surrounding tissues. Previous studies have demonstrated that pixel-to-pixel segmentation outperforms traditional segmentation. In this study U-net is used for segmentation. U-net outperforms conventional sliding window methods by using fewer images and data augmentation to increase model performance. The U-Net is named based on its structure. Two networks and convolutional layers form the 'U' model. The encoder comes first, followed by the decoder. Encoder network is contracting network. This network learns the input image's feature map to detect objects. Unlike convolutional neural networks, U-Nets don't have fully connected layers at the end since the output needed is a mask that has the identical dimensions as the input image.

This encoder network has 4 blocks. Each block consists of two convolutional layers with a 3×3 kernel size, valid padding, and a Rectified Linear Unit (ReLU) activation function. This is fed into a max pooling layer with a 2×2 kernel size. The max pooling layer halves spatial dimensions, simplifying model training costs. Between the encoder and decoder networks, a bottleneck layer is situated. The model's architecture depicts this as the lowest layer as shown in figure 4. There are two convolutional layers in it, followed by ReLU. The final feature map

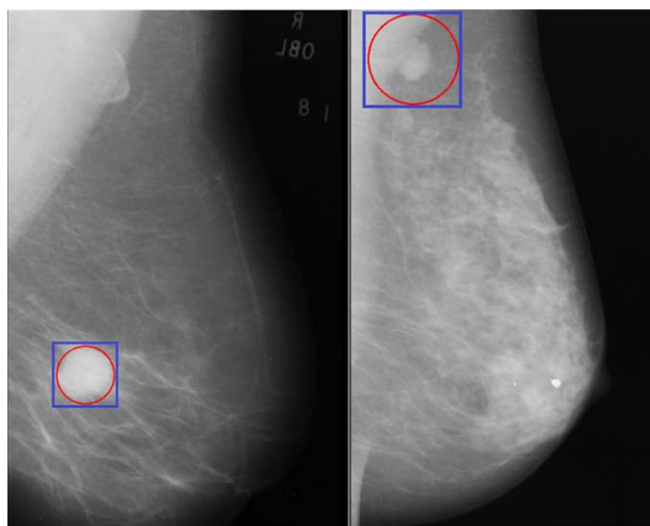


Figure 3. Mass detection using YOLO showing detected mass region (blue) and ground truth (red).

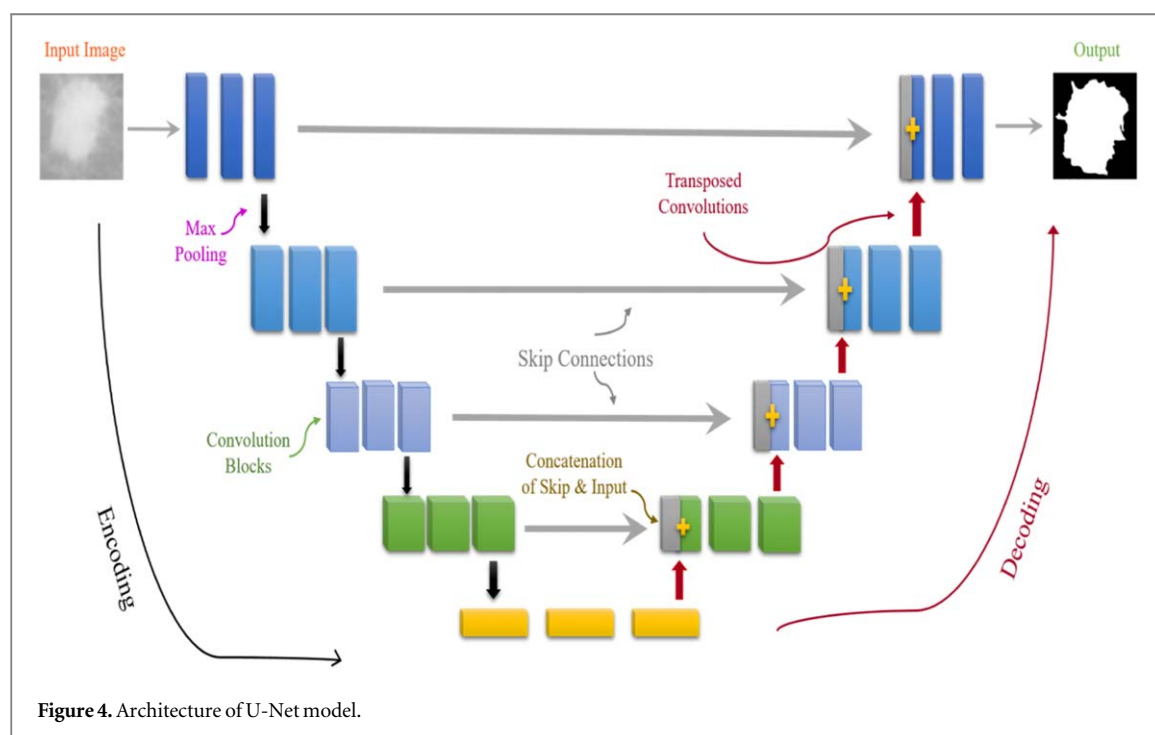


Figure 4. Architecture of U-Net model.

representation is the output of the bottleneck. The U-net exhibits enhanced image segmentation capabilities due to its integration of skip connections and decoder networks.

The detected mass is segmented using different segmentation techniques and are presented in figure 5. The figure depicts the segmentation results of Fully Convolutional Network (FCN), and Segmented Neural Network (SegNet), and U-Net with their Dice index and overall accuracy for three sample images. It also shows the segmentation results of FCN, SegNet and U-Net with their Dice index and overall accuracy for three sample images.

3.5. Mass classification using CNN

Segmented masses are resized to a size of 40×40 by using bicubic interpolation technique [12, 16, 23, 27, 52, 54, 55]. The resized masses are then fed into the final stage of the proposed CAD system i.e., classification. Deep CNN based AlexNet [54] is used for the classification purpose. CNN has the ability to extract deep features from the input images [25]. CNN has two FC (Fully Connected) layers and five convolutional layers for the classification model. First and second layers, contains 20 and 64 filters with size 5×5 each, whereas

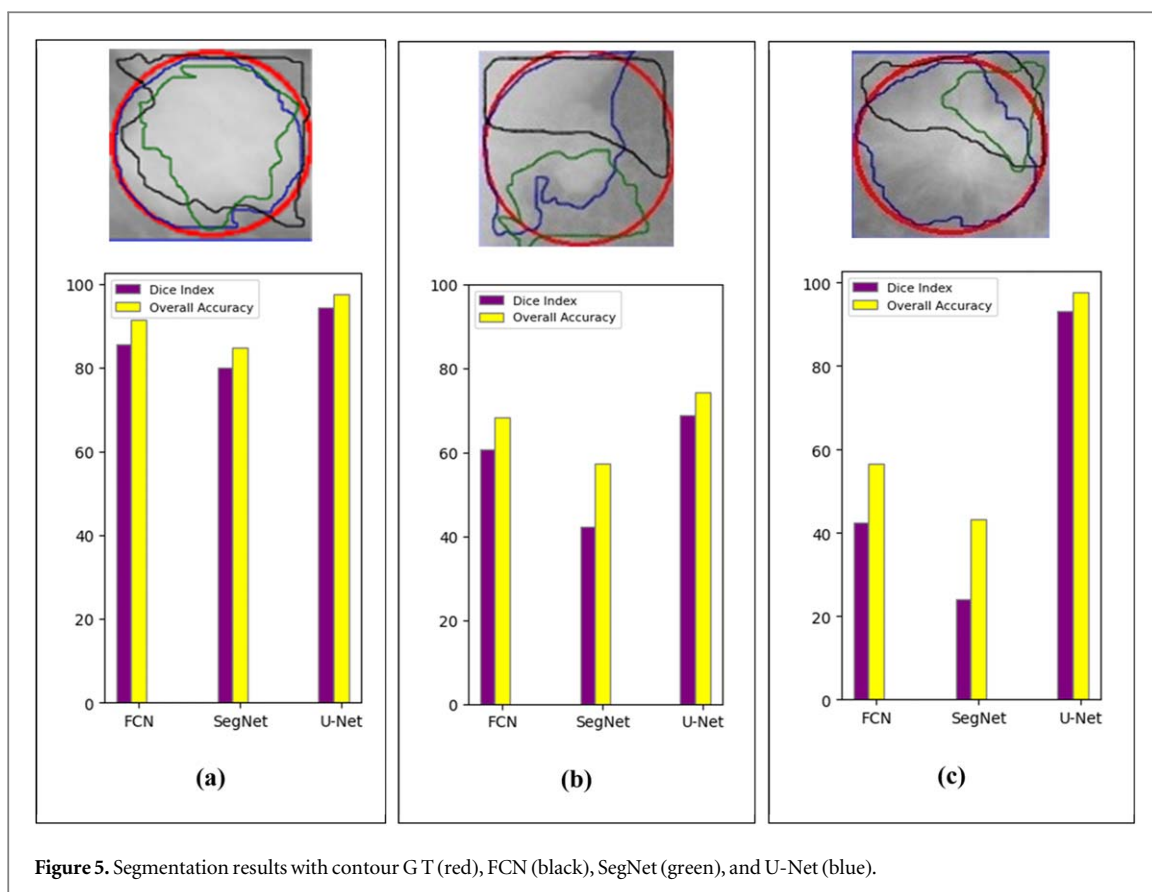


Figure 5. Segmentation results with contour G T (red), FCN (black), SegNet (green), and U-Net (blue).

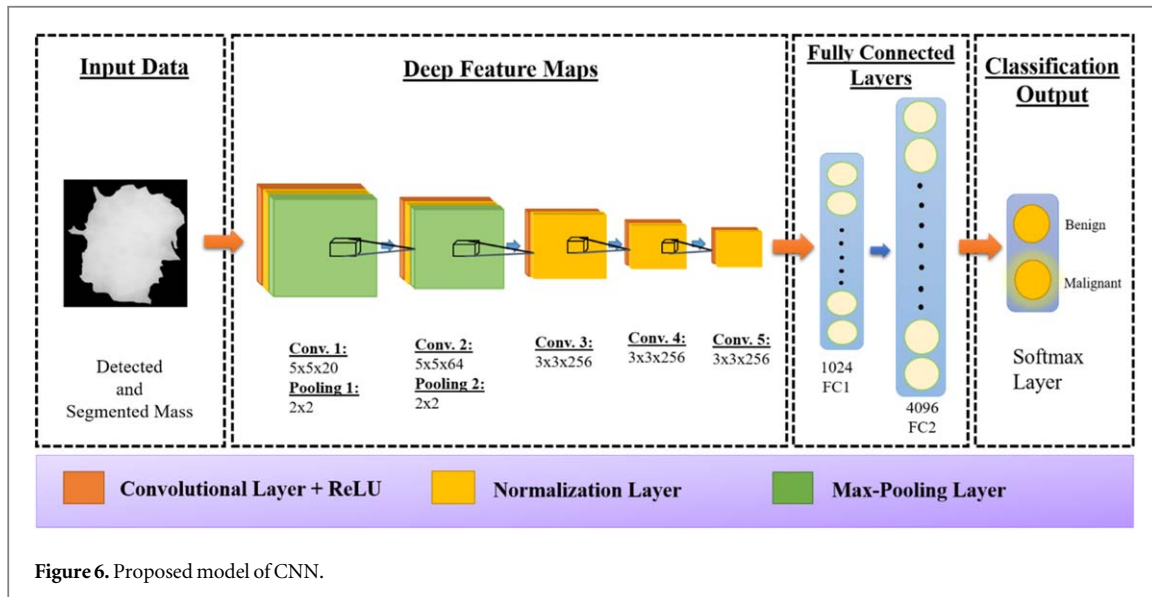


Figure 6. Proposed model of CNN.

the remaining layers have 256 filters with size 3×3 each. The input patch is subsampled by a factor of 2 using non-overlapping max-pooling with a 2×2 size. To improve the CNN's performance, the normalization layer is applied after each convolutional layer. Two fully connected layers are utilized, one with 1024 nodes and the other with 4096 nodes. Following these layers, a Softmax layer is included with two nodes to classify between benign and malignant categories. The proposed CNN is shown in figure 6. The ReLU activation function is applied after every level except the last layer. ReLU is commonly employed in deep learning models due to its rapid saturation relative to other activation functions, resulting in faster training times. As per earlier studies, deep learning models with ReLU is faster and performs better.

Table 1. Assessment of mass detection performance.

Folds	Benign		Malignant		Total		Evaluation metrics		
	TN	FP	TP	FN	TRUE	FALSE	ACC	MCC	F1-Score
1st Fold	104	0	125	1	229	1	99.57	99.13	99.52
	100%	0.00%	99%	1%	99.57%	0.43%	—	—	—
2nd Fold	101	3	126	0	227	3	98.70	97.39	98.54
	97%	3%	100%	0.00%	98.70%	1.30%	—	—	—
3rd Fold	103	1	124	2	227	3	98.70	97.37	98.56
	99.04%	0.96%	98.42%	1.58%	98.70%	1.30%	—	—	—
Average (%)	98.72%	1.28%	99.21%	0.79%	98.99%	1.01%	98.99%	97.96%	98.87%

4. Experimental settings

In the proposed CAD-based model, the MIAS dataset is used to evaluate the proposed BCanD system at all its three phases, namely mass detection, mass segmentation, and mass classification.

The three-fold cross-validation test with the training, test, and validation datasets are applied. To prevent the bias error these datasets were generated through stratified partitioning [12, 16, 18, 56]. The three-fold cross-validation means all the three stages of the proposed BCanD model are trained three times for the performance evaluation of the model. For the experiments, the dataset is divided into 3 groups, 75% for training which includes benign: 378 and malignant: 312, and rest of the 25% of the dataset is divided equally (12.5% each) among validation and testing data which includes benign: 63 and malignant: 52, for each of the two groups. The following conditions were employed to prevent bias that could arise during training for all deep learning models—detection, segmentation, and classification—due to an imbalance in the training data. Initially, the training set undergoes shuffling within each of the mini-batches to ensure that one image is used only once each epoch, adopting the proposed method by Badrinarayanan *et al* [57]. Additionally, earlier research has involved the use of a weighted cross-entropy loss function for estimating parameters in deep learning models, achieved through a minimization process [12, 54, 57]. A dual cross-validation approach is used to identify optimal parameters for all deep learning models through training and validation datasets. The final evaluation of these models is exclusively carried out using the test dataset.

For each stage of the proposed CAD system three-fold cross-validation test is done. The obtained results for the cross-validation of the detection stage is shown in table 1. Similarly, for the segmentation stage the deep learning models FCN [55], SegNet [57] and U-Net are trained using the learning rate of 0.001 with Adam optimizer. And total numbers of 100 epochs is used with 24 mini-batches. To analyze the effect of segmentation stage the system is tested with two scenarios, i.e. classification without the segmentation and with the segmentation. In first scenario the correctly detected masses are inputted into the classification directly, without the segmentation stage. Whereas in the second scenario, the detected masses are segmented and then fed into the classification stage. The same three-fold cross-validation test is performed for both scenarios. A learning rate of 0.0001 is utilized with the Adam optimizer, total number of epochs is set to 100 with a mini-batch size of 20. A dropout of 0.3 is used on the fully connected layers to avoid overfitting [16, 52, 56–58]. The results of both scenario are presented in tables 5 and 6.

For the implementation of the proposed BCanD model, deep learning libraries like Keras, Theano are used. Also, for the identification and segmentation Tensorflow environment was created. For all the experiments, the hardware setup included an Intel Dual-Core i5 2.7 GHz processor, 8 GB of RAM, and an Intel Iris Graphics 6100 GPU. The suggested model was developed using Python 3.9.12 on a system running macOS Monterey version 12.7.2.

4.1. Evaluation metrics

All the stages of the proposed methodology are evaluated using standard evaluation methods. For the detection stage overall accuracy is used, also the F1-score and MCC (Matthews Correlation Coefficient) was used. The F1-score is the harmonic mean of precision and sensitivity, sometimes referred to as the Dice similarity coefficient [59]. The Matthews Correlation Coefficient (MCC) serves as a reliable measure for assessing classification accuracy, particularly in situations where class sizes are imbalanced. It incorporates True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). For appraising the efficacy of the segmentation method, metrics like sensitivity, specificity, overall accuracy, F1-Score, MCC, and the Jaccard index are employed. Optimal specificity and sensitivity are attained through accurate segmentation of all masses and surrounding tissues. The Dice coefficient (F1-score) and Jaccard index measure the similarity between predicted and ground-truth regions by calculating the ratio of false positive pixels to true positive pixels [16, 59]. Matthew

Table 2. Mass detection performance of YOLO with other models.

References	Method	Dataset	Accuracy
Ali <i>et al</i> 2020 [61]	Fully Convolutional Network	MIAS	96%
Shen <i>et al</i> 2020 [43]	Deep Belief Network	MIAS	91.5%
Zhang <i>et al</i> 2021 [62]	Net-5	MIAS	96.1%
Proposed CAD Model BCanD	YOLO	MIAS	98.99%

Table 3. Classification performance (%) of CAD system.

Folds	Benign		Malignant		Malignant					
	TN	FP	TP	FN	Sensitivity	Specificity	Jaccard	Dice	ACC	MCC
1st Fold	62	1	50	2	96.15	98.41	94.34	97.09	97.39	94.74
2nd Fold	59	4	47	5	90.38	93.65	83.93	91.26	92.17	84.19
3rd Fold	58	5	49	3	94.23	92.06	85.96	92.45	93.04	86.06
Average (%)	94.71	5.29	93.59	6.41	93.59	94.71	88.08	93.60	94.20	88.33

Table 4. Classification performance (%) of YOLO.

Folds	Benign		Malignant		Evaluation Metrics					
	TN	FP	TP	FN	Sensitivity	Specificity	Jaccard	Dice	ACC	MCC
1st Fold	47	16	40	12	76.92	74.60	58.82	74.07	75.65	51.31
2nd Fold	49	14	36	16	69.23	77.78	54.55	70.59	73.91	47.20
3rd Fold	48	15	37	15	71.15	76.19	55.22	71.15	73.91	47.34
Average (%)	76.19	23.81	59.79	22.75	72.44	76.19	56.20	71.94	74.49	48.62

Table 5. Classification performance (%) without segmentation.

Folds	Benign		Malignant		Evaluation Metrics					
	TN	FP	TP	FN	Sensitivity	Specificity	Jaccard	Dice	ACC	MCC
1st Fold	57	6	45	7	86.54	90.48	77.59	87.38	88.70	77.16
2nd Fold	54	9	42	10	80.77	85.71	68.85	81.55	83.48	66.61
3rd Fold	55	8	43	9	82.69	87.30	71.67	83.50	85.22	70.12
Average (%)	87.83	12.17	83.33	16.67	83.33	87.83	72.70	84.14	85.80	71.29

correlation coefficient (MCC) metric quantifies the correlation between the pixels representing segmented mass and the corresponding ground-truth values [60]. In the process of image classification, the performance metrics employed for each image (input Region of Interest) include the ROC curve with AUC, MCC, and F1-score, as opposed to evaluating each pixel individually, which is common in segmentation tasks [15, 18, 23, 36, 56].

5. Results and discussion

For the performance evaluation of mass detection using YOLO, a three-fold cross-validation on the test dataset was performed. On the test sets of the MIAS dataset, table 1 depicts the performance of mass detection over three-fold cross-validation using YOLO. The detected mass region is considered accurate if the IoU $\geq 50\%$, otherwise incorrect in each fold. The average overall accuracy of the detection phase is reported as 98.99%, F1-Score is reported as 98.87%, and MCC as 97.96%. The results indicate that the YOLO detector outperforms others in terms of accuracy. Table 2 shows the comparison of YOLO with the other models. The classification result of CAD and YOLO on test data is shown in table 3 and 4 respectively.

Table 4. As per the performance evaluation of YOLO, the ACC is 74.4.9%. For the detection of the mass YOLO has shown high accuracy and outperformed other existing model (ref: table 2), however for the classification of the mass the results are not extremely satisfactory. Therefore, to improve the accuracy of the classification, segmentation and classification stage is required in the CAD system.

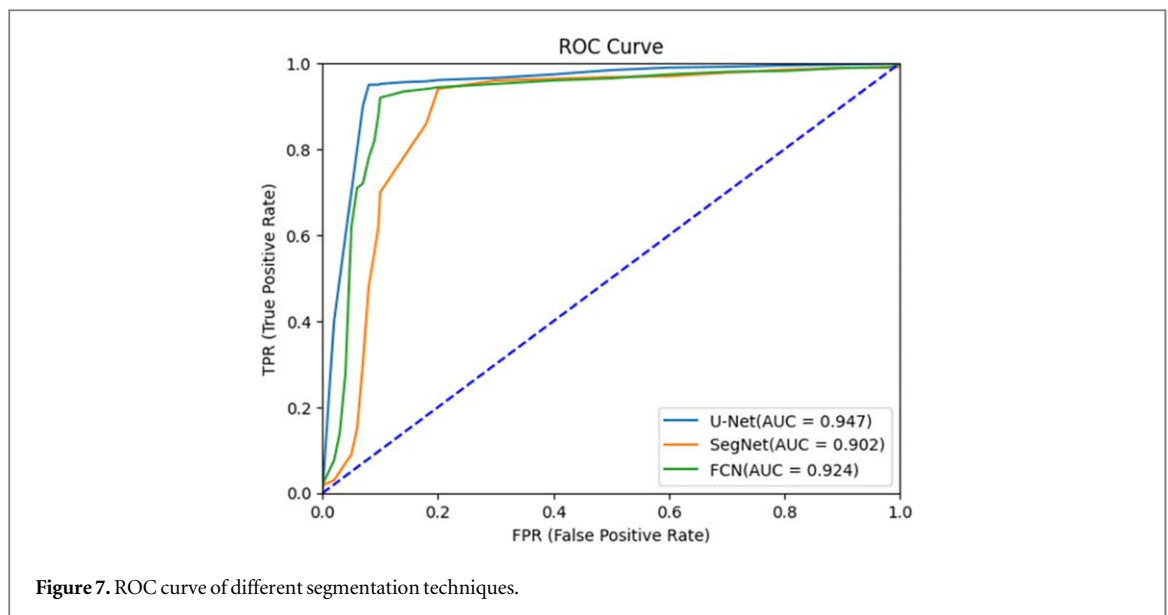


Table 6. Classification of CAD with and without segmentation.

Folds	Class	CAD without segmentation		CAD with segmentation	
		Predicted classes		Predicted Classes	
		Benign	Malignant	Benign	Malignant
1st Fold	Benign	57	6	62	1
		90.48	9.52	98.41	1.59
	Malignant	7	45	2	50
2nd Fold	Benign	13.46	86.54	3.85	96.15
		54	9	59	4
	Malignant	85.71	14.29	93.65	6.35
3rd Fold	Benign	10	42	5	47
		19.23	80.77	9.62	90.38
	Malignant	55	8	58	5
Average (%)	Benign	87.30	12.70	92.06	7.94
		9	43	3	49
	Malignant	17.31	82.69	5.77	94.23
Average (%)	Benign	87.83	12.17	94.71	5.29
	Malignant	16.67	83.33	6.41	93.59

U-Net performs better than Fully Convolutional Network (FCN), and Segmented Neural Network (SegNet) with an overall average accuracy of 95.96% and $AUC \approx 0.95$. Figure 5 shows the segmentation results of FCN, SegNet, and U-Net with their Dice index and overall accuracy for three sample images and figure 7 shows the ROC curve of all these three segmentation techniques. The U-Net outperformed other techniques due to that it has been used in CAD system for further study. The output of segmentation stage is fed into the final stage of the model, i.e., the classification stage. Using the ROC curve, MCC, Dice values, sensitivity, specificity, and total accuracy, the performance of the final classification stage was evaluated. These evaluation metrics were applied for the model with and without segmentation. The confusion metrics with all three folds of without segmentation and with segmentation are tabulated in table 6 and also shown in figure 8. The results of performance evaluation of CAD without and with segmentation is tabulated in tables 3 and 5 respectively. From the performance evaluation result it is clear that the BCanD performed better with the segmentation with the overall average accuracy of 94.20%, as compared to the model without the segmentation stage in which the accuracy is only 85.80%. It is also observed that the correct classification of benign and malignant classes are 94.71% and 93.59% respectively, whereas the false classification of these two classes are 5.29% and 6.42% respectively as shown in table 3. Hence, the BCanD system gives better result with the segmentation in all the three-folds of cross-validation. Figure 9 shows the ROC curve with and without segmentation and their respective AUCs. It is clearly visible that the TPR has increased whereas the FPR has decreased, therefore the

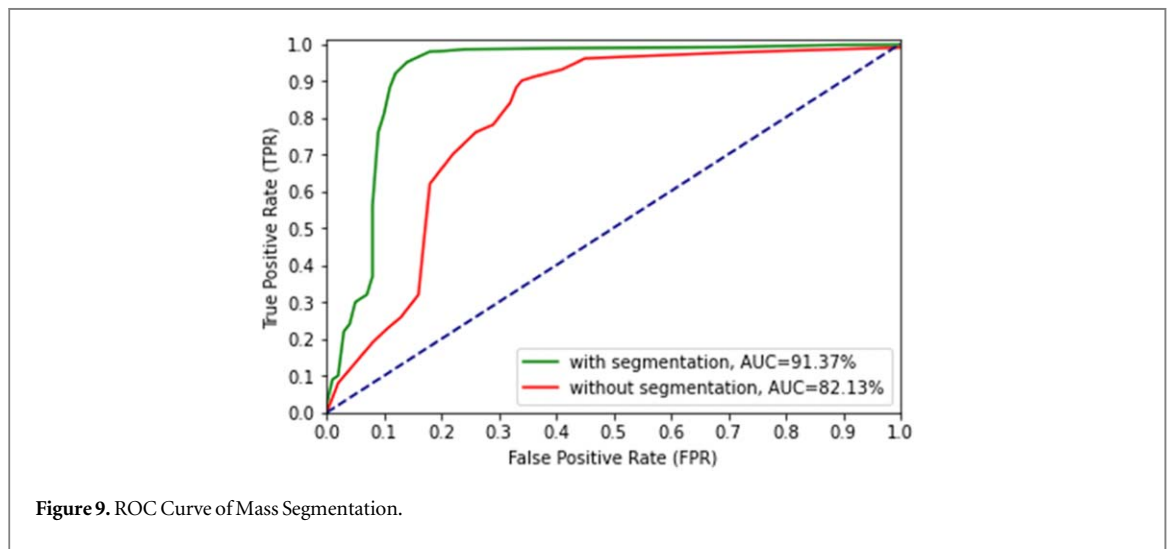
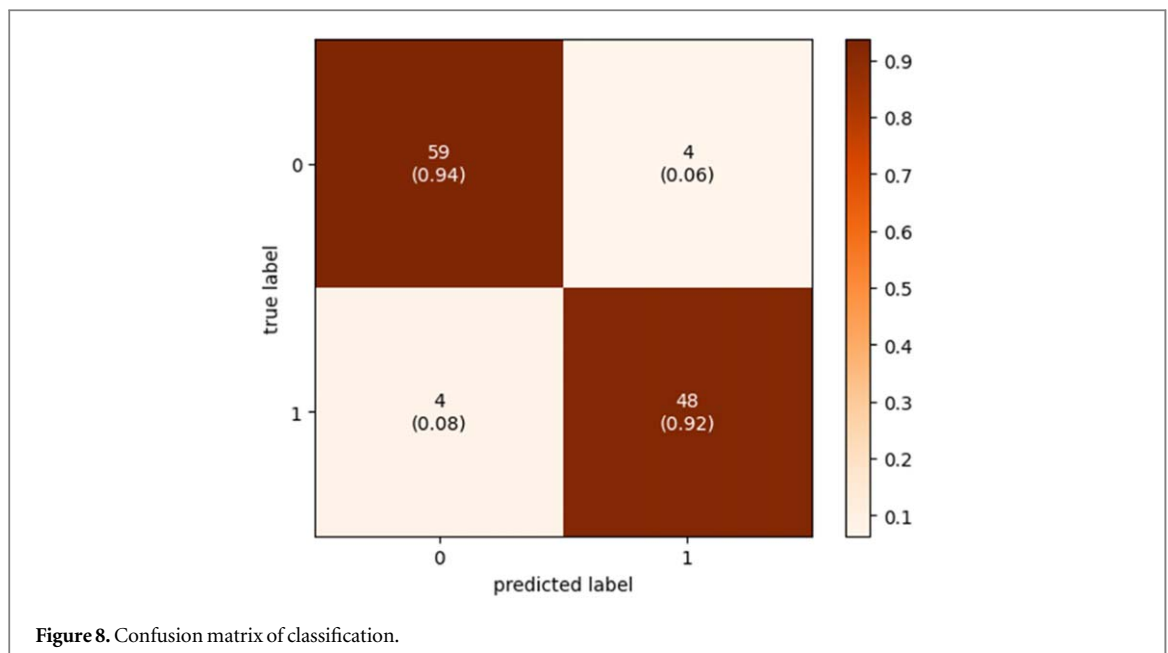


Table 7. Evaluation of the proposed system's performance compared to existing state-of-art systems.

Reference	Dataset	Prediction class	Overall accuracy
Charan <i>et al</i> 2018 [66]	MIAS	Normal/Abnormal	65%
Goudarzi <i>et al</i> 2018 [67]	MIAS	Normal/Abnormal	89.37%
Ting, Tan, and Sim 2019 [68]	MIAS	Benign/Malignant	90.5%
Yu <i>et al</i> 2019 [24]	MIAS	Benign/Malignant	81%
Khalid <i>et al</i> 2021 [69]	MIAS	Normal/Cancerous	87.5%
Alruwaili and Gouda, 2022 [70]	MIAS	Benign/Malignant	89.5%
Proposed CAD Model BCanD	MIAS	Benign/Malignant	94.20%

proposed BCanD system performs better. The earlier deep learning based models majorly focused on the diagnosis of the extracted area of mammograms, however the proposed deep-learning based model is capable of detection, segmentation and classification of the mammogram at the same time. The dataset used in this study is dense, hence the CAD solves the issue of detection and classification of mass in dense regions of the breast.

Deep learning-based CNNs have shown some promising results in medical image analysis [14, 59, 63–65]. The proposed CAD system outperforms any of the state-of-art systems shown in table 7. In earlier literatures, it has been observed that mass detection in dense regions of the breast is a challenging task [12, 15, 16], YOLO-based mass detection played a crucial role in the accurate detection of the mass. The challenges related with

segmentation was resolved using the U-net algorithm, which performed much better than the other techniques. And for the classification of the mass, CNN is used. Although the performance of YOLO for the detection is appreciably good, but in the case of classification it is not up to the mark. Therefore, there is a quite need for an integrated CAD system with segmentation and classification stage. The proposed CAD model performs better with the segmentation as compared to without segmentation, as shown in tables 3 and 5. The proposed CAD achieves the highest accuracy, sensitivity, specificity, jaccard, dice, and MCC with 97.39%, 96.15%, 98.41%, 94.34%, 97.09%, and 94.74% respectively, and average accuracy, sensitivity, specificity, Jaccard, dice, and MCC with 94.20%, 93.59%, 94.71%, 88.08%, 93.60%, and 88.33% respectively. The high performance of the CAD system is due to high detection accuracy (table 2), robust segmentation (figures 5 and 7), and better performance of the proposed CNN (table 3). The contribution of all the stages resulted in a high-performing system that has outperformed all other state-of-the-art systems on the MIAS dataset. The comparative results of the proposed CAD with state-of-the-art systems are shown in table 7. The proposed CAD system is capable of handling all the stages of detection, segmentation, and classification of breast cancer.

6. Conclusion

YOLO-based deep learning models are capable enough to detect mass in the denser tissues or even in pectoral muscles. Therefore, YOLO was selected for mass detection. The U-net was chosen for segmentation stage as it works well with small training set, high training accuracy and low training time And generates pixel level segmentation map. CNN was used for mass classification.

The proposed system is capable of breast mass detection, its segmentation and also classification of the breast mass into benign and malignant. The overall accuracy of the system is 94.20% which is considered as high performance. The BCanD CAD system is efficient in handling the all the stages as compared to other CAD system as reported in table 7. The performance of the proposed integrated CAD, BCanD system is satisfactory and ready to be used by the radiologists for all the stages from detection to diagnosis of breast mass, and can be implemented for clinical applications.

Data availability statement

No new data were created or analysed in this study.

Statements and declarations

Declaration of interest's statement

The authors affirm that they do not possess any identifiable conflicting financial interests or personal ties that could have potentially influenced the findings presented in this paper.

ORCID iDs

Mehar Arfi  <https://orcid.org/0000-0001-9513-6645>

Suman Lata Tripathi  <https://orcid.org/0000-0002-1684-8204>

References

- [1] Eroğlu Y, Yildirim M and Çinar A 2021 Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR *Comput. Biol. Med.* **133** 104407
- [2] Whang J S, Baker S R, Patel R, Luk L and Castro A 2013 The causes of medical malpractice suits against radiologists in the united states *Radiology* **266** 548–54
- [3] Sathishkumar K, Chaturvedi M, Das P, Stephen S and Mathur P 2022 Cancer incidence estimates for 2022 & projection for 2025: result from national cancer registry programme, India *Indian J. Med. Res.* **156** 598–607
- [4] DeSantis C E, Ma J, Goding Sauer A, Newman L A and Jemal A 2017 Breast cancer statistics, 2017, racial disparity in mortality by state *CA Cancer J. Clin.* **67** 439–48
- [5] Xu S, Wu H and Bie R 2019 CXNet-m1: anomaly detection on chest x-rays with image-based deep learning *IEEE Access.* **7** 4466–77
- [6] Gugulothu V K and Balaji S 2023 A novel deep learning approach for the detection and classification of lung nodules from CT images *Multimed Tools Appl.* **82** 47611–34
- [7] Min H, Chandra S S, Dhungel N, Crozier S and Bradley A P 2017 Multi-scale mass segmentation for mammograms via cascaded random forests 2017 *IEEE 14th International Symposium on Biomedical Imaging ISBI 2017* 113–7

- [8] Chakravarthy S R S, Bharanidharan N and Rajaguru H 2023 Processing of digital mammogram images using optimized ELM with deep transfer learning for breast cancer diagnosis *Multimed Tools Appl.* **82** 47585–609
- [9] Dong M, Lu X, Ma Y, Guo Y, Ma Y and Wang K 2015 An efficient approach for automated mass segmentation and classification in mammograms *J. Digit Imaging.* **28** 613–25
- [10] Tang J, Rangayyan R M, Xu J, Naq I E and Yang Y 2009 Computer-aided detection and diagnosis of breast cancer with mammography: recent advances *IEEE Trans. Inf. Technol. Biomed.* **13** 236–51
- [11] American Cancer Society 2022 Limitations of mammograms | how accurate are mammograms? <https://cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/limitations-of-mammograms.html>
- [12] Carneiro G, Nascimento J and Bradley A P 2017 Automated analysis of unregistered multi-view mammograms with deep learning *IEEE Trans. Med. Imaging* **36** 2355–65
- [13] Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V and Garcia-Rodriguez J 2017 A review on deep learning techniques applied to semantic segmentation arXiv preprint arXiv:1704.06857
- [14] Litjens G et al 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88
- [15] Al-antari M A et al 2018 An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network *J. Med. Biol. Eng.* **38** 443–56
- [16] Dhungel N, Carneiro G and Bradley A P 2017 A deep learning approach for the analysis of masses in mammograms with minimal user intervention *Med. Image Anal.* **37** 114–28
- [17] Suckling J, Parker J, Dance D, Astley S and Hutt I 2015 *Mammographic Image Analysis Society (MIAS) Database V1.21* Apollo - University of Cambridge Repository <https://doi.org/10.17863/CAM.105113>
- [18] Al-Masni M A et al 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
- [19] Rajinikanth V and Satapathy S C 2018 Segmentation of ischemic stroke lesion in brain MRI Based on social group optimization and fuzzy-tsallis entropy *Arab. J. Sci. Eng.* **43** 4365–78
- [20] Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R and Barkan E 2016 A region based convolutional network for tumor detection and classification in breast mammography *Deep Learning and Data Labeling for Medical Applications* editors, ed G Carneiro et al (Springer International Publishing) 197–205
- [21] Min H et al 2020 Fully automatic computer-aided mass detection and segmentation via pseudo-color mammograms and mask R-CNN 2020 *IEEE 17th International Symposium on Biomedical Imaging (ISBI)* 1111–5
- [22] Arevalo J, González F A, Ramos-Pollán R, Oliveira J L and Guevara Lopez M A 2016 Representation learning for mammography mass lesion classification with convolutional neural networks *Comput. Methods Programs Biomed.* **127** 248–57
- [23] Jiao Z, Gao X, Wang Y and Li J 2016 A deep feature based framework for breast masses classification *Neurocomputing* **197** 221–31
- [24] Yu X, Zeng N, Liu S and Zhang Y-D 2019 Utilization of DenseNet201 for diagnosis of breast abnormality *Mach. Vis. Appl.* **30** 1135–44
- [25] Krithiga R and Geetha P 2021 Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review *Arch Computat Methods Eng.* **28** 2607–19
- [26] Abhisheka B, Biswas S K and Purkayastha B 2023 A comprehensive review on breast cancer detection, classification and segmentation using deep learning *Arch Computat Methods Eng.* **30** 5023–52
- [27] Dhungel N, Carneiro G and Bradley A P 2015 Deep learning and structured prediction for the segmentation of mass in mammograms *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* 9349 ed N Navab et al (Springer International Publishing) 605–12
- [28] Kumar N and Yadav S C 2023 Comprehensive analysis of deep learning based text classification models and applications *International Journal of Applied Science and Engineering* **20** 2022342
- [29] Iqbal I, Shahzad G, Rafiq N, Mustafa G and Ma J 2020 Deep learning-based automated detection of human knee joint's synovial fluid from magnetic resonance images with transfer learning *IET Image Process.* **14** 1990–8
- [30] Oliveira L L G, Silva S A E, Ribeiro L H V, de Oliveira R M, Coelho C J and Andrade S 2008 ALS. Computer-aided diagnosis in chest radiography for detection of childhood pneumonia *Int. J. Med. Inform.* **77** 555–64
- [31] Kooi T et al 2016 A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography *Breast Imaging* ed A Tingberg et al (Springer International Publishing) 51–6
- [32] Farhan A H and Kamil M Y 2020 Texture analysis of mammogram using histogram of oriented gradients method *IOP Conf. Ser.: Mater. Sci. Eng.* **881** 012149
- [33] Iqbal I, Younus M, Walayat K, Kakar M U and Ma J 2021 Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images *Comput. Med. Imaging Graph.* **88** 101843
- [34] Dhungel N, Carneiro G and Bradley A P 2015 Automated mass detection in mammograms using cascaded deep learning and random forests 2015 *Int. Conf. on Digital Image Computing: Techniques and Applications (DICTA)* 2015, 1–8
- [35] Nanda N, Kakkar P and Nagpal S 2019 Computer-aided segmentation of liver lesions in CT Scans using cascaded convolutional neural networks and genetically optimised classifier *Arab. J. Sci. Eng.* **44** 4049–62
- [36] Al-Masni M A et al 2017 Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2017, 1230–3
- [37] Iqbal A and Sharif M 2023 BTS-ST: Swin transformer network for segmentation and classification of multimodality breast cancer images *Knowl.-Based Syst.* **267** 110393
- [38] Alom M Z, Hasan M, Yakopcic C, Taha T M and Asari V K 2018 Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation arXiv preprint arXiv:1802.06955
- [39] Ma S, Tang J and Guo F 2021 Multi-task deep supervision on attention R2U-net for brain tumor segmentation *Frontiers in Oncology.* **11** 704850
- [40] Lan L, Cai P, Jiang L, Liu X, Li Y and Zhang Y 2024 BRAU-Net++: U-shaped hybrid CNN-transformer network for medical image segmentation arXiv preprint arXiv:2401.00722
- [41] Hossain M S 2022 Micro calcification segmentation using modified u-net segmentation network from mammogram images *Journal of King Saud University-Computer and Information Sciences* **34** 86–94
- [42] Jin Q, Cui H, Sun C, Meng Z and Su R 2021 Cascade knowledge diffusion network for skin lesion diagnosis and segmentation *Appl. Soft Comput.* **99** 106881
- [43] Shen L, He M, Shen N, Yousefi N, Wang C and Liu G 2020 Optimal breast tumor diagnosis using discrete wavelet transform and deep belief network based on improved sunflower optimization method *Biomed. Signal Process. Control* **60** 101953
- [44] Ramesh S, Sasikala S, Gomathi S, Geetha V and Anbumani V 2022 Segmentation and classification of breast cancer using novel deep learning architecture *Neural Comput & Applic.* **34** 16533–45

- [45] Kriti, Virmani J, Dey N and Kumar V 2016 PCA-PNN and PCA-SVM Based CAD systems for breast density classification *Applications of Intelligent Optimization in Biology and Medicine: Current Trends and Open Problems* 96 (Springer) [159–80](#)
- [46] Muramatsu C, Hara T, Endo T and Fujita H 2016 Breast mass classification on mammograms using radial local ternary patterns *Comput. Biol. Med.* **72** 43–53
- [47] Li H, Meng X, Wang T, Tang Y and Yin Y 2017 Breast masses in mammography classification with local contour features *Biomed. Eng. Online* **16** 44
- [48] Yassin N I R, Omran S, El Houby E M F and Allam H 2018 Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review *Comput. Methods Programs Biomed.* **156** 25–45
- [49] Su J, Yu X, Wang X, Wang Z and Chao G 2024 Enhanced transfer learning with data augmentation *Eng. Appl. Artif. Intell.* **129** 107602
- [50] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: unified, real-time object detection *Proc. of the IEEE conference on computer vision and pattern recognition* 779–88
- [51] Aly G H, Marey M A E-R, El-Sayed Amin S and Tolba M F 2021 YOLO V3 and YOLO V4 for masses detection in mammograms with resnet and inception for masses classification *Advanced Machine Learning Technologies and Applications* 1339 ed A-E Hassanien et al (Springer International Publishing) [145–53](#)
- [52] Al-Antari M A, Al-Masni M A, Choi M-T, Han S-M and 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** 44–54
- [53] Kozegar E, Soryani M, Minaei B and Domingues I 2013 Assessment of a novel mass detection algorithm in mammograms *Journal of Cancer Research and Therapeutics* **9** 592–600
- [54] Krizhevsky A, Sutskever I and Hinton G E 2017 ImageNet classification with deep convolutional neural networks *Commun. ACM* **60** 84–90
- [55] Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation *Proc. of the IEEE conference on computer vision and pattern recognition* 3431–40
- [56] Kooi T et al 2017 Large scale deep learning for computer aided detection of mammographic lesions *Med. Image Anal.* **35** 303–12
- [57] Badrinarayanan V, Kendall A and Cipolla R 2017 SegNet: a deep convolutional encoder-decoder architecture for image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 2481–95
- [58] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting *The Journal of Machine Learning Research* **15** 1929–58
- [59] Yu L, Chen H, Dou Q, Qin J and Heng P-A 2017 Automated melanoma recognition in dermoscopy images via very deep residual networks *IEEE Trans. Med. Imaging* **36** 994–1004
- [60] Powers D M 2020 Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation arXiv preprint arXiv:[2010.16061](#)
- [61] Ali M J et al 2020 Enhancing breast pectoral muscle segmentation performance by using skip connections in fully convolutional network *Int. J. Imaging Syst. Technol.* **30** 1108–18
- [62] Zhang Y-D, Satapathy S C, Guttery D S, Górriz J M and Wang S-H 2021 Improved breast cancer classification through combining graph convolutional network and convolutional neural network *Inf. Process. Manage.* **58** 102439
- [63] Arfi M and Yadav S C 2021 A survey on applications of deep learning techniques in medical diagnostics for cancer detection *Drugs and Cell Therapies in Hematology* **10** 1359–67
- [64] Chen X et al 2022 Recent advances and clinical applications of deep learning in medical image analysis *Med. Image Anal.* **79** 102444
- [65] Suganyadevi S, Seethalakshmi V and Balasamy K 2022 A review on deep learning in medical image analysis *Int J Multimed Info Retr.* **11** 19–38
- [66] Charan S, Khan M J and Khurshid K 2018 Breast cancer detection in mammograms using convolutional neural network 2018 *International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* [1–5](#)
- [67] Goudarzi M and Maghooli K 2018 Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer *Biocybernetics and Biomedical Engineering* **38** 1004–14
- [68] Ting F F, Tan Y J and Sim K S 2019 Convolutional neural network improvement for breast cancer classification *Expert Syst. Appl.* **120** 103–15
- [69] Khalid S, Krishnan S and Thanki R M 2021 Deep learning model for classification of breast cancer in *Artificial Intelligence in Breast Cancer Early Detection and Diagnosis* (Springer) 93–100
- [70] Alruwaili M and Gouda W 2022 Automated breast cancer detection models based on transfer learning *Sensors* **22** 876