DIR

Diagn Interv Radiol 2025; DOI: 10.4274/dir.2024.242923

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

Breast cancer detection and classification with digital breast tomosynthesis: a two-stage deep learning approach

Yazeed Alashban

King Saud University, College of Applied Medical Sciences, Department of Radiological Sciences, Riyadh, Saudi Arabia

PURPOSE
The purpose of this study was to propose a new computer-assisted two-staged diagnosis system that combines a modified deep learning (DL) architecture (VGG19) for the classification of digital breast tomosynthesis (DBT) images with the detection of tumors as benign or cancerous using the You Only Look Once version 5 (YOLOv5) model combined with the convolutional block attention module (CBAM) (known as YOLOv5-CBAM).

METHODS
In the modified version of VGG19, eight additional layers were integrated, comprising four batch normalization layers and four additional pooling layers (two max pooling and two average pooling). The CBAM was incorporated into the YOLOv5 model structure after each feature fusion. The experiment was carried out using a sizable benchmark dataset of breast tomography images. A total of 22,032 DBT examinations from 5,060 patients were included in the data.

RESULTS
Test accuracy, training loss, and training accuracy showed better performance with our proposed architecture than with previous models. Hence, the modified VGG19 classified DBT images more accurately than previously possible using pre-trained model-based architectures. Furthermore, a YOLOv5-based CBAM precisely discriminated between benign lesions and those that were malignant.

CONCLUSION
DBT images can be classified using modified VGG19 with accuracy greater than the previously available pre-trained models-based architectures. Furthermore, a YOLOv5-based CBAM can precisely distinguish between benign and cancerous lesions.

CLINICAL SIGNIFICANCE
The proposed two-tier DL algorithm, combining a modified VGG19 model for image classification and YOLOv5-CBAM for lesion detection, can improve the accuracy, efficiency, and reliability of breast cancer screening and diagnosis through innovative artificial intelligence-driven methodologies.

KEYWORDS
Breast imaging, artificial intelligence, mammography, breast cancer

Corresponding author: Yazeed Alashban

E-mail: Yalashban@ksu.edu.sa

Received 02 July 2024; revision requested 05 September 2024; accepted 12 October 2024.

Epub: 09.12.2024

Publication date: 28.04.2025

DOI: 10.4274/dir.2024.242923

Breast cancer (BC) is one of the main causes of mortality in women and a major global health concern.[1] According to data from the World Health Organization, in 2022, 665,684 women worldwide lost their lives due to malignancy in the breast, accounting for 2.3 million new cases of the disease. BC is the most common cancer globally among women; between 2015 and 2021, 7.8 million women were diagnosed with the disease.[2] In the US, BC ranks as the second most common malignancy after lung cancer, according to the Surveillance, Epidemiology, and End Results Program.[3] Globally, according to available data,[3,4] one in eight women will contract BC. As a result, BC screening is one of the most significant and common

You may cite this article as: Alashban Y. Breast cancer detection and classification with digital breast tomosynthesis: a two-stage deep learning approach. *Diagn Interv Radiol.* 2025;31(3):206-214.

medical imaging prerequisites, with over 39 million examinations carried out annually.[5] Early identification and discovery are essential for therapy, rehabilitation, and a decrease in death rates.[6] The prognosis and survival rate of cancer vary greatly depending on its stage. Cancer treatment is more successful the earlier the disease is discovered.[7]

Radiologists examine and annotate images generated by screening techniques to identify tumors.[8] The gold standard for this cancer screening has been supplanted by the relatively new imaging technique known as digital breast tomosynthesis (DBT), which has taken the place of mammography.[8] This is a type of three-dimensional (3D) mammography that aims to increase abnormality detection.[9] DBT, which recreates multiple low-dose picture projections from a moving digital X-ray source over a restricted arc angle, is used to build the 3D model.[10] Since a two-dimensional (2D) mammography examines every tissue in the breast at once, there is a risk that certain tissue features will overlap and produce inaccurate results. By allowing radiologists to view multiple layered images prior to classifying tumors, DBT helps address some of the problems associated with 2D mammograms.[11] Compared with traditional mammography, DBT often requires longer image acquisition and processing times, as well as increased radiation exposure (though it is still within safe limits).[10,11]

Radiologists are already using computer-aided diagnosis tools to help them make decisions.[12] These technologies have the potential to reduce significantly the time and energy required to evaluate a lesion in clinical practice.[13] They may also reduce the occurrence of false positives, which lead to unnecessary and uncomfortable biopsies.[14] Recent technological advancements in deep learning (DL), such as artificial neural networks and transfer learning, have outperformed several machine learning algorithms in tasks such as classifying and identifying lesions.[15] Unlike traditional machine learning methods, which require a manual feature extraction and selection step, DL algorithms adaptively learn the optimal feature extraction process from the input data.[13,14]

However, although DL techniques for lesion detection and classification have been used extensively using mammography, there have been few studies using DBT. This could be attributed to the computer memory constraints associated with DL methods, which are linked to the higher dimensionality of the data. In previous studies, breast tumors from DBT data have been segmented, classified, and detected using DL. Li et al.[16] carried out deep convolutional neural network (DCNN)-based mass classification of BC using DBT and assessed different transfer learning strategies. They collected data on 441 patients who had undergone DBT and conducted three different experiments to compare 2D and 3D DCNNs trained on volumetric DBT. The 2D convolutional neural network (CNN) that was trained on both DBT and full-field digital mammography achieved better results, with a change in area under the curve of 0.009.[16]

Ricciardi et al.[17] developed a DCNN-based detection system for the automatic classification of the presence or absence of mass lesions in DBT-annotated images. Background correction, data augmentation, and normalization were basic pre-processing steps. Three DCNN architectures trained on two distinct datasets were compared: 1) built from scratch (DBT-DCNN); 2) pre-trained (AlexNet and VGG19); 3) optimized using a transfer learning approach. Additionally, a Grad-CAM technique was used to provide a position indication for the lesion in the DBT. The accuracy of the DBT-DCNN network was 90% ± 4%, and the sensitivity was 96% ± 3%.[17]

Lotter et al.[8] presented a DL method that was annotation-efficient and accurate; the method achieved maximum performance in classification, detected cancers in clinically negative mammograms, was effectively applicable to a population with low screening participation, and outperformed five full-time breast-imaging radiologists, with an average 14% increase in sensitivity. The model used a multiple-instance learning approach in which it was progressively and effectively trained on DBT using only breast-level labels. The authors were successful in maintaining localization-based interpretability by generating new "maximum suspicion projection" images from DBT data.[8]

For the prediction of Ki-67 expression in DBT images, Oba et al.[18] developed a model based on DL. The Ki-67 expression of 126 patients with pathologically proven BC was chosen and assessed. The DL model employed the Xception architecture to forecast the levels of Ki-67 expression. The accuracy, on average, was 0.912. The findings point to the possible use of their model to predict Ki-67 expression from DBT, which is useful in deciding on a BC treatment plan prior to surgery.[18] Buda et al.[19] shared a large-scale publicly available DBT examination dataset, which included information for 5,060 patients, and used it to train a detection model. One hundred twenty-four images having bounding boxes for malignant and 175 images having bounding boxes for benign lesions were used to develop a detection algorithm based on a 2D DenseNet. There was no pretraining on alternative datasets or comparable modalities, such as mammography. The free-response receiver operating curve, displaying the sensitivity of the model in relation to false-positive predictions, was utilized for the ultimate assessment of the baseline detection algorithm.[19]

Earlier studies have proposed classification or detection using DBT with notable contributions. However, they are limited by fewer images in the datasets,[16] lack of external validation and clinical assessment,[17] limited comparison of advanced architectures, and lack of diversity in training data.[8,18] The DBT data used in this study has also been utilized in several studies;[19-25] however, these studies have either focused on the classification or detection of the lesions as benign or malignant. Table 1 provides a summary of the studies conducted on the Duke Dataset from the Cancer Imaging Archive (TCIA).[26] The model in the current study is the first state-of-the-art model that classifies a DBT scan into one of three classes: normal, actionable, and tumor. Moreover, it detects the lesion as benign or cancerous. The model incorporates a modified VGG19 DL architecture. The batch normalization layers are placed in every fourth convolutional layer to enhance the model's training efficiency

Main points

- The modified VGG19 architecture classified digital breast tomosynthesis (DBT) images more accurately than previously available pre-trained models.
- The You Only Look Once version 5 (YOLOv5)-based convolutional block attention module (CBAM) precisely discriminated between benign and malignant lesions, showing better performance metrics such as test accuracy and training loss.
- The research fills a significant gap in breast cancer diagnosis by utilizing advanced deep-learning strategies for DBT images. The two-tier deep learning algorithm, combining a modified VGG19 model for image classification and YOLOv5-CBAM for lesion detection, demonstrated good outcomes in terms of accuracy and time.
- The study highlights the advantages of DBT over traditional two-dimensional mammography, emphasizing its improved accuracy in screening due to the three-dimensional view it provides.

Table 1. Summary of the previous studies utilizing data from the Cancer Imaging Archive

Citation	Architecture	Pre-processing	Training/testing dataset	Outcome	Results
21	ResNet-18, AlexNet, MobileNetV2, GoogleNet, DenseNet-201, VGG-16,	DBT augmentation; image enhancement techniques; color feature mapping	Patients – 5,060 Slices – 22,032	Classification into normal, benign, and malignant	Acc.: 56.52
22	R-CNN	Conversion of volume intensities to 8-bits depth; extraction of breast mask area; flipping to convert all the volumes into same orientation	Patients – 5,060 Slices – 22,032	Lesion detection	IOU: 0.85
23	ResNet	Cropping; reduction of pixels; transformation	Cancer + actionable - 100 Normal + benign - 100	Classification	Acc.: 86
24	Inception v3	Cropping; reduction of pixels; augmentation	Normal – 1,680 Tumor – 223	Lesion detection	Acc.: 91.4
25	Faster R-CNN	Data augmentation; image flipping; image translation; channel reception augmentation	Patients – 985 Scans – 1,000	Detection	Acc.: 83.08
26	2 Layer DenseNet	Cropping; downscaling	Patients – 5,060 Scans – 22,032	Lesion detection	Sensitivity: 78
27	Faster R-CNN	Cropping; normalization; masking and background suppression	VICTRE + Patients – 5,060 Scans – 22,032	Lesion detection	Sensitivity: 60

DBT, digital breast tomosynthesis.

by reducing internal covariant shifts. The tumor is detected using a You Only Look Once version 5 (YOLOv5)-based convolutional block attention module (CBAM) architecture, utilizing the two submodules of CBAM: channel attention and spatial attention. This study explores the integration of YOLOv5 (a state-of-the-art object detection model) with CBAM (a mechanism that enhances feature representation) to improve detection accuracy and efficiency. Thus, this model has applications in both the screening and diagnosis of BC.

Methods

Dataset

The dataset available on TCIA website was used in this investigation; it was acquired from the Duke Health System using the Duke Enterprise Data Unified Content Explorer tool between January 1, 2014, and January 30, 2018.[26] The data included a total of 22,032 DBT examinations from 5,060 patients. The dataset included DBT images from four different views along with four categories of cases: normal (no sign of cancer and a biopsy was never performed), actionable (cancer may be present, but no biopsy was performed), biopsy-proven benign (a biopsy was performed, and the tumor was determined to be benign), and biopsy-proven cancer (a biopsy was performed, and the tumor was classified as malignant).[24] The Digital Imaging and Communications in Medicine (DICOM) images consisted of a collection of 2D slices taken from four different views: left-mediolateral oblique, right-mediolateral oblique, left-craniocaudal, and right-craniocaudal.

Ethics

This investigation utilized data from the TCIA website, which was obtained from the Duke Health System. Since the data is publicly available and patient consent is not required, ethical approval was not necessary for this study.

Methodology

The overall methodology consisted of two stages: classification and detection. First, the images were classified as normal, actionable, benign, or cancer. In the second stage, the lesion was detected as benign or cancerous using the annotated images containing bounding boxes on the tumor area. The step-by-step methodology for each stage is shown in Figure 1, which summarizes the entire architecture utilized in this study.

Data pre-processing

Certain pre-processing steps were applied at both stages. The following sections describe all the steps that were applied to prepare the dataset for modeling.

Classification

The following steps were carried out to prepare the DBT images for classification into normal, actionable, or tumor. The images were changed from DICOM to JPEG format, a transformation that not only simplified the data format but also allowed compatibility with the next stages of processing. The intensity rescaling was done to standardize the pixel intensity values in all the images so the uniformity of the image could be ensured and the effect of the different illuminations or contrasts could be eliminated. Color space conversion was also carried out to improve the understandability and the discriminative capacities of the images, which, in turn, facilitates the extraction of more significant features for the classification of the images. Resizing was done to adjust the spatial sizes of the images. Hence, the spatial dimensions of the images were harmonized, which improved consistency and removed possible distortions that could affect the analysis. Normalization-the scaling of pixel values to a standard range-was also performed.

Detection

To prepare the data for the detection stage of determining the tumor as benign or cancerous, the images were first augmented, and then the pre-processing techniques mentioned in the classification were applied. The process of purposefully increasing the volume and complexity of already-existing data is known as data augmentation. Data augmentation has become a necessary pre-processing step in DL.

Because a significant number of training samples are needed for neural networks and medical datasets are sometimes scarce, the first step in increasing the diversity of the dataset is data augmentation; in this study, the Roboflow tool was used for this activity. The following steps were performed: text files were generated to contain essential annotations, and all the generated text files were imported to Roboflow. Five types of augmentation techniques were applied: horizontal and vertical flips, 90-degree rotations (clockwise, anticlockwise, upside down), cropping (ranging from 0% to 25% maximum zoom), rotations (–15 to +15 degrees), and shears (10-degree vertical and horizontal). The total number of images before and after augmentation is given in Table 2.

Data splitting

The dataset was split into three subsets for the classification and detection stages: training (number of images for classification: 19,148, number of images for detection: 2,116), validation (number of images for classification: 1,163, number of images for detection: 604), and testing (number of images for classification: 1,721, number of images for detection: 303) in the 70, 20, and 10 ratios. The number of instances in each split for each category for the classification framework is given in Table 3.

Experimental setup

The chosen equipment, including an NVIDIA RTX 4090 GPU and AMD EPYC 7R12 48-Core Processor, provided high computational power (1.8 TFLOPS and 24.0/192 CPU cores, respectively), which is essential for intensive model calculations. The motherboard ROME2D32GM supports PCIe 4.0, enhancing data transfer speeds (22.8 GB/s), which is crucial for handling large datasets. With 516 GB of memory and a 4TB Predator SSD, the system ensured ample storage and quick data access (3,830 MB/s), supporting efficient model training and analysis. The equipment's high-performance specifications were used to optimize model development and execution. The pre-processed DBT images were classified using VGG19, and detection was based on YOLOv5-CBAM.

Modelling

The modified VGG19 model was used to classify the DBT slice images into normal, actionable, and tumor. A YOLOv5-CBAM model was used for the detection of lesions as benign or cancerous.

Modified VGG19

Transfer learning involves transferring the learned parameters of the pre-trained CNN model. It involves shifting the weights (as given in Table 4) of a CNN model that was trained on additional sizable datasets.[27] Scientists are creating deeper learning models to increase performance as DL models have become more and more popular in image classification and recognition applications. VGG19 is a neural network comprising 43 layers, namely the input, 16 convolutional layers, 16 ReLU layers, 5 max pool layers, 3 full-connected layers, 1 softmax layer, and the output. In this way, the modified version of VGG19 consisted of 8 complementary layers, which were 4 batch normalization layers and 4 extra pooling layers. The batch normalization layers consisted of 2 max pooling and 2 average pooling layers. The layering of batch normalization layers between every 4th convolution layer was interpreted to improve training efficiency by reducing internal instability. This modification produces not only a smaller scale or initial values of the gradient that parameters rely on for modifying but also a better and more natural flow of data between the intermediate layers of the neural network, which greatly reduces the number of iterations required for training. As to extra pooling layers to further the 5th and the 10th convolutional layers of the DL model, crucial low-level details are passed through the learning model, and this helps capture sharp features integrally. The size of the input image was 512 × 512.

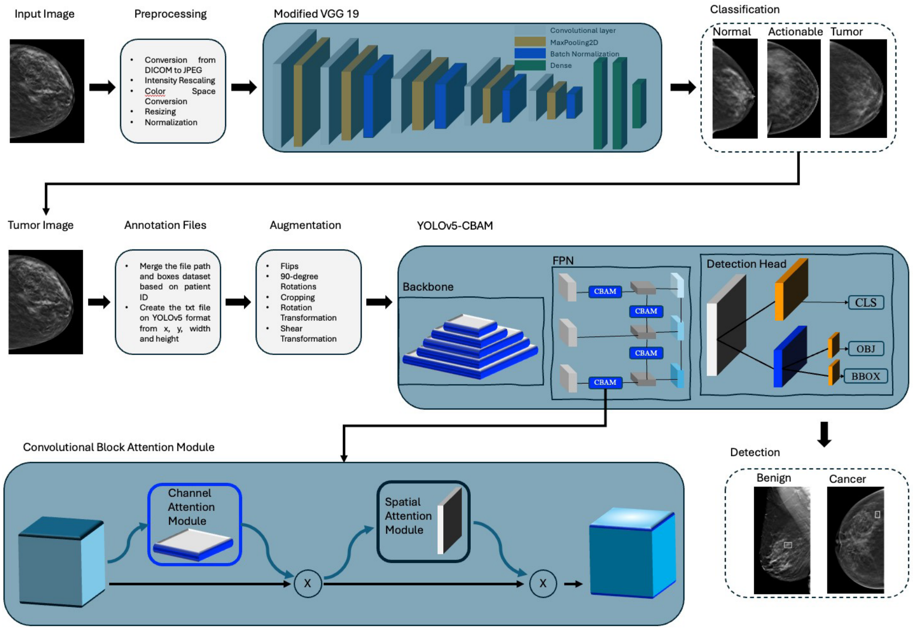

Figure 1. The schematic depicts the organization of the suggested framework. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Table 2. Total number of images before and after augmentation for the application of the YOLOv5-CBAM architecture

Dataset	Before augmentation	After augmentation
Training	293	2,293
Validation	58	456
Testing	35	274
Total	386	3,023

YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Table 3. Summary of the dataset splitting for each class

Category	Training	Testing	Validation
Normal	18,232	1,356	928
Actionable	716	244	160
Tumor	200	121	75

YOLOv5-CBAM

The attention mechanism makes it possible for models to prioritize and process information selectively, focusing only on the most crucial details and ignoring the rest. Convolutional block attention modules are one type of attention mechanism meant to enhance CNN performance. As can be seen in Figure 2, the CBAM is incorporated into the proposed model structure after each feature fusion, or "concat." In an image or feature map, it mainly enhances feature extraction and records meaningful spatial and channel-wise dependencies. The efficacy of this module is demonstrated in the experiments reported in the study,[28] where the performance of the module is significantly improved by integrating the CBAM into various models across a variety of classification and detection datasets.

Convolutional block attention modules are made up of two sub-modules: the channel attention module and the spatial attention module. The primary focus of channel attention is on locating the essential traits or features needed to identify a lesion in an image. However, it is crucial to remember that the lesion is a relatively small and sparse component within the entire image when it comes to particular tasks, such as lesion detection. In these situations, the value of the individual pixels in the entire image is not equal. At this point, spatial attention is applied to solve the "where" issue, which involves locating the lesion in the image. Functioning alongside channel attention, spatial attention gathers data from various spatial regions of the image. By giving these spatial features weights, it essentially highlights the areas of the picture where lesions are present. Applying channel and spatial attention in that order achieves this. Figure 3 illustrates how channel attention can compute channel weights represented as WCA ∈ RC × 1 × 1, and spatial attention can compute spatial weights (WS) denoted as WS ∈ RH × W × 1, given the input feature map F ∈ RW × H × C.

Channel attention refers to a multi-step process that is applied to an input feature map (F). Global max pooling (GMP) and global average pooling (GAP) are carried out to record the highest and lowest spatial responses. These responses are then processed by a multi-layer perceptron. Then, element-wise addition is used to integrate the results of GMP and GAP. After that, a sigmoid activation function is applied to the combined data, resulting in a channel weight feature map that assigns a weight to each channel based on its significance. Finally, an element-wise multiplication is performed between the channel weight matrix and the original feature map (F) as:

(1)

$$F' = F \times W_{CA}$$

Table 4. Parameter values at each layer of the modified VGG19 model

Layer name	Activation maps	Learnable parameters	Total learnable parameters
Input	512 × 512 × 3	-	0
block1_conv1	512 × 512 × 64	Weights: 3 × 3 × 3 × 64, bias: 64	1,792
block1_conv2	512 × 512 × 64	Weights: 3 × 3 × 64 × 64, bias: 64	36,928
block1_pool	256 × 256 × 64	-	0
block2_conv1	256 × 256 × 128	Weights: 3 × 3 × 64 × 128, bias: 128	73,856
block2_conv2	256 × 256 × 128	Weights: 3 × 3 × 128 × 128, bias: 128	147,584
batch_normalization_1	256 × 256 × 128	Offset: 128, scale: 128	512
block2_pool	128 × 128 × 128	-	0
block3_conv1	128 × 128 × 256	Weights: 3 × 3 × 128 × 256, bias: 256	295,168
average_pooling2d_1	64 × 64 × 256	-	0
block3_conv2	64 × 64 × 256	Weights: 3 × 3 × 256 × 256, bias: 256	590,080
block3_conv3	64 × 64 × 256	Weights: 3 × 3 × 256 × 256, bias: 256	590,080
block3_conv4	64 × 64 × 256	Weights: 3 × 3 × 256 × 256, bias: 256	590,080
batch_normalization_2	64 × 64 × 256	Offset: 256, scale: 256	1,024
block3_pool	64 × 64 × 256	-	0
block4_conv1	32 × 32 × 512	Weights: 3 × 3 × 256 × 512, bias: 512	1,180,160
block4_conv2	32 × 32 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
max_pooling2d_1	16 × 16 × 512	-	0
block4_conv3	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block4_conv4	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
batch_normalization_3	16 × 16 × 512	Offset: 512, scale: 512	2,048
block4_pool	16 × 16 × 512	-	0
block5_conv1	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block5_conv2	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block5_conv3	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
block5_conv4	16 × 16 × 512	Weights: 3 × 3 × 512 × 512, bias: 512	2,359,808
batch_normalization_4	8 × 8 × 512	Offset: 512, scale: 512	2,048
block5_pool	8 × 8 × 512	-	0
flatten_1	8,192	-	0
dense_3	4,096	Weights: 8,192 × 4,096, bias: 4,096	33,558,528
dense_4	4,096	Weights: 4,096 × 4,096, bias: 4,096	16,781,312
dense_5	3	Weights: 4,096 × 3, bias: 3	12,291
SoftMa ×	1 × 1 × 3	-	0
Classification Output	1 × 1 × 3	-	0
Number of total learnable parameters			70,379,331

where F′ is the weighted feature map, F is the input feature map, and W_{CA} is the channel weight matrix. The channel weight matrix is computed as follows:

(2)

$$\boldsymbol{W_{CA}(F) = \sigma(f_{cencoder}(AvgPool(F)) + f_{cencoder}(MaxPool(F)))}$$

where the global max-pooling operation is represented by MaxPool, the average pooling operation is AvgPool, σ is the sigmoid function, and f_c is the channel encoder.

Spatial attention functions were analyzed using GAP and GMP to compute the average and maximum spatial responses on the input feature map. These resulting responses are utilized to combine into a set of descriptive features. A spatial weight feature map (WS) is produced by activation with a sigmoid function and is multiplied element-wise by the original feature map. This approach distills the model's focus to important regions of the network, thus identifying the spatial attention process, given as:

(3)

$$\boldsymbol{F'' = F' \times W_{SA}}$$

where W_{SA} is the spatial weight matrix and is calculated as:

(4)

$$\boldsymbol{W_{SA}(F) = \sigma(f_{cencoder}(AvgPool(F))©f_{cencoder}(MaxPool(F)))}$$

The proposed model leverages YOLOv5 to preserve the original network topology while extracting features from the three feature layers of the backbone network. The head network receives these features after they have been concatenated and sent for object detection. The head network's ability to comprehend complex spatial feature arrangements in the data is improved by integrating the CBAM. When dealing with small objects or intricate details, such as tiny lesions within an image, this problem becomes extremely helpful. Through the refinement of the network's understanding of semantic and spatial nuances, the CBAM enhances detection performance. It increases the model's ability to locate and identify small teeth with greater accuracy and generates a stronger recognition effect without increasing the training cost.

Results

This study's primary goal was to develop and evaluate a model for BC screening and diagnosis from DBT data with greater accuracy. A two-stage architecture was developed for this purpose.

Classification

The task of optimizing DL models to achieve the highest level of accuracy and detection in computer vision is still the main concern in BC screening. In this study, the performance of the VGG19 architecture-known for its impact on image processing- is shown, and the DBT images are classified into tumor, normal, and actionable classes. The effect of different optimizers and batch sizes on the accuracy and loss of the model is extensively studied. Specifically, the influence of these parameters on the performance of the VGG19 architecture in classifying DBT images into tumor, normal, and actionable categories. Table 5 presents the findings of the modified VGG19. It can be seen that with the increase in batch size, the performance, accuracy, and loss increase. The Adam optimizer shows better performance than the other two, with the highest accuracy and minimum loss.

Our findings reflect the complex relationship between optimization techniques and the model's performance; thus, we demonstrated that the Adam optimizer is superior in achieving high accuracy and minimizing the loss in different batch sizes. Furthermore, the confusion matrix shown in Figure 4 not only gives more weight to our classification results but also explains the model's ability to correctly distinguish between the classes. The model shows a greater accuracy in classifying normal, actionable, and tumor, with all three classes having a true positive rate >89%.

Detection

Figure 5 compares model-to-model performance (YOLOv5 vs. YOLOv5-CBAM); the 100-epoch training period is assessed. The metrics plotted are the precision, recall, and mean average precision (mAP), and the threshold is 0.5. The YOLOv5-CBAM model gave higher sums than the standard YOLOv5 for all metrics. This is a clear indicator of im-

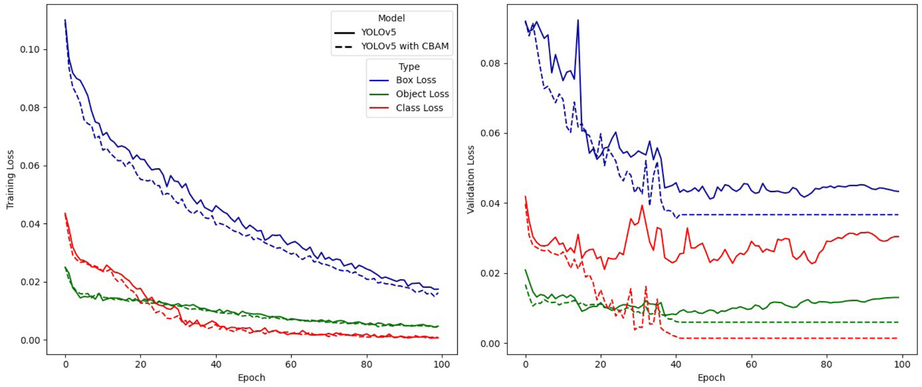

Figure 2. Comparison of the validation loss and training loss for YOLOv5-CBAM and YOLOv5. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

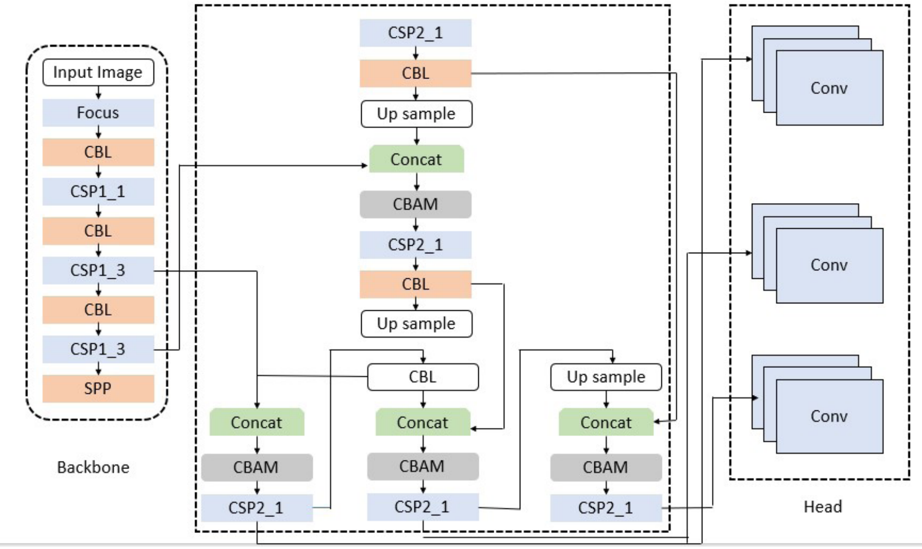

Figure 3. YOLOv5-CBAM. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

proved detection efficiency. The graph presents variations mostly at the beginning (0 to 20) since the model takes time to increase the weights. With time, the metrics reach steady states, where mAP is gradually getting better, which plays a role in the model's convergence. The recall for both models is usually more stable and consistently remains fairly high, whereas precision has a greater degree of fluctuation compared with recall. Thus, the model is most accurate when it predicts the relevant cases, specifically those instances where abnormalities are present in the DBT images.

Similarly, Figure 2 shows the comparison of the box loss, object loss, and class loss. The addition of CBAM to YOLOv5 shows a reduction in all three parameters in both training and validation. This shows that the YOLOv5 based-CBAM can learn from the data well and generalize it.

Table 6 displays the performance metrics for two configurations of YOLOv5. It indicates that the YOLOv5 model enhanced with the CBAM significantly outperforms the standard YOLOv5 across all metrics for both benign and cancerous classes, suggesting that the CBAM addition effectively improves the model's detection and classification capabilities in these specific medical imaging tasks.

Similar results can be observed from the confusion matrices. The YOLOv5-CBAM model shows a significant improvement over the standard YOLOv5 in both classes. It has higher true positive rates for both benign (0.84 vs. 0.71) and cancerous (0.89 vs. 0.78). It also has lower false positive and false negative rates, indicating better overall accuracy and reliability in classification. The comparison of the confusion matrices for YOLOv5 and YOLOv5-CBAM is shown in Figure 6.

Through the analysis, the variations in performance are unveiled, which are the focus of the CBAM in the detection efficiency and model convergence. Furthermore, we conduct a detailed comparison of loss parameters between the two models, and thus, we get to the issues of their learning dynamics and generalization capabilities. Furthermore, a thorough analysis of the performance metrics of both YOLOv5 versions is presented, which helps explain their effectiveness for different classes.

Table 5. Performance metrics for the VGG19 model

	Batch size	Training accuracy	Training loss	Validation accuracy	Validation loss	Testing accuracy	Testing loss
SGDM	32	88%	0.32	86%	0.34	85%	0.35
	64	85%	0.35	83%	0.37	82%	0.39
	512	82%	0.39	80%	0.42	78%	0.44
Adam	32	87%	0.3	86%	0.32	85%	0.33
	64	90%	0.27	89%	0.29	88%	0.3
	512	95%	0.2	94%	0.22	93%	0.23
RMSProp	32	87%	0.31	85%	0.33	84%	0.35
	64	84%	0.34	82%	0.36	81%	0.38
	512	80%	0.38	78%	0.4	77%	0.42

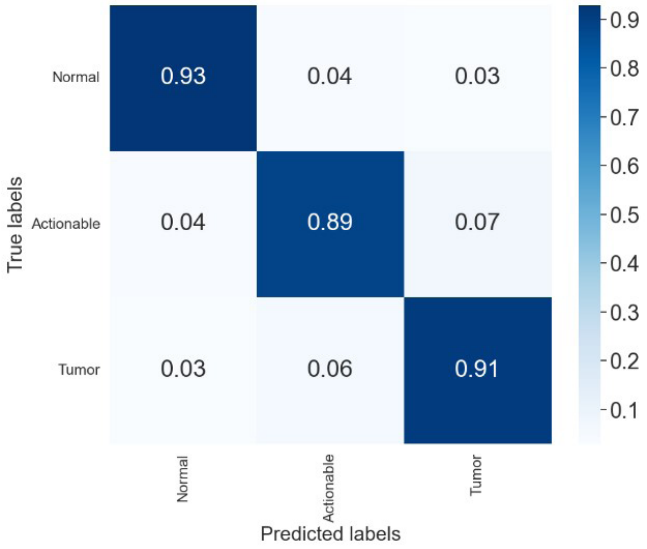

Figure 4. Confusion matrix for the VGG19 model.

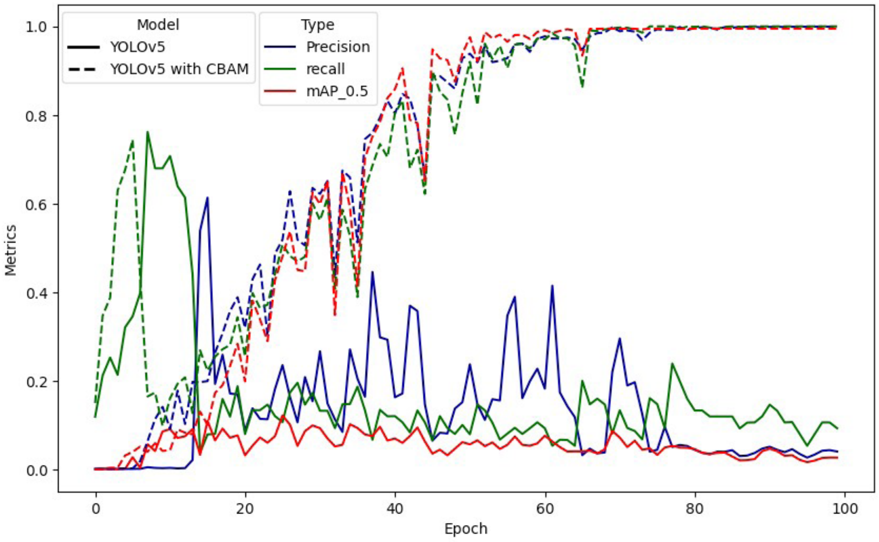

Figure 5. Comparison of the precision, recall, and mAP of YOLOv5 and YOLOv5-CBAM. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module; mAP, mean average precision.

Table 6. Performance metrics for the two configurations of the YOLOv5 model

	mAP		Recall		F1-score	
Classes	YOLOv5	YOLOv5-CBAM	YOLOv5	YOLOv5-CBAM	YOLOv5	YOLOv5-CBAM
Benign	0.815	0.915	0.716	0.879	0.760	0.890
Cancer	0.793	0.891	0.751	0.881	0.713	0.856
Overall	0.785	0.887	0.796	0.896	0.775	0.878

YOLOv5, You Only Look Once version 5; mAP, mean average precision; CBAM, convolutional block attention module.

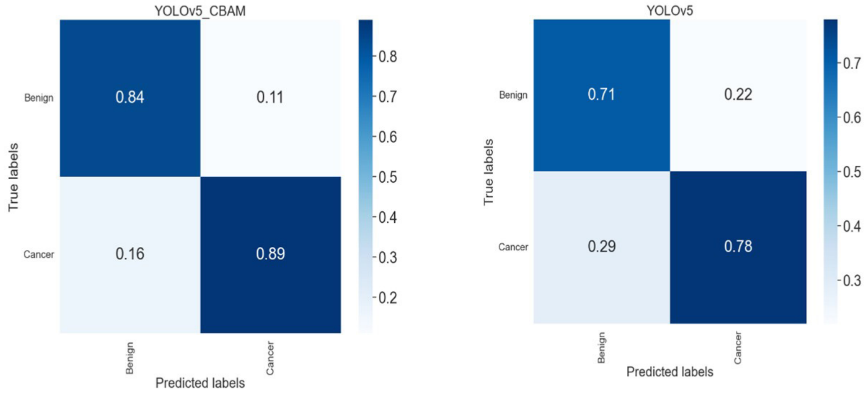

Figure 6. The comparison of the confusion matrices for YOLOv5 and YOLOv5-CBAM. YOLOv5, You Only Look Once version 5; CBAM, convolutional block attention module.

Discussion

This research work fills a significant void in BC diagnosis through the use of enhanced DL strategies for DBT images. Today, BC is still considered one of the most widespread health issues affecting women worldwide; its early diagnosis can contribute to enhancing the effectiveness of its treatment and, consequently, the increase in female survival rates. DBT has brought improvements in the accuracy of BC screening by giving a 3D view, which has some discrepancies as compared with traditional 2D mammography. The developed two-tier DL algorithm includes a modified VGG19 model for image classification and YOLOv5-CBAM for identification of lesions, which demonstrates good outcomes in terms of accuracy and time.

In the classification stage, better results from the previous pre-trained models are achieved by the modified VGG19 with extra layers, such as batch normalization and pooling layers, to improve feature extraction and training. This change not only enhances the performance of classification models in normal, actionable, and tumor types but also highlights the versatility of DL approaches in medical imaging where conventional methods may fall short. The inclusion of YOLOv5-CBAM in the detection stage adds to the model's effectiveness in the identification of malignant and benign lesions based on the attention mechanisms that align data highlights in the image. The YOLOv5-CBAM model improves performance through its attention mechanisms, which focus on the most informative features in the image. The CBAM enhances the model's ability to prioritize relevant areas by applying both channel and spatial attention, thereby reducing false positives and improving the detection accuracy of malignant and benign lesions.

Based on the experimental outcomes, the applicability of the philosophy of the proposed work has been demonstrated, and new achievements in the analysis of DBT using comparable methodologies have been established. To achieve this, the study adopted TCIA, which provides a large dataset to minimize the likelihood of model overfitting, which is detrimental in clinical applications. Furthermore, the study presents methodological reflections, including such aspects as data preprocessing, dataset enlargement, and computational environment; these are crucial for reproducing the presented study and expanding similar research. Altogether, this study advances the knowledge base of artificial intelligence (AI)-supported BC diagnosis and lays down the foundation for effective diagnosis models that can enhance identification processes globally, hence boosting patients' survival.

Despite the promising results, the study faced several limitations. The reliance on a specific dataset from TCIA may limit generalizability to other populations. Additionally, the model's performance in real-world clinical settings needs further validation. Challenges also included handling variability in image quality and computational resources required for model training. Future work will be able to extend to various associated modalities, add multiple imaging data, and carry out studies with clinical materials to validate the performance in practical scenarios. Moreover, the role of DL models in the diagnostics of medical conditions, together with methods that enable the interpretability and explainability of the results, deserves more attention and development to earn the trust of physicians. Finally, harnessing such advanced AI technologies as the ones discussed in this study has the potential to significantly improve the practice of BC screening and its management and, thus, the state of global healthcare.

In conclusion, over 39 million examinations are performed yearly as part of the BC screening program. However, BC screening has been one of the most difficult applications of AI in medical imaging. DBT can create 3D images where tissue overlapping is reduced, making it simpler for radiologists to spot abnormalities and resulting in a more precise diagnosis. This study suggested the use of a new computer-aided multi-class diagnosis system that uses YOLOv5-CBAM to identify benign or malignant tumors and a modified DL architecture (VGG19) for classifying DBT images. A large set of breast tomography images was used in the experiment. Test accuracy, training loss, and training accuracy showed better performance of our proposed architecture than the previous models. Hence, the modified VGG19 classified DBT images more accurately than previously possible using pre-trained model-based architectures. Second, YOLOv5-based CBAM precisely discriminated between benign lesions and those that are malignant.

Footnotes

Conflict of interest disclosure

The author declared no conflicts of interest.

Funding

The author would like to extend their sincere appreciation to the researchers supporting the program for funding this work under the Researchers Supporting Project number

(RSPD2024R780), King Saud University, P.O. Box 145111, Riyadh 4545, Saudi Arabia.

References

1. Sharma MP, Shukla S, Misra G. Recent advances in breast cancer cell line research. *Int J Cancer*. 2024;154(10):1683-1693. [CrossRef]
2. da Costa Nunes GG, de Freitas LM, Monte N, et al. Genomic variants and worldwide epidemiology of breast cancer: a genome-wide association studies correlation analysis. *Genes (Basel)*. 2024;15(2):145. [CrossRef]
3. Giaquinto AN, Sung H, Miller KD, et al. Breast cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(6):524-541. [CrossRef]
4. Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)*. 2022;135(5):584-590. [CrossRef]
5. Yi A, Jang MJ, Yim D, Kwon BR, Shin SU, Chang JM. Addition of screening breast US to digital mammography and digital breast tomosynthesis for breast cancer screening in women at average risk. *Radiology*. 2021;298(3):568-575. [CrossRef]
6. Zheng J, Lin D, Gao Z, et al. Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access*. 2020;8:96946-96954. [CrossRef]
7. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. *Science*. 2022;375(6586):9040. [CrossRef]
8. Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med*. 2021;27(2):244-249. [CrossRef]
9. Rezaei Z. A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Syst Appl*. 2021;182:115204. [CrossRef]
10. Dhamija E, Gulati M, Deo SVS, Gogia A, Hari S. Digital breast tomosynthesis: an overview. *Indian J Surg Oncol*. 2021;12(2):315-329. [CrossRef]
11. Gao Y, Moy L, Heller SL. Digital breast tomosynthesis: update on technology, evidence, and clinical practice. *Radiographics*. 2021;41(2):321-337. [CrossRef]
12. Wani IM, Arora S. Computer-aided diagnosis systems for osteoporosis detection: a comprehensive survey. *Med Biol Eng Comput*. 2020;58(9):1873-1917. [CrossRef]
13. Zebari DA, Ibrahim DA, Zeebaree DQ, et al. Systematic review of computing approaches for breast cancer detection based computer aided diagnosis using mammogram images. *Appl Artif Intell*. 2021;35(15):2157-2203. [CrossRef]
14. Ramadan SZ. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *J Healthc Eng*. 2020;2020:9162464. [CrossRef]
15. Xu L, Gao J, Wang Q, et al. Computer-aided diagnosis systems in diagnosing malignant thyroid nodules on ultrasonography: a systematic review and meta-analysis. *Eur Thyroid J*. 2020;9(4):186-193. [CrossRef]
16. Li X, Qin G, He Q, et al. Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. *Eur Radiol*. 2020;30(2):778-788. [CrossRef]
17. Ricciardi R, Mettivier G, Staffa M, et al. A deep learning classifier for digital breast tomosynthesis. *Phys Med*. 2021;83:184-193. [CrossRef]
18. Oba K, Adachi M, Kobayashi T, et al. Deep learning model to predict Ki-67 expression of breast cancer using digital breast tomosynthesis. *Breast Cancer*. 2024. [CrossRef]
19. Buda M, Saha A, Walsh R, et al. A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images. *JAMA Netw Open*. 2021;4(8):2119100. [CrossRef]
20. El-Shazli AMA, Youssef SM, Soliman AH. Intelligent computer-aided model for efficient diagnosis of digital breast tomosynthesis 3D imaging using deep learning. *Applied Sciences (Switzerland)*. 2022;12(11):5736. [CrossRef]
21. Martí R, del Campo PG, Vidal J, et al. Lesion detection in digital breast tomosynthesis: method, experiences and results of participating to the DBTex challenge. Proceedings volume 12286, 16th International Workshop on Breast Imaging (IWBI2022); 122860W, 2022. [CrossRef]
22. Nogay HS, Akinci TC, Yilmaz M. Comparative experimental investigation and application of five classic pre-trained deep convolutional neural networks via transfer learning for diagnosis of breast cancer. *Adv Sci Technol Res J*. 2021;15(3):1-8. [CrossRef]
23. Maciej Serda, Becker FG, Cleary M, et al. Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza. G. Balint, Antala B, Carty C, Mabieme JMA, Amar IB, Kaplanova A, eds. Uniwersytet śląski. 2013;7(1):343-354. [CrossRef]
24. Hassan L, Saleh A, Singh VK, Puig D, Abdel-Nasser M. Detecting breast tumors in tomosynthesis images utilizing deep learning-based dynamic ensemble approach. *Computers 2023*. 2023;12(11):220. [CrossRef]
25. Mota AM, Clarkson MJ, Almeida P, Matela N. Detection of microcalcifications in digital breast tomosynthesis using faster R-CNN and 3D volume rendering. In proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022). *Bioimaging*. 2022;2:80-89. [CrossRef]
26. Konz N, Buda M, Gu H, et al. A competition, benchmark, code, and data for using artificial intelligence to detect lesions in digital breast tomosynthesis. *JAMA Netw Open*. 2023;6(2):230524. [CrossRef]
27. Wang X, Chen G, Qian G, et al. Large-scale multi-modal pre-trained models: a comprehensive survey. *Mach Intell Res*. 2023;20(4):447-482. [CrossRef]
28. Akbarnezhad E, Naserizadeh F. Improving camouflage object detection using U-NET and VGG16 deep neural networks and CBAM attention mechanism. *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)*. 2024:56-62. [CrossRef]