Segmentation for mammography classification utilizing deep convolutional neural network

Dip Kumar Saha[1], Tuhin Hossain[2], Mejdl Safran[3*], Sultan Alfarhood[3], M. F. Mridha[4*] and Dunren Che[5]

Abstract

Background Mammography for the diagnosis of early breast cancer (BC) relies heavily on the identification of breast masses. However, in the early stages, it might be challenging to ascertain whether a breast mass is benign or malignant. Consequently, many deep learning (DL)-based computer-aided diagnosis (CAD) approaches for BC classification have been developed.

Methods Recently, the transformer model has emerged as a method for overcoming the constraints of convolutional neural networks (CNN). Thus, our primary goal was to determine how well an improved transformer model could distinguish between benign and malignant breast tissues. In this instance, we drew on the Mendeley data repository's INbreast dataset, which includes benign and malignant breast types. Additionally, the segmentation anything model (SAM) method was used to generate the optimized cutoff for region of interest (ROI) extraction from all mammograms. We implemented a successful architecture modification at the bottom layer of a pyramid transformer (PTr) to identify BC from mammography images.

Results The proposed PTr model using a transfer learning (TL) approach with a segmentation technique achieved the best accuracy of 99.96% for binary classifications with an area under the curve (AUC) score of 99.98%, respectively. We also compared the performance of the proposed model with other transformer model vision transformers (ViT) and DL models, MobileNetV3 and EfficientNetB7, respectively.

Conclusions In this study, a modified transformer model is proposed for BC prediction and mammography image classification using segmentation approaches. Data segmentation techniques accurately identify the regions affected by BC. Finally, the proposed transformer model accurately classified benign and malignant breast tissues, which is vital for radiologists to guide future treatment.

Keywords Mammography, Breast cancer, Segmentation, Classification, SAM

*Correspondence:
Mejdl Safran
mejdl@ksu.edu.sa
M. F. Mridha
firoz.mridha@aiub.edu
Full list of author information is available at the end of the article

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 2 of 18

Introduction

Background

Cancer is a fatal disease affecting people worldwide. Since the year 1,900, researchers have been gathering information about cancer, and it has been widely acknowledged that it is an incurable disease [1]. By 2030, it is possible that the incidence of this condition will increase by more than 50 percent in the United States [2]. This disorder involves the uncontrolled growth and dissemination of atypical cells within the organism. Based on benignity and malignancy, the two principal forms of cancer are malignant and benign, respectively. Benign cancer cells are typically non-cancerous and have a very slow development rate. In contrast, malignant tumors exhibit rapid growth, pose a significant threat, and can easily spread through the bloodstream to other areas of the body [3]. Women commonly experience various types of cancer, including lung, bone, blood, brain, liver, and BC. The first reported case of BC, which is considered to be one of the earliest forms of cancer seen anywhere around the globe, was found in Egypt in the year 1,600 BC [4, 5]. It is the second greatest cause of death among females, according to data that was exposed by the American Cancer Society in the year 2017 [6, 7]. About 2.1 million women have BC every year, and it may be deadly if left untreated, says the World Health Organisation [8]. If BC is identified early, the patient has a 70% better chance of living for five years than if the cancer is more advanced [9]. Hence, timely identification and management of BC are of utmost significance for patients.

Motivation

There are many approaches for detecting BC, some of which have been discussed in [10]. Mammography is the most extensively used and successful tool for detecting BC because it is inexpensive and satisfies standards for utilization by practitioners [11]. When performing mammography, low-energy X-rays are employed in the process of acquiring images of the breast, which are subsequently analyzed to identify any signs of cancer. Typically, doctors obtain diagnostic results from mammography analyses; however, these results are susceptible to human error owing to factors such as doctors' subjective experiences and exhaustion. Even for experienced doctors, it is still difficult to determine what is wrong with a breast mass on a mammogram because the lumps do not appear at the start. According to relevant research, a reliable CAD system can assist doctors in making accurate decisions while efficiently alleviating patient difficulties [12]. The idea of DL emerged from recent developments in ML and image processing [13], which has attracted significant interest from researchers [14]. The first step in the classification process involves manually extracting features from the mammograms. Subsequently, all extracted features were fed into the ML classifier for further classification [15]. Although the conventional pattern recognition approach has had a few successes in mammogram classification, it is reliant on the researcher's deliberately generated features and lacks the possibility of autonomous learning. This deficiency can be efficiently addressed by utilizing a technique known as a CNN. It performs well in natural image analysis and can automatically pick out and extract image features. As a result, CNN has piqued the interest of many researchers, who have sought to use it in the analysis and diagnosis of medical images such as CT lung images [16], MRI brain images [17], and thyroid ultrasound images [18]. Although CNN-based classification models have proven to be effective, recent developments have led to the development of a unique vision model known as a ViT. This approach has been proved as more accurate for various public benchmarks [19]. However, only a small number of research have examined the usage of PTr in BC identification so far, and the potential of PTr in this area has not been fully investigated [20]. Therefore, to effectively detect BC, we propose a novel PTr model using a data segmentation technique.

Contribution

The primary contribution to this field is that our methodology can help achieve more accurate early identification of BC. The dataset used in this study has two categories: benign and malignant. A mismatch between classes may lead to learning bias in the model. We propose a segmentation-dependent class balance as a solution to this problem. Second, we developed an improved PTr-based transformer method for mammography classification. This enhanced transformer-based strategy improves upon the shortcomings of CNN methods by employing the TL technique [21]. The optimal cutoff for removing the ROI from mammography was determined using the SAM approach. To detect BC in the INbreast dataset, we successfully modified the design at the bottom of the PTr method. Throughout the model, spatial reduction attention (SRA) progressively decreased the physical dimensions of its features. Assigning an embedding location to each transformer component further strengthens the concept of order. BC can be classified from INbreast images by successfully modifying the architecture of the bottom layer of PTr. Therefore, the primary insights of this research are the following:

1. We successfully changed the design in the bottom layer of the PTr using the TL approach to identify BC in mammography images.

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 3 of 18

2. We preprocessed the data and segmented the images using the SAM model to create a more flexible and better-performing model.

3. To contrast the suggested model's performance in binary classification tasks, ViT and two DL models EfficientNetB7 and MobileNetV3 were used.

"Related works" section is a review of the proposed literature, and "Materials and methods" section presents the methodology. "Materials and methods" section is further divided into five subsections: image data pre-processing, the architecture of the SAM model, hyper-parameter tuning for SAM, data splitting, the proposed method, and an enhanced pyramid transformer using the TL approach. "Experimental evaluation and results" section represents a comprehensive evaluation of the execution and outcomes, which are also divided into six parts: implementation details, environment specifications, dataset description, performance evaluation metrics, segmentation results using SAM, results analysis employing the proposed architecture, and the computational complexity of all evaluated models. Finally, the paper is concluded in "Conclusion and future work" section with recommendations for further study.

Related works
Deep learning models
Maqsood et al. [22] suggested a DL approach called transferable texture CNN which can detect the early stages of BC in screening images taken by mammography detectors. After applying this method to the mammography, INbreast, and MIAS datasets, the following results were obtained: accuracy of 99.08%, 96.82%, and 96.57%, respectively. Ragab et al. [23] classified benign and malignant tissues using mammograms. This study proposed two segmentation methods. Manual determination of the ROI is required for the first strategy, whereas the threshold with region-based approach called DCNN-SVM-AlexNet is utilized for the second approach. The results showed promise for the second segmentation method, SVM with a linear kernel function, compared to the first method using the mammography dataset. The accuracy was 80.5%, and the F1 score, AUC, sensitivity, specificity, and precision were 81.5%, 88%, 77.4%, and 84.2%, respectively. Salama et al. [24] demonstrated a novel technique for the segmentation and classification of BC images and various models have been utilized to classify benign and malignant tissues using the mammography, CBIS-mammography, and MIAS datasets. These models included InceptionV3, DenseNet121, MobileNetV2, VGG16, and ResNet50. By employing the mammography database, the proposed method of implementing data augmentation using InceptionV3 and a customized U-Net

framework yielded optimal results. This method achieved 98.87% accuracy, 98.98% sensitivity, 98.88% AUC, 98.79% precision, 97.99% F1 score, and 1.2134 processing times with the mammography datasets. Shen et al. [25] demonstrated several DL models, including Resnet50 and VGG16 for accurate detection of BC. When tested against pre-existing approaches on the CBIS-mammography dataset, this neural network approach for screening mammography classifications delivered superior results. After 198 epochs, the Resnet50 model has reached the highest accuracy of 97%. Murtaza et al. [26] presented a comprehensive review of the classification, segmentation, and grading of a wide variety of cancer forms, including BC, using conventional ML techniques and hand-engineered characteristics. This classification was accomplished using medical imaging multimodalities and cutting-edge artificial deep neural network techniques. To reduce the number of inconsistencies that are present in BC images, pre-processing strategies such as image normalization, scaling, and augmentation are utilized. Park et al. [27] suggested an artificial intelligence-based CAD for mammography screenings to improve cancer diagnosis. A total of 204 cases: 137 were determined to be truly negative, 33 to have mild symptoms, and 34 to have missed cancer. AI-CAD exhibited sensitivity, specificity, and diagnostic accuracy rates of 84.7%, 91.5%, and 86.3% for diagnostic mammograms, and 67.2%, 91.2%, and 83.38% for previous mammograms, respectively. Lotter et al. [28] utilized a DL method that works well with both strongly and poorly labeled data by training it in stages and maintaining its ability to be understood based on the location. The fact that the system was trained on a diverse range of sources, including a total of five datasets, is one of the reasons for its successful generalization. Through the application of DL, these findings demonstrate significant potential for the diagnosis of cancer in its early stages and enhancement of access to mammography screening programs. Dar et al. [29] demonstrated a significant potential for early cancer diagnosis and enhanced access to screening mammography through the application of DL. They employed multiple DL methods to identify and classify BC through diverse imaging modalities such as histopathology, mammography, thermography, PET/CT, MRI, and ultrasound. Frazer et al. [30] used the BreastScreen Victoria dataset, which was labeled by breast imaging radiologists and validated by surgical histopathology to detect the presence of BC based on DL strategies. Three DL models were used in this study: EfficientNetB6, NASNetLarge, and Inception-ResNet-V2. Additionally, two global + local techniques, GL1 and GL2, were employed, with GL1 utilizing distinct models and GL2 utilizing the same model with various hyper-parameter settings. Based on 349 test

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 4 of 18

samples and 930 test images, the best assessment result produced from GL2-ResNet22/18-Setting1 was an ACC of 0.8178 [95% CI 0.785, 0.850], and AUC of 0.8979 [95% CI 0.873, 0.923]. Li et al. [31] suggested a two-view model for classifying mammograms as either benign or malignant using CNN and recurrent neural networks (RNNs). Mammogram breast mass features were extracted from craniocaudal and mediolateral oblique (MLO) views using two modified ResNets that constitute the model. After applying the suggested two-view neural network (TV-NN) framework to the mammography database, the model generated a recall of 94.1%, an accuracy of 94.7%, and a maximum AUC of 0.968. Al-Masni et al. [32] represented a CAD architecture that uses the You Only Look Once (YOLO) principle to detect breast masses and categorize cancers. Using convolutional layers and fully connected neural networks, the proposed CAD system can detect the exact position of a mass and differentiate between benign and malignant tumors using an ROI-based CNN technique. Finally, the proposed CNN approach achieved 99.7 % accuracy for mass location detection, 97.0 % accuracy for classifying benign from malignant lesions, and an AUC value of 0.97 on the mammography dataset. Das et al. [33] compared shallow and deep CNN and presented a pre-trained approach that can categorize full mammogram images as either benign or malignant. The output knowledge is then input into three distinct types of shallow CNNs that vary in their representation. Two additional methods utilize fine-tuned transfer learning to feed the same set of images into pre-trained CNNs: VGG19, MobileNet-v2, ResNet50, Xception, Inceptionv3, and Inception-ResNet-v2. The experiment with two datasets revealed that the CBIS-DDSM dataset had an accuracy of 80.4%, while the INbreast dataset had an accuracy of 89.2%, 87.8%, and 95.1%. To segment mammography images, Saffari et al. [34] presented the Full-Resolution Convolutional Network (FrCN), an innovative segmentation architecture. Furthermore, the researchers used three conventional DL models-ResNet-50, InceptionResNet-V2, and a regular feedforward CNN-to categorize the identified and segmented breast lesions that are either benign or cancerous. With mammography images of the INbreast database, the FrCN-based breast lesion segmentation approach obtained impressive results, including a Dice coefficient of 92.69%, an overall accuracy of 92.97%, a Matthews Correlation Coefficient (MCC) of 85.93%, and a Jaccard similarity coefficient of 86.37%. Hama et al. [35] proposed a smart computer-based BC detection system that makes use of digital mammography and the transfer learning model to reduce the occurrence of these mistakes. Two steps make up the suggested method. First, they fine-tune the pre-trained Mask R-CNN model employing the COCO dataset to identify and segment breast lesions. The segmented lesions are then analyzed using various convolutional deep-learning models to determine whether they are benign or malignant. These models include ResNet101, VGG16, ResNet34, VGG19, DenseNet121, and AlexNet. On the INbreast dataset, the DenseNet121 model attained a breast lesion classification accuracy of 99.44%, while the average precision for the lesion detection and segmentation phases was 96.26%.

Transformer models

Few studies have explored the potential of ViT in mammography classification for early BC detection. Jaehwan et al. [36] proposed a transformer-based DL system that can normalize mammograms and account for inter-reader variability in the grading. They suggested a method that predicts the input mammography normalization parameters using a photometric transformer network (PTN) as a programmable normalization module. With easy integration with the main prediction network, it is possible to simultaneously learn optimal normalization and density grades. Comes et al. [37] suggested a model to build a clinical support tool for predicting the pathological complete response to nonadjacent chemotherapy early on. To start, in order to avoid manual feature extraction, a pre-trained CNN named AlexNET was employed to automatically extract low-level features that were associated with the image's local structure. The next step was to identify the most stable features and utilize them to build a support vector machine(SVM) classifier. Finally, the suggested model generated an accuracy of 91.4% for the fine-tuning dataset and 92.3% for the independent test by incorporating the optimal features extracted from both early treatment, and pre-treatment tests with certain clinical features, progesterone receptor(PgR), human epidermal growth factor receptor 2(HER2), and molecular subtype. The AUC value was 0.93 on the former and 0.90 on the latter. In addition, Van et al. [38] proposed and tested a new cross-view transformer method that was both token-based and pixel-wise using two publicly available mammography datasets. To avoid the need for pixel-by-pixel correspondence, a method that relies on transformers to combine views at the feature map level is suggested. In contrast to the traditional method of processing data within a single sequence using conventional transformers, they employed cross-view attention to transfer information across views. In another study, Tummala et al. [39] recommended a combined classifier of four swin transformer models for binary classification and eight-class classification using 7909 histopathological images obtained from the BreaKHis dataset at numerous zoom factors. The ensemble of Swin transformers surpassed

prior research for both eight-class and two-class BC classifications. The ensemble achieved 96.0% testing accuracy for eight-class classification, and for two-class classification, the testing accuracy is 99.6% without employing any pre-processing or augmentation methods. Khamparia et al. [40] also developed a hybrid transfer-learning framework that combines a modified Visual Geometry Group (VGG) and ImageNet model for detecting BC using both two-dimensional and three-dimensional mammography images. According to the experimental findings, the proposed hybrid transfer learning model has an accuracy of 94.3% and an AUC value of 93.3%, which are higher than those of the other examined CNNs. Garrucho et al. [41] developed eight cutting-edge identification methods, one of which was based on a transformer approach, and tested them in five previously unexplored areas to determine the effectiveness of generalizing mammography models. When it came to generalizing the mammography domain, they found that deformable-detection transformer (DETR) models based on transformers were the most reliable and effective. Wang et al. [42] also applied a proprietary semi-supervised learning framework, the integration of a ViT with an adaptive token sampler (ATS), and a consistency training (CT) model to classify BC using two separate datasets containing ultrasound and histology images. By comparing it with previous studies, the proposed model outperforms it in every one of the four metrics: accuracy 98.12%, recall 98.65%, precision 98.17%, and F1 factor 98.41%. In a different study, Dey et al. [43] presented a strategy that made use of transfer learning for detecting BC from the mammogram images. This approach utilizes the DenseNet121 pre-trained model and is implemented on the DMR-IR dataset. Furthermore, edge information was extracted from the thermal breast images using two detectors: Prewitt and Roberts. The experimental findings demonstrate that the proposed approach achieves a classification accuracy of 98.80%, which is higher than that of any other model. Boumaraf et al. [44] presented a novel transfer-learning-based automated approach for BC identification utilizing histopathological images. This method performs well for both eight-class and magnification-independent (MI) binary classification. To make the knowledge gained from ImageNet images more task-specific while freezing the remaining initial residual blocks, they presented a transfer learning approach that depended on a block-wise fine-tuning technique. Further enhancement of the scalability of the suggested method was achieved by utilizing a GCN derived from the data values of the target and three-fold data augmentation. The results demonstrated the efficacy of the suggested approach, as the MD binary classification accuracy ranged between 98.08% to 99.25%, and the MD

eight-class classification accuracy ranged between 89.56% to 94.49%. Additionally, the accuracies for the eight-class and binary classifications in the MI task were obtained as 92.03% and 98.42%, respectively. The current diagnostic building process for CADs often ignores all of the annotations in a mammography image; Chougrad et al. [45] proposed a technique that incorporates all of these comments to construct a CAD system that can provide a thorough BC diagnosis. The proposed VGG-FTED model performed well across all four benchmark datasets, with AUC values of 0.93, 0.89, 0.86, and 0.94 respectively. Shobayo et al. [46] developed a method for automated BC detection in mammography by using a combination of ViT and several pre-trained CNNs, ResNet50, and VGG16. They also use ViTBase, a custom model architecture, to identify BC mammograms as benign or malignant. Finally, they compared DL models based on ViT with those based on CNNs to find the best one for correctly categorizing mammography images of BC. With an accuracy of 99.9% and a precision of 99.8%, the Swin transformer demonstrates exceptional performance. Khan et al. [47] presented a TE-inspired deepgene transformer. It incorporates one-dimensional (1D) convolution layers along with a multi-head self-attention mechanism to create a hybrid architecture. This hybrid architecture was designed to analyze high-dimensional gene expression datasets and to determine whether the representation learned from the attention mechanism can surpass existing approaches. The comparative investigation demonstrated that the suggested DeepGene Transformer beat the existing conventional and state-of-the-art classification methods. This indicates that the DeepGene Transformer can be considered an effective method for classifying cancer and its various aspects. Su et al. [48] suggested a dual-shot architectural model that combines the YOLO and LOGO architectures to identify and segment masses at the same time. In the first step, they used YoloV5L6, the most advanced object detection framework, to find and crop the breast mass in high-resolution mammograms. In the second step, they changed the LOGO training strategy to improve both training efficiency and segmentation performance. They achieved this by training the entire image and cropping it separately using the global and local transformer branches. They reached the ultimate segmentation choice by merging the two branches. The proposed YOLO-LOGO model was tested employing two separate mammography datasets, INbreast and CBIS-DDSM. Prior research has not been able to match the performance of the suggested model. It achieves a mean average mass detection accuracy of 65.0% on the CBIS-DDSM dataset, as well as a rate of true positives of 95.7%. Also, the approach obtained an F1 score of 74.5% as well as an IoU score of

64.0% in mass segmentation. In a different dataset called INBreast, they also got the same results.

We found that the segmentation model on the mammography dataset was not correctly applied in a previous study. Segmented images were incorporated into the training stage. In addition, no one has employed an enhanced transformer model, which would have improved the precision of the procedure. For mammogram classification at an early stage in the affected region, our proposed methodology achieves greater precision.

Materials and methods

This section describes our proposed approach for correctly identifying the two categories of benign and malignant masses in the BC dataset. The main steps of the process are shown in Fig. 1. Our methodology, which consists of evaluating the PTr model for mammogram image analysis and comparing its performance via a segmentation technique with well-known CNN models. We chose PTr and ViT [49] as the transformer model for this study, and compared its performance with other popular pre-trained models, such as EfficientNetB7 and MobileNetV3. To maximize the performance of the models, preprocessing, and image data segmentation [50] processes were applied to the dataset prior to model training on these architectures. These steps are necessary to provide

reliable and accurate classification of BC using mammography dataset.

Image data preprocessing

Data Preprocessing enhances the performance of ML models. It helps with feature improvement, noise reduction, and normalization and standardization of pixel values, all of which help the model become more broadly applicable. Preprocessing also makes it easier to uncover important patterns from the images to make prediction accurately. It also aids in overcoming challenges such as class disparities. This collection contained 3,878 images, not all of which are properly focused. There are also many images of different sized. We applied a cropping and resizing technique that reduces the image's size to 224×224 pixels while maintaining any noteworthy disease areas to guarantee consistency. This size works for every model employed in the process.

Architecture of segmentation anything model (SAM)

SAM [51] begins by converting the concept of a prompt from natural language processing (NLP) [52] to segmentation. A prompt can be any information that indicates what to segment in an image, such as a group of foreground or background points, a rough box or mask or any combination of these. A natural pre-training
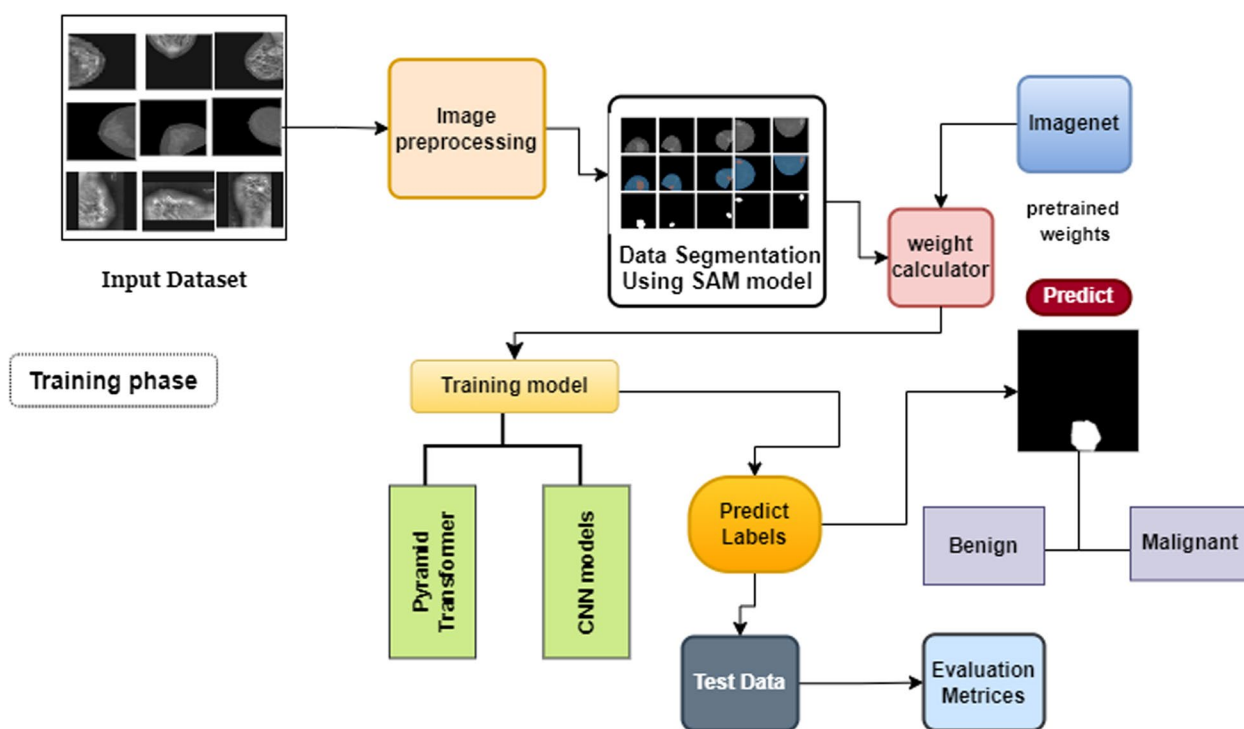

Fig. 1 The suggested methods entail the following steps: (i) Loading training data (ii) Preprocessing images (iii) SAM-based segmentation of the preprocessed images (iv) Training dataset using the Transformer and CNN models (v) Predict mammograms class (vi) Best Model evaluation

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 7 of 18

technique that mimics a series of prompts for every training sample and contrasts the model's mask predictions with the actual truth is suggested by the promptable segmentation task. In Fig. 2 illustrates the architecture of the SAM model. Also an explanation of the blocks the SAM architecture, including important settings and parameters inside these vital parts have described in Table 1. It provides a detailed summary of this data, including an in-depth examination of the operators, strides, channels, and associated layers.

Image encoder, The pretrained ViT, which has been slightly modified to handle precise supplies, is used by the image encoder. The image encoder can be applied before activating the model; it executes once for each image. The ViT-$\frac{H}{32}$ variation, with eight equally spaced global attention blocks and a 30×30 attention window [53], is specifically used. An embedding of the input image that has been downsampled by a factor of 32 is the image encoder's results.

Prompt encoder, Depending on the type of prompt, the prompt encoder uses a different encoding process to turn the prompts into vectors with features. Both tiny

Table 1 An explanation of the blocks the SAM architecture, including important settings and parameters inside these vital parts

Block	Operator	Strides	Channel	Layer
SERes	Conv, 64×64	1	32	256
SAM encoder	Conv, 64×64	1	32	1024
SERes + Unsampling	Conv, 128×128	2	64	128
	Conv, 256×256	2	16	64

points, boxes with bounds, the text or masks prompts are possible.

Mask decoder, The image embedding, prompt embedding, and output token are successfully mapped to a mask via the Transformer-based two-layer decoding that makes up the mask decoder [54]. The decoder component updates all embeddings by using self-attention and cross-attention in two directions. Following the execution of two blocks, the output identifiers are transformed into the Multilayer Perceptron (MLP) of the adaptable
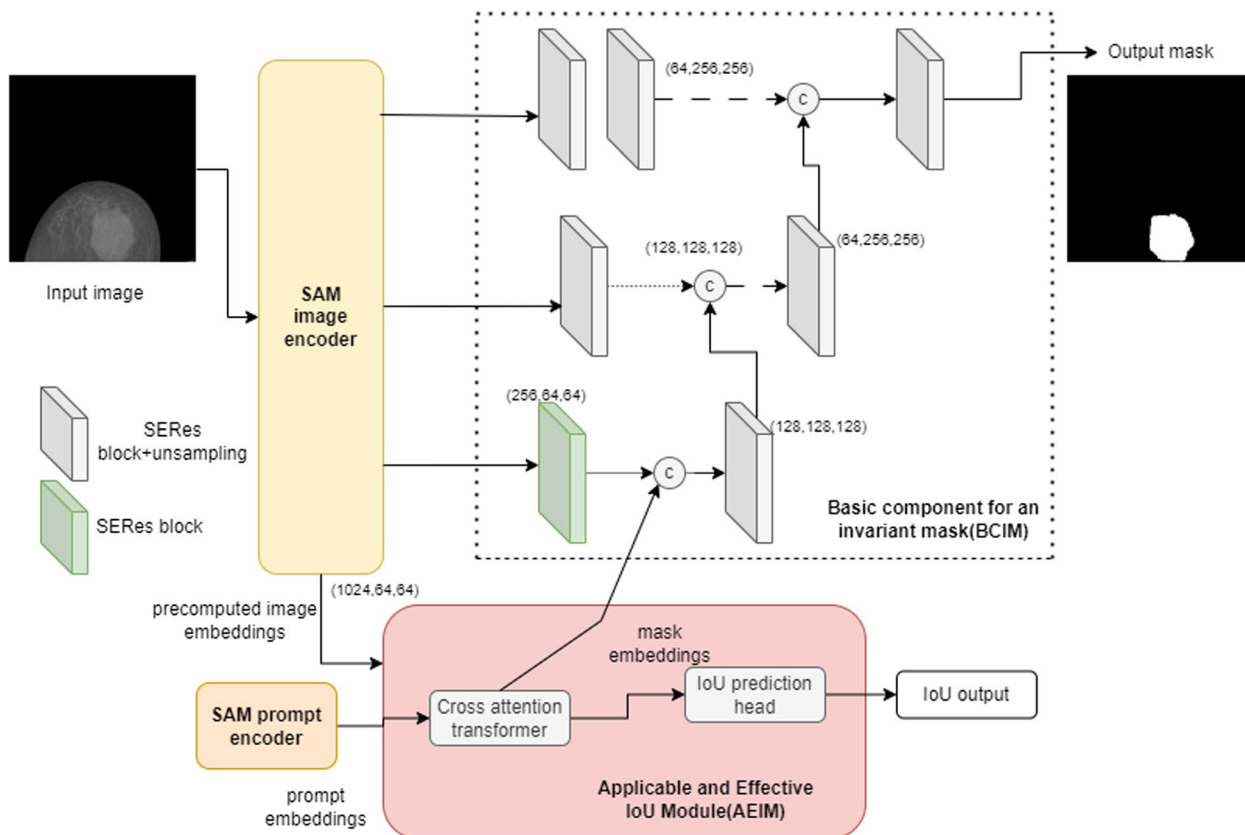

Fig. 2 The basic component of an invariant mask (BCIM) can be used to fuse image incorporation and mask-embedded data, resulting in a mask. This can be achieved by using the SAM in conjunction with an appropriate and efficient IoU module (AEIM) to construct mask embeddings in response to prompts

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 8 of 18

linear method, the mask chances for each image are computed, and the visual consolidation is unsampled.

Improving robustness over various prompts, in prompt mode, the final segmentation result is solely dependent on the prompt, and the model is still more sensitive to inaccurate prompts.Partitioning the SAM mask decoder [55] into two components was the goal: the efficient and suitable IoU module (AEIM), which creates mask embeddings based on the given prompt. Apart from the basic component of the invariant mask (BCIM), which creates the mask by fusing the mask embeddings from the AEIM with the image-embedded data from the image encoder. Employing a cross-attention transformer to calculate the variance of SAM predictions and a variety of multibox prompts to produce unique predictions. This model can generate unresolved maps that show where segmentation is challenging. These maps enable further clinical investigation and offer vital insight regarding possible segmentation errors.

Hyperparameter tuning for segmentation anything model

We performed an investigation to fine-tune the SAM's hyperparameters for the segmentation task. We examined CNN's decay, optimizer, learning rate, dropout, batch size, stride, and padding, among other parameters. The purpose of this experiment was to evaluate the effects of various factors on our model. The hyperparameter-improving experiment of the SAM model is presented in Table 2. We used the decay of 0.0001 and a learning rate of $2e-4$ which is suitable for mammography segmentation task. In addition to the batch size of 64 stride of 1×1 and the dropout rate was 0.2. The output of modified SAM model is a significance contribution in this research.

Dataset splitting

The procedure for dividing a dataset is crucial for maintaining its state of balance. Before beginning the training process, we separated the data into training and testing

Table 2 Experiment to tune hyperparameters for segmentation anything model

Parameter	Search area	Chosen value
Decay	[0.001, 0.0001]	0.0001
Optimizer	[Adam, RMSprop]	Adam
Learning rate	[1e-4, 2e-4]	2e-4
Dropout	[0.2, 0.1]	0.2
Batch size	[32,64,128,256]	64
Stride	[1 × 1, 2 × 2]	1 × 1
Padding	[Same, Valid]	Same

sets. The ratios are 85% and 15% for training and testing of the full dataset, respectively. Consequently, 415 images were used for the model's testing while 2355 images from the complete dataset were used for training.

Proposed method

In this work, we used transformer-based CNN technique to identify benign and malignant tissues in mammograms. To classifying mammograms, an improved PTr model that used TL approach [24] was employed in this study. We implemented a successful architecture modification at this bottom layer of PTr to identify BC from mammography images. We preprocessed and segmented the data before applying the models in training mode. Two of the four models we implemented EfficientNetB7 [56] and MobileNetV3 [57] are pre-trained models. Other variants of transformer model ViT were also employed in this study, which achieved higher accuracy in computer vision tasks. We used the mammography images to train all of the models to identify the top model, which will aid in more accurate illness prediction.

Pyramid Transformer (PTr)

In computer vision, a recent development that puts CNN architectures to challenge is the transformer model [58]. Utilizing hierarchical structures, PTr gathers information at several levels and is optimized for image classification applications. The study [49] first presented the vit architecture of a transformer based. PTr's primary job is to separate the input image into distinct local windows at skewed frame angles. The framework can accurately gather data at both the regional and global levels because of the separate handling of each frame. The transformer layers in the model's design are stacked hierarchically, allowing it to gather data at different geographic resolutions.

PTr primary surface: In this section the basic concept of PTr [59] is to break down the given image into distinct local frames of varying sizes, or what are known as "skewed frames". Because every aperture is treated independently it can gather local and global data. Figure 3 describes a thorough description of the PTr design structure for the foundation version of the TL approach. The dimension of the origin image was set at 224×224 because that was the size that PTr. which had been trained previously and optimized needed. Additionally, a given RGB image with source parameters of H × W × 3 was divided into smaller sections with lengths equal to 3 × 3, starting with the first patch measurement in the original approach. Applying several PTr sections with modified self-attention to these linear patch insertions of size C preserves the number of tokens at approximately $\frac{H}{3} \times \frac{W}{3}$.
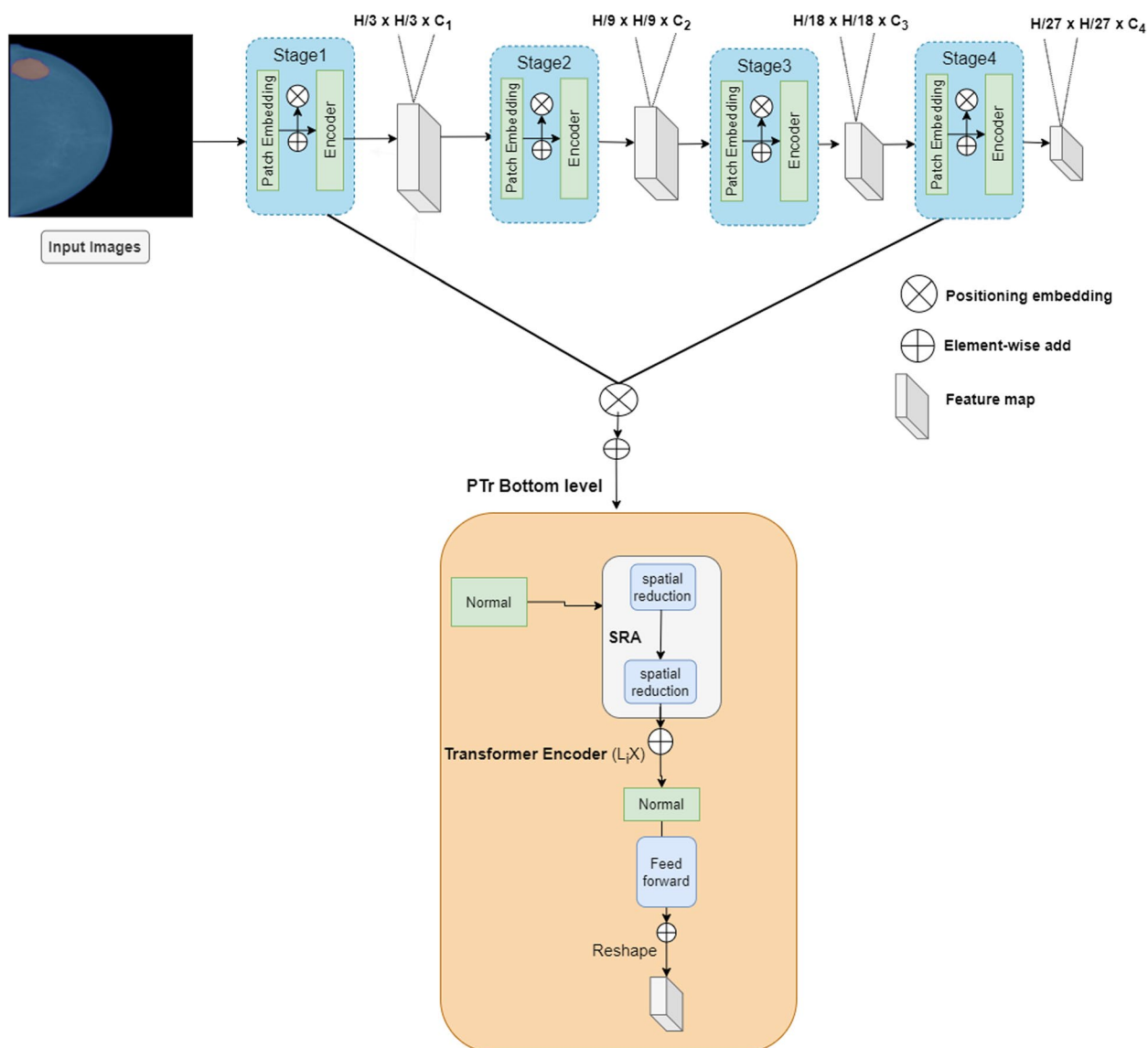
Fig. 3 The architecture of the pyramid transformer (PTr) using TL approach is represented by the letters H, W, C, Pi, i-th stage patch size, F, feature map, L_i, transformer-encoder layer, and SRA, spatial-reduction attention

Second step: The patch fusion layer concatenates the features of each pair of contiguous two-by-two patches, and the 6C spatial concatenated characteristics are subjected to a linear level. Consequently, the number of different patches multiplied, and 3C is the final depth of the nonlinear layer. Additionally, PTr modules are used to transform the features, and the number of Stage 2 outcome patches was maintained at $\frac{H}{9} \times \frac{W}{9}$.

PTr bottom level: PTr uses SRA, a type of individual attention, to calculate the linear magnitude of the neural network. A decrease in both keys and values is its defining feature [60]. The SRA gradually reduced the physical proportions of its features throughout the model.

This further reinforced the idea of order by giving each transformer component an embedding position. In this bottom layer of PTr, we successfully modified the architecture to detect BC from mammogram images.

Here we utilized an enhanced PTr to train the mammography using TL [61] approach. The objective was to use transformers from the big natural image collection to classify mammograms into two groups: benign and malignant images. To do this, we removed the pre-trained prediction head and installed a $D \times K$ feedforward layer, where $K = 4$ represents the entire number of classes downwards. Here, we used TL to improve [62] the target function $f_t(.)$ learning in the desired domain

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 10 of 18

Table 3 Parameters and environment requirements

Environment	Parameters
Platform	Google Colab
Language version	Python 3.0
Graphics card	TPU, 12.7 GB
Available RAM	12 GB
Disk space	19.2 GB

D_t by applying knowledge gathered from the learning task T_s and the original domain D_s. There are m samples for training $((x^1 \ y^1)..., (x^i \ y^i) ....,(x^m \ y^m))$ in the ImageNet dataset, where x_i and y_i present the $i_t h$ input and label, respectively. Subsequently, TL was performed to create W_1 using the weights of the ImageNet pretrained transformer W_0 as the base, where b is the bias. Here we have mentioned the softmax output function in Eq. (1) and the objective function of TL approach is described in Eq. (2).

$$Softmax \ output = \left(y^{i,j} \middle| x^{i,j}, w_0, w_1, b \right) \tag{1}$$

$$J(w_1, b|w_0) = \frac{-1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{m} y^{ij} log \left(p \left(softmax \ output \right) \right) \tag{2}$$

Experimental evaluation and results
Implementation details
Using the Adam optimizer and a learning rate of 0.001, we trained our models across 40 epochs. An exponential decline and batch count of 16 were used for training. We used an 8.50:1.50 ratio to divide the data into training and testing groups, respectively. GELU was employed as a means of activation with an L2 regularizer for PTr models. For the CNNs, ReLu was employed in conjunction with an L2 regularizer. Identical parameter values were employed in all experiments, especially the comparison with the recommended approach, to avoid bias in the outcomes.

Environment specifications
Considerable computing power required for image analysis and classification can be provided by GPUs [63] (Graphics computing Units). A GPU installation is more expensive and requires more gear to support the processing activity. Consequently, we trained our model on the Google Colab platform, which provides access to powerful cloud GPUs. The configuration and specification of the environment we used for our research are listed in Table 3. Google Colab comes with a v3 TPU chip that has two TensorCores, available RAM of 12 GB, and 19.2 GB

of disc space. DL models can be trained in a large-scale computer environment using these criteria.

Dataset
In this research, we used the INbreast dataset [64] to train and evaluate our Transformer and CNN model based on early BC detection. A total of 7632 images of benign and malignant tissues were included in the dataset. Each image had 227x227 pixels and was RGB. Due to the limitation of disk space in Google Colab, we used 2770 INbreast images in this research, of which 1410 images indicated malignant masses and 1360 scans indicated benign masses. In addition, many images contain noise and are incorrectly focused. Figure 4 shows the sample images taken from the dataset. The distribution of the dataset is presented in Table 4[1].

Performance evaluation metrics
This section is considered when assessing how well the system's examined trials work. A model is used to forecast classifications that may be true or incorrect. There are four possible formats for the output that results from classifying data related to various classifications [65]. Firstly, true positive (T_p) and true negative (T_n) indicate that all predictions are accurate, regardless of whether they are true or wrong. However, there may be an additional circumstance in which the prediction holds true in principle but not in practice, or vice versa. False positive (F_p) and false negative (F_n) terms are used to describe these two situations. In addition, we may be able to ascertain our models' classification performance by computing more exact metrics from the confusion matrix.

Accuracy: This was defined as the entire test session and the size of the correctly identified samples.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{3}$$

Precision: The number of expected positive samples and tests that have been rightfully declared positive are expressed.

$$Precision = \frac{T_p}{T_p + F_p} \tag{4}$$

Recall: Recall is the number of positive samples that might be appropriately classified as positive and legitimately identified.

$$Recall = \frac{T_p}{T_p + F_n} \tag{5}$$

---

1 https://data.mendeley.com/datasets/ywsbh3ndr8/2

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334
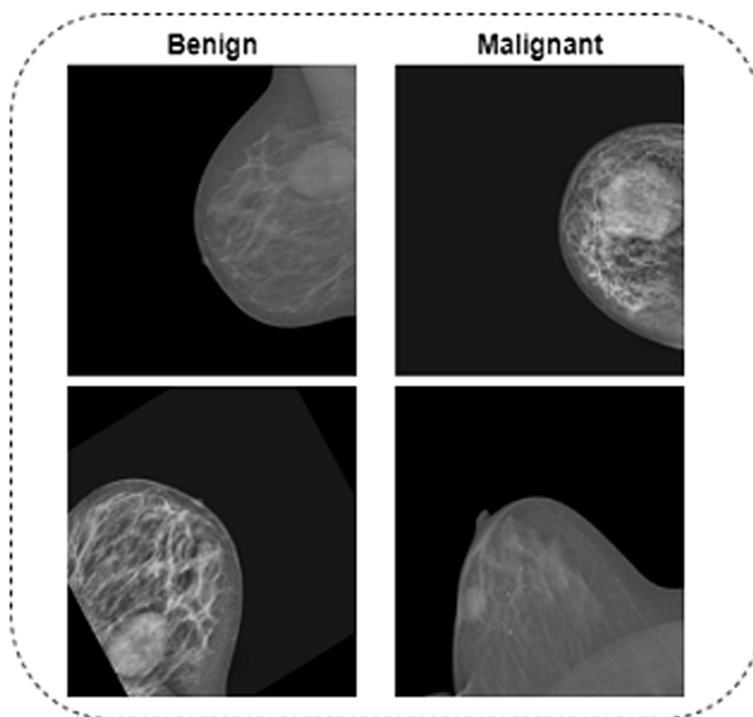
Page 11 of 18

Fig. 4 Sample images of the INbreast dataset used in this study

Table 4 Distribution of the dataset (INbreast) using in this research

Class	Training images	Testing images	Total images
Benign masses	1156	204	1360
Malignant masses	1199	211	1410
Total	**2355**	**415**	**2770**

F1 Score: The F1 score results from a linear combination of accuracy and recall, acting as a single assessment.

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

To evaluate models for predictions, accuracy, and F1-score might not always be sufficient. As a result, the evaluation procedure uses the receiver operating characteristics (ROC) curve as a secondary criterion. A curve known as the ROC curve yields the AUC. The ROC was determined by comparing the true positive rate (TPR) with the false positive rate (FPR). This formula determines the FPR and the recall is the sole part of the TPR.

$$\text{FPR} = \frac{F_p}{F_p + T_n} \tag{7}$$

Findings of SAM model

We individually labeled the diseases in 300 images of each class that we tracked. Appropriately annotated images were used to train the SAM model [66]. Using the Cvat website https://www.cvat.ai/, we manually annotated 300 images of each class. We then used the SAM model to train the images after annotation to more accurately identify the ground truth mask. The model successfully determined the ground truth for each image after the training. Next, we used all the segmented images to train the proposed Transformer and CNN models. Figure 5 shows the segmentation results. Figure 5a define the imput images from mammogram dataset. The images were then trained using the SAM model shown in Fig. 5b. Finally, we obtained the truth mask from the segmented images as shown in Fig. 5c. The training and validation accuracy of the SAM model has been shown in Fig. 6a. The validation accuracy of this model was 97.5%, which was promising. Figure 6b shows the training and validation loss of the SAM model. We observed that the training loss has decreased as the number of epochs increased. The outcome demonstrates that the model can effectively emphasize the mammography cancer tissues portion of the images and mask all the superfluous pixels.

The training and validation outcomes of the SAM model are presented in Table 5. Out of the twenty epochs utilized to train and validate the SAM model in this
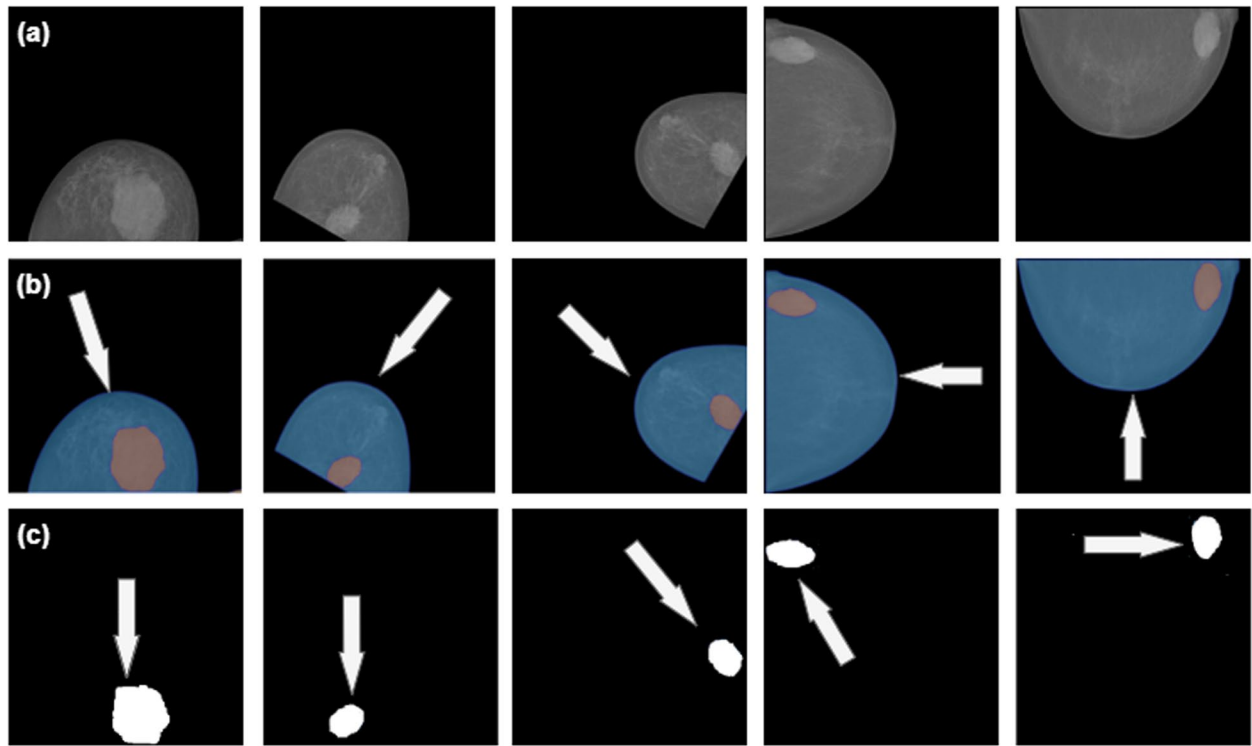
Fig. 5 Segmented images using trained SAM model. **a** Input images (**b**) Trained images using SAM model (**c**) Find Ground truth mask
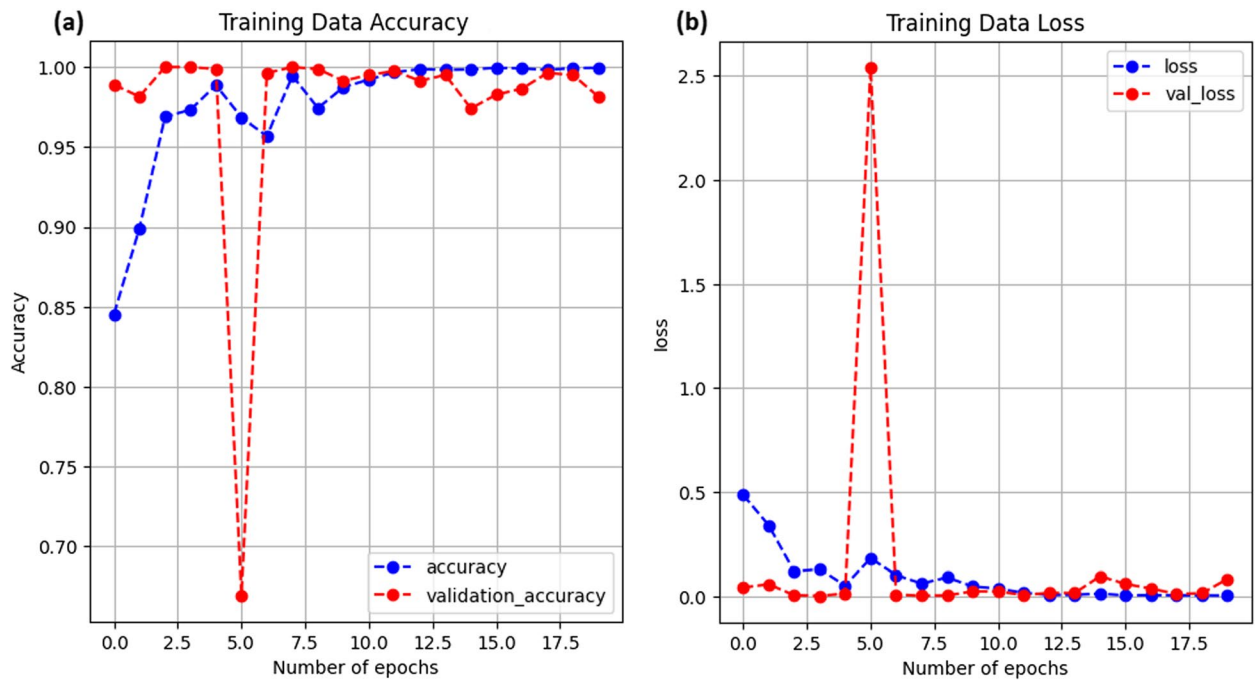

Fig. 6 Segmentation results using the SAM model, (**a**) Training accuracy and validation Accuracy, and (**b**) Training loss and validation loss

Table 5 The segmented results for the last ten epochs (11–20). Using the SAM model, we presented the training accuracy, validation accuracy, and validation loss

Epoch number	Training accuracy %	Validation accuracy %	Validation loss %
11	98.86	99.53	0.26
12	99.93	99.81	0.20
13	99.96	99.00	0.94
14	99.99	99.27	0.87
15	99.99	97.28	1.89
16	99.99	98.40	1.21
17	99.99	98.92	1.05
18	99.99	99.29	0.86
19	99.99	99.65	0.28
20	99.99	97.86	1.74

study, we highlighted the results from the last ten. With a minimal loss of 0.20%, we achieved the highest validation accuracy of 99.81% in the twelve-number epoch. After completing twenty epochs, the SAM model achieves a higher validation accuracy of 97.50%, which is an optimistic finding in this study.

Result analysis

Here, we provide the findings from all the evaluated models used in this research. The PTr design was proposed to examine the resilience of the models. Furthermore, this study used ViT, EfficientNetB7 and MobileNetV3 to assess and compare the efficacy of our recommendations. Finally, we provide a performance analysis of the proposed architectures, complete with multiple graphs and charts.

Classifiers performance matrix of suggested models

The classifier performance of the model is displayed in a table-like in a similar manner. Table 6 presents the categorizing outcomes of our proposed systems, PTr, Vit, EfficientNetB7, and MobileNetV3, which utilize data segmentation techniques in the mammography dataset. Table 6 shows that PTr performs exceptionally well in mammography classification, which is a significant finding of the present study.

In either case, it shows the forecasts outcome using the test set data. The actual class is shown in rows, while the forecasting for each class is shown in columns. Conversely, the inclination of the matrix shows how many images were successfully identified. The confusion matrix of the two DL models and the proposed PTr method utilized in this architecture is shown in Fig. 7. The confusion matrix is shown in Fig. 7a EfficientNetB7, b MobileNetV3, and c proposed PTr models generated for the

binary classification. Consequently, it is clear that the proposed research has created a high level of categorization accuracy. Compared with the DL models and all other previously proposed models, PTr performed better. An effective method for classifying mammography images is the transformer-based model. Despite its exceptional performance, the metrics of this model were lower than those of the Transformer system for the INbreast dataset.

Classifier parameter results

The ROC shown in Fig. 8, is a supplementary compliance indicator for the presented models. The highest documented micro average AUC score for the PTr model is 0.998 with 0.988, and the EfficientNetB7 model possesses the second-most nearly ROC. In our proposed methodology, the proposed model exhibited exceptional performance.

The comprehensive report of classification considering the mean accuracy, recall, and F-1 score for each approach is shown in Table 7. In this instance, PTr outperformed every other model that we explored. A precision of 99.97% indicated exceptional success using this method.

Prediction accuracy and loss

By applying our proposed techniques, this section displays the accuracy and loss of each of the four models. The concepts of accuracy and loss are presented in Table 8 for the evaluation and validation of the dataset. With a 0.17% loss, the PTr model using the data segmentation approach exhibited the best testing accuracy, of 99.87%. At 99.98%, PTr's highest accuracy is achieved while verifying the data. Furthermore, it proved to be incredibly accurate in testing and validation. MobileNetV3 had the lowest test accuracy of 94.0% and highest loss score of 1.23% when compared to the other methods.

A line chart showing both the precision and loss of the 40 epochs of the PTr and ViT models is shown in Fig. 9. Figure 9a illustrates the relatively high consistency and reliability of the PTr designs. Figure 9b shows the loss and accuracy for the ViT Model. Similar to PTr, the present study achieved high and stable accuracy scores for every epoch by employing the segmentation approach.

Table 9 also compares the most recent research on mammography dataset using segmentation techniques in several publications on the prediction and classification of mammograms. Regarding the AUC values, this research demonstrated that our transformer-based model with segmentation techniques using the TL approach performed better than any prior study. Our proposed model performed better in this specific section than in any other study.
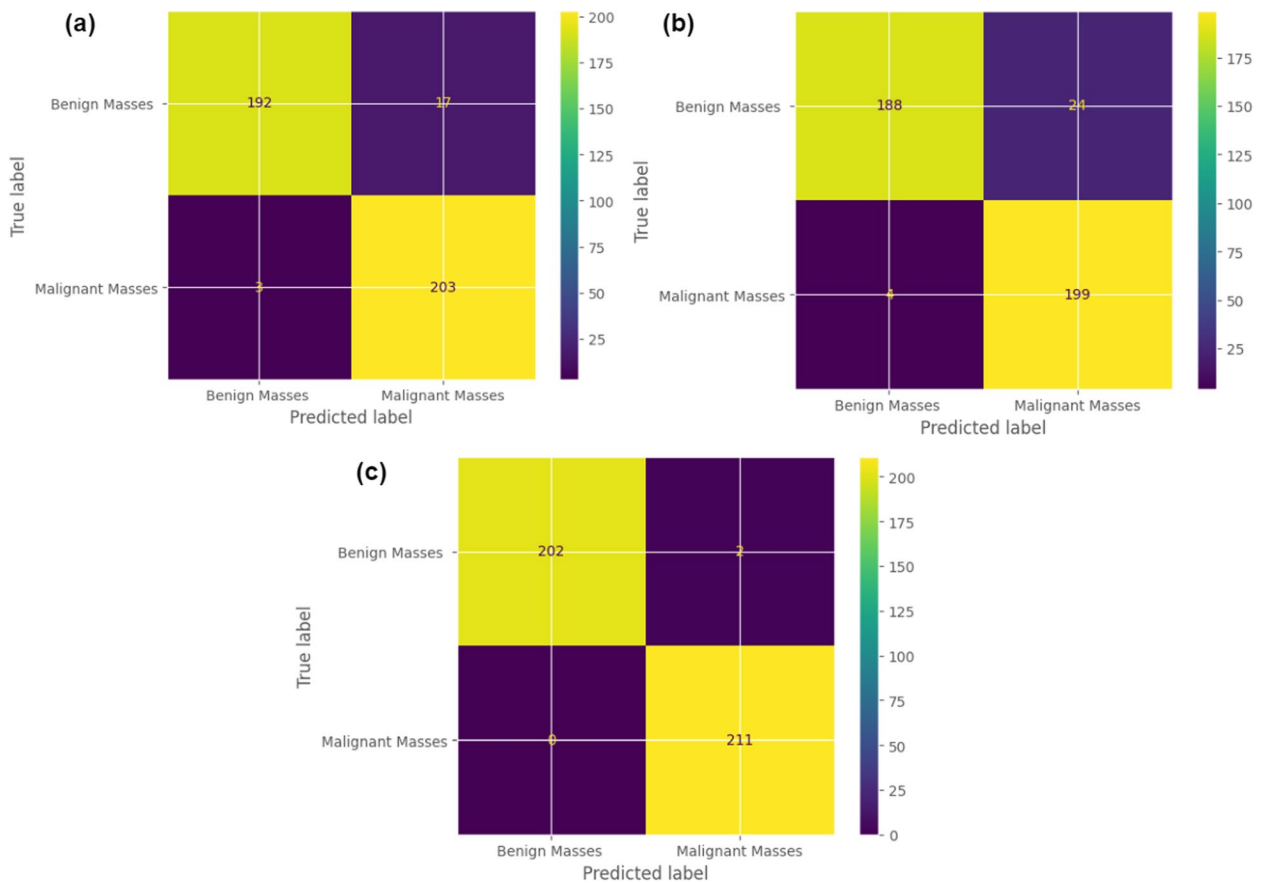
Fig. 7 Confusion matrix for correct and incorrect predictions made during the testing phase (**a**) EfficientNetB7, (**b**) MobileNetV3, and **c** proposed PTr methods on the Inbreast dataset
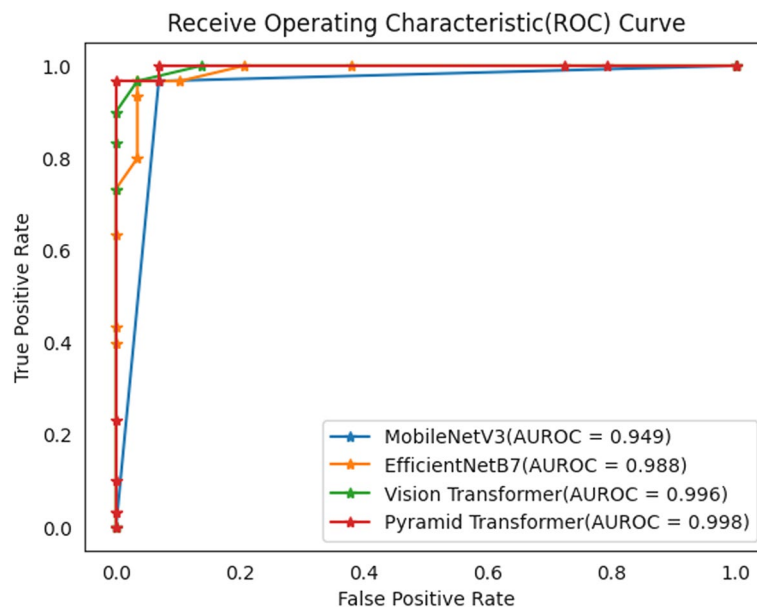

Fig. 8 ROC curve for all evaluated methods in mammography classification using segmentation techniques. Here are the micro areas under the ROC curve for the PTr, ViT, EfficientNetB7 and MobileNetV3 representations
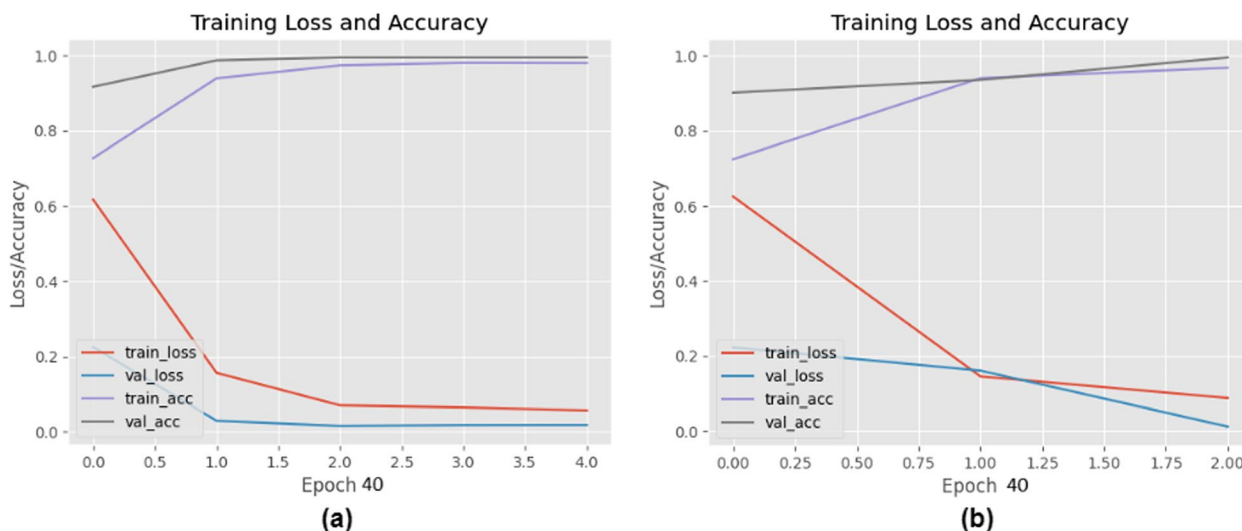
Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 15 of 18

Fig. 9 Training loss and accuracy of the transformer models used in this research, (**a**) Proposed pyramid transformer method and (**b**) Vision transformer model

Table 6 Class-wise classification results for PTr, ViT, EfficientNetB7, and MobileNetV3. The F1-score, Precision, and Recall of the Transformer and CNN models using data segmentation approaches, as well as the evaluation metrics values for each class, are displayed

Model	Class	Recall %	Precision %	F1-score %
PTr	Benign	99.96	99.00	99.95
	Malignant	99.00	99.98	99.99
ViT	Benign	99.00	99.00	99.00
	Malignant	98.00	99.00	99.00
EfficientNetB7	Benign	98.00	92.00	95.00
	Malignant	92.00	99.00	95.00
MobileNetV3	Benign	95.00	95.00	95.00
	Malignant	94.00	94.00	94.00

Table 7 Extensive report on classification using segmentation techniques. (The models PTr, ViT, EfficientNetB7, and MoileNetV3 were compared based on average precision, recall, and F1-score values)

Method	Recall %	Precision %	F1-score %
PTr	99.98	99.97	99.97
ViT	99.00	99.00	99.00
EfficientNetB7	95.00	95.00	95.00
MobileNetV3	94.00	94.00	94.00

Analysis of computational complexity

An essential trade-off between computing efficiency and model complexity is highlighted by comparing the proposed PTr with all models assessed in this study. PTr had

Table 8 Loss and accuracy for the best models

Model	Accuracy %		Loss%	
	Validation	Test	Train	Validation
PTr	99.96	99.87	0.13	0.15
ViT	98.67	98.43	0.23	0.31
EfficientNetB7	95.00	94.22	0.68	1.19
MobileNetV3	94.00	93.16	0.78	1.27

fewer layers than the other DL models except for Mobile-NetV3. Owing to their layers, they are intrinsically more complicated and result in more parameters. Table 10 illustrates that while MobileNetV3 has 28 layers and PTr has only 16 layers, these layers include a substantial number of parameters (21.5 million and 13 million, respectively) because of their architecture. They required less training time than the other models used in our research.

It is necessary to highlight that the PTr models performed exceptionally well in this study and required less computational power. As a result, there is a conflict between processing power and model performance with other DL models. The resultant compromise must be carefully considered by researchers and practitioners in light of their unique needs and limitations.

Conclusion and future work

Researchers worldwide are currently developing methods for screening BC patients at an early stage to minimize the high death rate associated with BC in women. Therefore, early detection tools for BC should be precise and cost-efficient. For that reason, researchers

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 16 of 18

Table 9 The efficacy of the suggested approach was compared with an earlier study that employed segmentation techniques to improve the accuracy of mammography class prediction on mammography datasets

Previous study	Dataset	Segmentation (Yes/No)	Purpose	Best method	AUC (score)
Garrucho et al. [39]	OPTIMAM	No	Detection	Defor-DETR	0.948
Li et al. [31]	DDSM	No	Classification	TV-NN	0.968
Al-masni et al. [32]	DDSM	No	Detection and Classification	CNN	0.97
Ragab et al. [23]	DDSM	Yes	Detection	DCNN-SVM-AlexNet	0.88
Salama et al. [24]	DDSM	Yes	Classification	InceptionV3	0.989
Khamparia et al. [38]	DDSM	Yes	Detection	Hybrid MVGG16 ImageNet	0.933
Das et al. [33]	INbreast	No	Detection	Xception	N\A
Saffari et al. [34]	INbreast	Yes	Classification	CNN	N\A
Hama et al. [35]	INbreast	Yes	Detection and Classification	DenseNet121	N\A
Proposed Model	**INbreast**	**Yes**	**Prediction and Classification**	**Pyramid Transformer**	**0.998**

Table 10 Complexity analysis of all evaluated models

Points	PTr	ViT	MobileNetV3	EfficientNetB7
Parameters (millions)	21.5	22.8	13	30
Layers	16	33	28	347
RAM used (GB)	4.5	6.7	3.1	11.3
GPU used (GB)	6.4	10.6	6.3	13
Training time (hours)	3	4.5	2	5

have focused on DL-based BC detection by using mammography data. Based on current research trends, we present an improved transformer model using segmentation techniques for BC detection and the classification of mammography images. Accurate identification of BC-affected regions was achieved using data segmentation approaches based on SAM architecture. In the segmentation challenge, the SAM model achieved an accuracy of 97.5%. At the bottom layer of the PTr, we successfully modified the architecture to distinguish BC from the mammography images. Finally, we compared our suggested PTr model to three existing DL models: ViT, MobileNetV3, and EfficientNetB7, and found that it performed better with an AUC score of 99.98% and an accuracy of 99.96%.

However, this result is restricted to training on a single dataset from a single source, owing to disc space limitations. Therefore, if we wish to draw broad conclusions from this experiment, it would be prudent to perform additional experiments using datasets from multiple sources. In addition, it would be beneficial for future research to examine the impact of different DL parameters on explainable artificial intelligence methods for breast mammography image classification.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Computer Science and Engineering, Stamford University Bangladesh, Siddeswari, Dhaka, Bangladesh. [2]Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. [3]Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia. [4]Department of Computer Science, American International University-Bangladesh, Kuratoli, Dhaka, Bangladesh. [5]Department of Electrical Engineering and Computer Science, Texas A&M University-Kingsville, Kingsville 78363, Texas, USA.

References
1.   Ponraj DN, Jenifer ME, Poongodi P, Manoharan JS. A survey on the preprocessing techniques of mammogram for the detection of breast cancer. J Emerg Trends Comput Inf Sci. 2011;2(12):656–64.

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 17 of 18

2.   Rosenberg PS, Barker KA, Anderson WF. Estrogen receptor status and the future burden of invasive and in situ breast cancers in the United States. J Natl Cancer Inst. 2015;107(9):djv159.

3.   Chaurasia V, Pal S, Tiwari B. Prediction of benign and malignant breast cancer using data mining techniques. J Algoritm Comput Technol. 2018;12(2):119–26.

4.   Mohammed MA, Al-Khateeb B, Rashid AN, Ibrahim DA, Abd Ghani MK, Mostafa SA. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. Comput Electr Eng. 2018;70:871–82.

5.   Obaid OI, Mohammed MA, Ghani MKA, Mostafa A, Taha F, et al. Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer. Int J Eng Technol. 2018;7(4.36):160–6.

6.   Panigrahi L, Verma K, Singh BK. Ultrasound image segmentation using a novel multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution. Expert Syst Appl. 2019;115:486–98.

7.   Singh VP, Srivastava S, Srivastava R. Automated and effective content-based image retrieval for digital mammography. J X-ray Sci Technol. 2018;26(1):29–49.

8.   Abdul Halim AA, Andrew AM, Yasin N, Abd Rahman MA, Jusoh M, Veeraperumal V, et al. Existing and Emerging Breast Cancer Detection Technologies and Its Challenges: A Review. Appl Sci. 2021;11(11):10753. https://doi.org/10.3390/app112210753.

9.   Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, et al. Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. Cancer. 2018;124(13):2785–800.

10. Borchartt TB, Conci A, Lima RC, Resmini R, Sanchez A. Breast thermography from an image processing viewpoint: a survey. Signal Process. 2013;93(10):2785–803.

11. Pedro RWD, Machado-Lima A, Nunes FL. Is mass classification in mammograms a solved problem?-a critical review over the last 20 years. Expert Syst Appl. 2019;119:90–103.

12. Kooi T, Litjens G, Van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal. 2017;35:303–12.

13. Zhong J, Wang L, Yan C, Xing Y, Hu Y, Ding D, et al. Deep learning image reconstruction generates thinner slice iodine maps with improved image quality to increase diagnostic acceptance and lesion conspicuity: a prospective study on abdominal dual-energy CT. BMC Med Imaging. 2024;24(1):159.

14. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.

15. Tsochatzidis L, Zagoris K, Arikidis N, Karahaliou A, Costaridou L, Pratikakis I. Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach. Pattern Recognit. 2017;71:106–17.

16. Moitra D, Mandal RK. Automated AJCC staging of non-small cell lung cancer (NSCLC) using deep convolutional neural network (CNN) and recurrent neural network (RNN). Health Inf Sci Syst. 2019;7:1–12.

17. Talo M, Baloglu UB, Yıldırım Ö, Acharya UR. Application of deep transfer learning for automated brain abnormality classification using MR images. Cogn Syst Res. 2019;54:176–88.

18. Poudel P, Illanes A, Sadeghi M, Friebe M, Patch based texture classification of thyroid ultrasound images using convolutional neural network. In: 2019 41st Annual international conference of the ieee engineering in medicine and biology society (EMBC). IEEE; 2019. pp. 5828–31.

19. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020. arXiv:2010.11929.

20. Gheflati B, Rivaz H, Vision transformers for classification of breast ultrasound images. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. pp. 480–3.

21. Srikantamurthy MM, Rallabandi VS, Dudekula DB, Natarajan S, Park J. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Med Imaging. 2023;23(1):19.

22. Maqsood S, Damaševičius R, Maskeliūnas R. TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages. Appl Sci. 2022;12(7):3273.

23. Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ. 2019;7:e6201.

24. Salama WM, Aly MH. Deep learning in mammography images segmentation and classification: Automated CNN approach. Alex Eng J. 2021;60(5):4701–9.

25. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. Sci Rep. 2019;9(1):12495.

26. Murtaza G, Shuib L, Abdul Wahab AW, Mujtaba G, Mujtaba G, Nweke HF, et al. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artif Intell Rev. 2020;53:1655–720.

27. Park GE, Kang BJ, Kim SH, Lee J. Retrospective review of missed cancer detection and its mammography findings with artificial-intelligence-based, computer-aided diagnosis. Diagnostics. 2022;12(2):387.

28. Lotter W, Diab AR, Haslam B, Kim JG, Grisot G, Wu E, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat Med. 2021;27(2):244–9.

29. Dar RA, Rasool M, Assad A, et al. Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. Comput Biol Med. 2022;149:106073.

30. Frazer HM, Qin AK, Pan H, Brotchie P. Evaluation of deep learning-based artificial intelligence techniques for breast cancer detection on mammograms: Results from a retrospective study using a BreastScreen Victoria dataset. J Med Imaging Radiat Oncol. 2021;65(5):529–37.

31. Li H, Niu J, Li D, Zhang C. Classification of breast mass in two-view mammograms via deep learning. IET Image Process. 2021;15(2):454–67.

32. Al-Masni MA, Al-Antari MA, Park JM, Gi G, Kim TY, Rivera P, et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Prog Biomed. 2018;157:85–94.

33. Das HS, Das A, Neog A, Mallik S, Bora K, Zhao Z. Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach. Front Genet. 2023;13:1097207.

34. Saffari N, Rashwan HA, Abdel-Nasser M, Kumar Singh V, Arenas M, Mangina E, et al. Fully automated breast density segmentation and classification using deep learning. Diagnostics. 2020;10(11):988.

35. Soltani H, Amroune M, Bendib I, Haouam MY, Benkhelifa E, Fraz MM. Breast lesions segmentation and classification in a two-stage process based on Mask-RCNN and Transfer Learning. Multimed Tools Appl. 2024;83(12):35763–80.

36. Jaehwan L, Donggeun Y, Hyo-Eun K. Photometric transformer networks and label adjustment for breast density prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul: IEEE; 2019. p. 0.

37. Comes MC, Fanizzi A, Bove S, Didonna V, Diotaiuti S, La Forgia D, et al. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs. Sci Rep. 2021;11(1):14123.

38. Van Tulder G, Tong Y, Marchiori E. Multi-view analysis of unregistered medical images using cross-view transformers. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24. Springer; 2021. pp. 104–113.

39. Tummala S, Kim J, Kadry S. Breast-net: Multi-class classification of breast cancer from histopathological images using ensemble of swin transformers. Mathematics. 2022;10(21):4109.

40. Khamparia A, Bharati S, Podder P, Gupta D, Khanna A, Phung TK, et al. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. Multidim Syst Sign Process. 2021;32:747–65.

41. Garrucho L, Kushibar K, Jouide S, Diaz O, Igual L, Lekadir K. Domain generalization in deep learning based mass detection in mammography: A large-scale multi-center study. Artif Intell Med. 2022;132:102386.

42. Wang W, Jiang R, Cui N, Li Q, Yuan F, Xiao Z. Semi-supervised vision transformer with adaptive token sampling for breast cancer classification. Front Pharmacol. 2022;13:929755.

43. Dey S, Roychoudhury R, Malakar S, Sarkar R. Screening of breast cancer from thermogram images by edge detection aided deep transfer learning model. Multimed Tools Appl. 2022;81(7):9331–49.

Kumar Saha *et al. BMC Medical Imaging*   (2024) 24:334

Page 18 of 18

44. Boumaraf S, Liu X, Zheng Z, Ma X, Ferkous C. A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. Biomed Signal Process Control. 2021;63:102192.

45. Chougrad H, Zouaki H, Alheyane O. Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomputing. 2020;392:168–80.

46. Shobayo O. Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images. 2024.

47. Khan A, Lee B. DeepGene transformer: Transformer for the gene expression-based classification of cancer subtypes. Expert Syst Appl. 2023;226:120047.

48. Su Y, Liu Q, Xie W, Hu P. YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms. Comput Methods Programs Biomed. 2022;221:106903.

49. Saha DK, Joy AM, Majumder A. YoTransViT: A transformer and CNN method for predicting and classifying skin diseases using segmentation techniques. Inform Med Unlocked. 2024;47:101495.

50. Müller D, Kramer F. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. BMC Med Imaging. 2021;21:1–11.

51. Osco LP, Wu Q, de Lemos EL, Gonçalves WN, Ramos APM, Li J, et al. The segment anything model (sam) for remote sensing applications: From zero to one shot. Int J Appl Earth Obs Geoinformation. 2023;124:103540.

52. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. J Biomed Inform. 2017;73:14–29.

53. Li Y, Wang D, Yuan C, Li H, Hu J. Enhancing agricultural image segmentation with an agricultural segment anything model adapter. Sensors. 2023;23(18):7884.

54. Zhang C, Puspitasari FD, Zheng S, Li C, Qiao Y, Kang T, et al. A survey on segment anything model (sam): Vision foundation model meets prompt engineering. 2023. arXiv preprint arXiv:2306.06211.

55. Zhang Y, Shen Z. Jiao R. Segment anything model for medical image segmentation: Current applications and future directions. Comput Biol Med. 2024;171:108238.

56. Raza R, Zulfiqar F, Khan MO, Arif M, Alvi A, Iftikhar MA, et al. Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Eng Appl Artif Intell. 2023;126:106902.

57. Xi Y, Zhang W, Zhou F, Tang X, Li Z, Zeng X, et al. Transmission line fault detection and classification based on SA-MobileNetV3. Energy Rep. 2023;9:955–68.

58. Ali H, Mohsen F, Shah Z. Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review. BMC Med Imaging. 2023;23(1):129.

59. Wang W, Xie E, Li X, Fan DP, Song K, Liang D, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision, 2021. pp. 568–578.

60. Ayana G, Dese K, Dereje Y, Kebede Y, Barki H, Amdissa D, et al. Vision-transformer-based transfer learning for mammogram classification. Diagnostics. MDPI; 2023;13(2):178.

61. Junzhe Z, Fuqiang J, Yupeng C, Weiyi W, Qing W. A water surface garbage recognition method based on transfer learning and image enhancement. Results Eng. 2023;19:101340.

62. Wee CY, Liu C, Lee A, Poh JS, Ji H, Qiu A, et al. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. NeuroImage: Clin. 2019;23:101929.

63. Krzywaniak A, Czarnul P, Proficz J. Dynamic GPU power capping with online performance tracing for energy efficient GPU computing using DEPO tool. Futur Gener Comput Syst. 2023;145:396–414.

64. Raiaan MAK, Fahad NM, Mukta MSH, Shatabda S. Mammo-light: a light-weight convolutional neural network for diagnosing breast cancer from mammography images. Biomed Signal Process Control. 2024;94:106279.

65. Saha DK. An extensive investigation of Convolutional Neural Network designs for the diagnosis of Lumpy skin disease in Dairy Cows. Heliyon. 2024;10.

66. Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y. Segment anything model for medical image analysis: an experimental study. Med Image Anal. 2023;89:102918.

Publisher's Note