Applied Intelligence (2024) 54:3448–3463
https://doi.org/10.1007/s10489-024-05358-5

A self-supervised learning model based on variational autoencoder for limited-sample mammogram classification

Meryem Altin Karagoz[1,2,3] · **O. Ufuk Nalbantoglu**[2,3,4]

Accepted: 21 February 2024 / Published online: 4 March 2024

Abstract

Deep learning models have found extensive application in medical imaging analysis, particularly in mammography classification. However, these models encounter challenges associated with limited annotated mammography public datasets. In recent years, self-supervised learning (SSL) has emerged as a noteworthy solution to addressing data scarcity by leveraging pretext and downstream tasks. Nevertheless, we recognize a notable scarcity of self-supervised learning models designed for the classification task in mammography. In this context, we propose a novel self-supervised learning model for limited-sample mammogram classification. Our proposed SSL model comprises two primary networks. The first is a pretext task network designed to learn discriminative features through mammogram reconstruction using a variational autoencoder (VAE). Subsequently, the downstream network, dedicated to the classification of mammograms, uses the encoded space extracted by the VAE as input through a simple convolutional neural network. The performance of the proposed model is assessed on public INbreast and MIAS datasets. Comparative analyzes are conducted for the proposed model against previous studies for the same classification task and dataset. The proposed SSL model demonstrates high performance with an AUC of 0.94 for density, 0.99 for malignant-nonmalignant classifications on INbreast, 0.97 for benign-malignant, 0.99 for density, and 0.99 for normal-benign-malignant classifications on MIAS. Additionally, the proposed model reduces computational costs with only 228 trainable parameters, 204.95K FLOPs, and a depth of 3 in mammogram classification. Overall, the proposed SSL model exhibits a robust network architecture characterized by repeatability, consistency, generalization ability, and transferability among datasets, providing less computational complexity than previous studies.

Keywords Self-supervised learning · Mammography · Classification · Variational autoencoder

1 Introduction

Breast cancer is the most prevalent type of cancer worldwide, with 2.26 million new cases and 685,000 reported deaths in 2020, according to the World Health Organization (WHO). The early diagnosis of breast cancer is critical for effective treatment and for reducing mortality rates [1]. X-ray mammography screening serves as the predominant method for the early diagnosis and examination of suspicious lesions in breast tissue due to its high-resolution imaging capabilities and the visualization of abnormalities. However, the diagnostic process based on mammography is labor intensive, time-consuming, and demanding in its need for specialized expertise for accurate interpretation. Moreover, the manual identification of certain lesions in mammography poses challenges, particularly in glandular and highly dense breast tissue [2]. Mammography also has limitations, failing to detect at least 25% of cancers, with approximately 10-15% of cases requiring additional screening modalities such as ultrasound to identify lesions [3]. In response to these challenges, and to mitigate costs associated with supplementary screening, computer-aided diagnosis systems (CADs) have emerged as valuable decision support tools for experts [4–6].

✉ Meryem Altin Karagoz
makaragoz@cumhuriyet.edu.tr

O. Ufuk Nalbantoglu
nalbantoglu@erciyes.edu.tr

1 Department of Computer Engineering, Sivas Cumhuriyet University, Sivas, Turkey

2 Department of Computer Engineering, Erciyes University, Kayseri, Turkey

3 Artificial Intelligence and Big Data Application and Research Center, Erciyes University, Kayseri, Turkey

4 Genome and Stem Cell Center (GenKök), Erciyes University, Kayseri, Turkey

Deep learning has gained substantial attention and endorsement from many researchers spanning diverse domains, encompassing computer vision [7], fault diagnosis [8], speech recognition [9], recommendation systems [10, 11], and medical image analysis [12, 13]. Foundational deep learning architectures include convolutional neural networks (CNNs), autoencoders such as the variational autoencoder [8, 14] and the stacked denoising autoencoder [15], generative adversarial networks (GANs) [16], deep belief networks (DBN) [17], long short-term memory (LSTM) networks, and transformer networks [10]. In addition, the hybrid versions of these basic structures are applied, such as the combination of LSTM and DBN networks [18]. Notably, CNNs are widely recognized as the most popular type of deep learning network for capturing robust, discriminative, and local features from raw datasets thanks to the convolutional mechanism. In summary, deep learning, particularly in the form of CNNs, has become a critical and successful application in automated mammogram classification. Deep learning assumes a pivotal role as a CAD system, providing valuable support to expert radiologists across various domains.

The scarcity of extensive, publicly available datasets poses a significant challenge for medical image analysis studies, particularly for deep learning approaches such as mammography imaging. This challenge arises because deep learning models require substantial data to effectively learn the numerous parameters involved. Furthermore, deep learning models often encounter overfitting issues when applied to limited datasets, thus falling short of their intended generalization capabilities. Addressing the shortage of mammogram data by collecting and publicly releasing a large volume poses practical challenges in the short term. These challenges include concerns related to patient privacy, time constraints, costs, and the expertise required for such endeavors. Consequently, the development of strategic approaches within deep learning models is imperative in order to surmount the limitations imposed by inadequate data in mammography screening. Various strategies have been employed to mitigate overfitting challenges in deep networks when dealing with limited data. These strategies include data augmentation, synthetic data generation, multi-task learning, and the use of transfer learning models [19]. Despite the application of these strategies, enhancing the performance of deep learning models under limited data scenarios remains a formidable task [20].

In the past few years, self-supervised learning (SSL) strategies have attracted remarkable attention because of their ability to handle learning tasks utilizing pretext tasks in data scarcity [21–23]. The pretext task is pre-designed as the preliminary task for learning discriminative features on unlabeled data to be used in downstream tasks, such as classification, detection, and segmentation. Commonly used pretext tasks are categorized as generation based, context based, free-semantic-label based, or cross-modal based [24]. The purpose of the generation-based pretext task is to learn the features of images during the generation process, such as image inpainting [23], image colorization [22], image generation, and image super resolutions [25], with GANs [16] or autoencoder-based networks [26]. On the other hand, other approaches aim to learn visual features by solving pretext tasks that are similar to context for context-based, generating semantic labels for the free-semantic-label based, and verifying inputs with the bi-channel network cross-modal-based. Chee et al. [27] designed AIRNet affine registration for 3D MRI brain images by leveraging discriminative visual features. Chen et al. [28] applied context restoration as a pretext task on 2D ultrasound images for detection, CT images for localization, and brain MRI for segmentation. According to the proposed strategy for learning semantic features, two small patches are randomly selected, swapped, and then reconstructed. While repeating this swapping process several times, the intensity distribution is preserved, but spatial contexts change [28]. Zhou et al. [29] introduced Model Genesis on CT and X-ray images using various transformers with an autoencoder-based network. Thus, the proposed model by [29] provides more generalization ability with different transformations of images to improve the performance of segmentation and classification [30]. Gildenblat et al. [31] defined the Siamese network for image similarity as the first self-supervised model of histopathology WSIs. Thus, a large set of pairs extracted from histopathology WSIs are automatically annotated by similarity learning without any manual annotation owing to self-learning strategy [31]. Talep et al. [32] proposed a multimodal puzzle task as a self-supervised model for brain tumor segmentation and for predicting survival days on multimodalities. To et al. [33] proposed a self-supervised model for anomaly detection and localization on brain MRI, which uses a variational encoder to reconstruct images and a Siamese U-net for pseudo-labeling. In summary, as a common approach, SSL has emerged as a robust, repeatable, and accurate strategy in light of the lack of annotated data utilizing semantic visual features from the data itself via a pretext task.

We found a noticeable lack of self-supervised learning models for the mammography-based classification task, resulting from limited annotated data. To this end, we propose the use of a self-supervised model for mammogram classification for various tasks and datasets. The model we propose involves two main steps. First, a pretext task network is generated with a VAE to capture discriminative features of whole mammography images during reconstruction mammograms. Second, the downstream task of classifying mammograms is carried out with encoded space extracted by VAE. The proposed SSL model has been evaluated on public INbreast and MIAS mammography datasets. The proposed SSL model is designed to handle various tasks, including density, malignant-nonmalignant, and normal-benign-malignant

classification tasks. Furthermore, the results were compared with related works designed in the same classification task on INbreast and MIAS datasets that used end-to-end deep learning models and a self-supervised learning model [20], as explained in Section 2. The main contributions of this study are as follows:

- We proposed a new self-supervised learning classification model for a limited number of mammogram samples.
- The proposed SSL provides higher performance than previous studies, including a self-supervised model [20] for whole image-based classification.
- The proposed SSL exhibits greater generalization ability on small mammography datasets without high variance.
- The proposed SSL provides ease of computation complexity with lower dimensional representations by encoded space and lower computational complexity suitable for small datasets.
- In contrast to findings by [34] et al., which suggest that deep learning models perform well on a specific mammogram dataset but lack transferability to unknown external datasets, our research demonstrates the transferability of the encoded space from the INbreast dataset during training to an unknown external dataset (MIAS).

2 Related works on mammogram classification

Gong et al. [20] proposed a task-driven self-supervised bi-channel networks (TSBN) framework using self- supervised learning strategies to be trained on the INbreast dataset. TSBN has two main networks: (1) a pretext task designed with new contributions gray-scale image mapping (GSIM) for the reconstruction of mammograms using UNET [35] and the Residual Dense Network (RDN), and (2) a downstream network for malignant and nonmalignant classification constructed using ResNet50 as the backbone network. The combined features generated by the two networks are fed into the collaborative transfer module to improve classification performance based on the transfer learning mechanism between the bi-channel networks. As a result, the TSBN achieved higher performance (85.78% accuracy) than fine tuning-based SSL algorithms for the classification of mammograms on a limited INbreast dataset.

Shen et al. [36] proposed an "end-to-end" deep learning model for ROI-based and whole image-based breast cancer detection on CBIS-DDSM [37] and the INbreast [38] datasets. They used Resnet50 and VGG16 networks by combining the mediolateral oblique (MLO) and cranial-caudal (CC) views with a simple implementation that considered the average of model evaluations on separate views. As a result, the proposed four-model averaging improved the AUC from 0.88 to 0.91 on CBIS-DDSM and from 0.95 to 0.98 on INbreast, surpassing the best single model for ROI-based classification. Regarding whole image classification, the models trained on CBIS-DDSM and transferred to INbreast achieved AUC between 0.87 and 0.92 on independent training subsets of the INbreast for fine-tuning.

Wang et al. [34] focused on the inconsistency of deep models. They tested different deep models (e.g., AlexNet [39], VGG16 [40], and ResNet50 [41] to extract mammogram features, followed by classifier networks built using CNN. They also compared the performances of AlexNet, VGG16, and ResNet50 as end-to-end networks. As a result, they highlighted that different deep learning models trained on the DDSM dataset with high performance were not found to be transferable to testing on another public dataset, resulting in worse performance in each case.

Zhao et al. [42] introduced a novel bilateral adaptive spatial and channel attention network (BASCNet) that used Resnet as the backbone for breast density classification using single view (CC or MLO) and multi-view (CC and MLO) images. They also compared different Resnet series, which are Resnet18, Resnet 34, Resnet 50, Resnet 101, Resnet 152, regarding single views and multi-views on INbreast and DDSM. Resnet18-based BASCNet has achieved the best performance with multi-view mammograms on INbreast, with 90.51 accuracy. Zhao et al. also focused on transferring the models trained on different datasets, and the models on INbreast with pre-training on DDSM performed worse than those trained only on INbreast. On the other hand, they observed that the deeper networks lacked generalization ability. As a result, they emphasized that the generalization and transfer among deep models' datasets is an unsolved issue.

Li et al. [43] introduced a model based on dilated and attention-guided residual learning for multi-view density classification of mammograms. They assessed their proposed model on private clinical and INbreast datasets, demonstrating classification accuracies of 88.7% and 70.0%, respectively. Incorporating multi-view perspectives (CC and MLO), as well as sharing parameters across different streams, significantly enhanced the model's performance on clinical datasets. Despite the fact this proposed model outperformed both naive residual networks and recently developed deep learning methods on the private clinical dataset, it produced less accurate results on the public dataset, which the researchers attributed to its smaller size.

Houby et al. [44] presented CNN models for malignant and nonmalignant classification by using patches of the region of interest (ROI) and whole images on MIAS, INbreast, and DDSM mammography benchmark datasets. They reported satisfactory performance on both ROI-based and whole image-based tasks. The proposed pipeline performs preprocessing steps, augmentation, and resizing before classification. In whole image classification for MIAS,

their proposed model achieved 93.39 accuracy, 0.945 AUC, 92.72 sensitivity, 94.12 specificity, 93.58 f1; for INbreast, it achieved 93.04 accuracy, 94.6 AUC, 94.83 sensitivity, 91.23 specificity, and 93.22 f1.

Razali et al. [45] integrated artificial intelligence (AI) into breast screening processes, acknowledging challenges related to diverse patient demographics and non-standardized configurations of intelligence models. They proposed an enhanced classification model using a deep learning approach with a CNN) and a support vector machine (SVM). The method outperforms existing approaches in classifying breast density regions by utilizing pre-trained CNN models such as GoogleNet, ResNet50, ResNet101, and AlexNet, with SVM as a classifier. Notably, ResNet50 and GoogleNet combined with SVM exhibit significant improvements, with over 94% accuracy and an AUC exceeding 0.95 on MIAS. The model demonstrates strong feature extraction capabilities, suggesting potential applications in detecting malignancy from screening mammogram images.

Lou et al. [46] proposed a two-stage method that combines image preprocessing and model optimization. The first step is designed for preprocessing mammograms to improve the signal-to-noise ratio (SNR) and physiological characteristics. Their proposed model reduces manual labeling requirements by eliminating reliance on labor-intensive ROI annotations. The second step presents an ECA-Net50 model based on ResNet50 that incorporates an efficient channel attention (ECA) module for benign and malignant classification on an imbalanced mammogram dataset. ECA-Net50 was evaluated on the INbreast dataset and achieved an AUC value of 0.960, accuracy of 0.929, recall of 0.928, and precision of 0.883.

Jiang et al. [47] introduced a three-stage deep learning framework for breast cancer detection and classification by leveraging the probabilistic anchor assignment (PAA) algorithm. In the first stage, a PAA-based detector identifies suspicious breast lesions in mammograms. Following this, a two-branch ROI detector is introduced to classify and regress lesions, effectively minimizing false positives with a threshold-adaptive post-processing algorithm. The final stage involves an ROI classifier that combines local-ROI and global-image features to classify lesions as benign or malignant. Additionally, an image classifier is introduced to perform whole mammogram classification, utilizing the same global image features. Their proposed model integrates with three public mammogram datasets (CBIS-DDSM, INbreast, MIAS). It also enhances diagnostic efficiency by automatically detecting and classifying breast lesions for benign and malignant mammograms compared to recent state-of-the-art methods.

In summary, the general trend adopted by deep learning studies is to use end-to-end transfer learning networks trained by large amounts of natural data, such as ImageNet [48], to enhance generalization on limited mammogram datasets. Some studies employed a two- or three-stage approach to enhance the performance of classification, addressing challenges associated with insufficient dataset size. On the other hand, recently, SSL models have been applied to mammography-based classification tasks [20]. We performed a comparative study of mammogram classification on INbreast and MIAS datasets, evaluating the performance and computational complexity of our proposed SSL model against those of previous models.

3 Materials and methods

The proposed SSL model consists of two steps for mammography-based classification. The proposed SSL model is given in Fig. 1.

1. The first step is image reconstruction with a variational autoencoder (VAE), defined as a pretext task before the main task. Thus, deep probabilistic features have been extracted from high-dimensional data into low-dimensional space by VAE in an unsupervised manner.
2. The second step is classification, using the encoding variables extracted by VAE as a downstream task.

3.1 Pretext task network with a variational autoencoder

A variational autoencoder (VAE) is an unsupervised and generative deep learning model introduced by Kingma et al. in 2013 [49]. The VAE is formed by a probabilistic encoder network and decoder network based on Bayesian inference [49, 50]. First, the encoder network encodes image (x) to latent vector (z) by computing the approximated posterior distribution $q_\varphi\ (z|x)$. The latent space is calculated with mean (μ), standard deviations (σ) and random variable $\varepsilon\ \sim N\ (0, 1)$ as follows:

$$z_i = (\mu_i + \sigma_i \cdot \varepsilon) \sim q_\phi(z|x) \tag{1}$$

The decoder network decodes latent vector z to x' (reconstructed image) by maximizing the marginal log-likelihood $p_\theta(x|z)$. The basic VAE uses reconstruction loss (L_{rec}) and Kullback-Leibler divergence (L_{KL}) loss for minimizing reconstruction error between x and x', which is calculated as follow :

$$L_{rec} = -E_{q_\varphi(z|x)}[\log p_\theta(x|z)] \tag{2}$$

$$L_{KL} = D_{KL}(q_\varphi\ (z|x)\ \|_\theta(x|z) \tag{3}$$

$$L = L_{rec} + L_{KL} \tag{4}$$

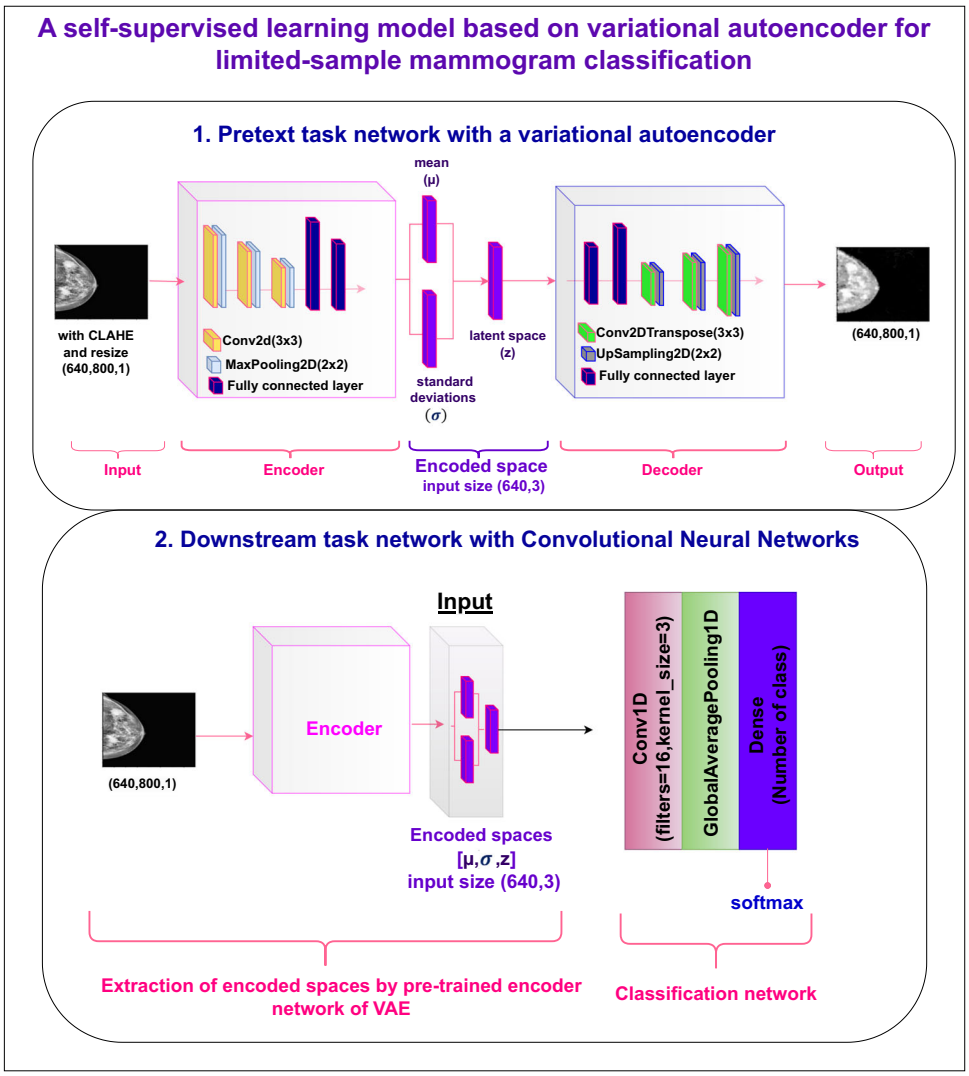

Fig. 1 The proposed self-supervised model for classifying mammogram images via two networks: The pretext task network and downstream network via a Variational Autoencoder (VAE) to extract features while reconstructing mammograms. VAE extracts profound probabilistic features from high-dimensional data, converting them into a reduced-dimensional space through unsupervised approaches. Then, the downstream network engages in classification using the encoding variables obtained from the VAE in a subsequent task facilitated by a CNN network

The pretext task network, incorporating a VAE, is illustrated in Fig. 2. Within the VAE network, the encoder tries to map an image of fixed dimensions (640, 800, 1) onto a latent space with dimensions (640, 1). Simultaneously, the decoder network strives to reconstruct the image to its original size. The encoder, receiving an input image of shape (640, 800, 1), systematically diminishes its spatial dimensions through three 2D convolutional layers employing a (3,3) kernel size, ReLU activation, and subsequent max-pooling layers. The resulting flattened representation undergoes dense layers, yielding mean (μ) and log variance (σ) vectors. A Lambda layer computes the final latent vector (z). The decoder gets the latent vector (z) as input and reconstructs the original image through a dense layer, transforming the latent vector into a format compatible with the decoder. A reshape layer then converts the 1D output to a 4D tensor. Consequently, the decoder network comprises three 2D transposed convolutional layers with a (3,3) kernel size and ReLU activation, followed by up-sampling layers that progressively decrease filters. The ultimate output shape signifies the reconstructed image as (640, 800, 1). Additionally, L1 and L2 regularization are applied to the dense layers in both the encoder and decoder, contributing to the model's generalization and mitigating overfitting concerns.

In summary, the VAE facilitates image representation and compresses data from a high-dimensional dataset into an allowable dimensional latent space using a learned non-linear map. In the encoding process, the encoder network of VAE endeavors to extract deep probabilistic features from high-dimensional data and transform them into a low-dimensional space. This study concentrates explicitly on the latent space, characterized by mean (μ) and standard deviation (σ) vectors, within the context of a classification task. Consequently, the VAE is configured as a pretext task preceding the main classification task and trained in an unsupervised manner with MIAS and INbreast datasets to extract self-features for

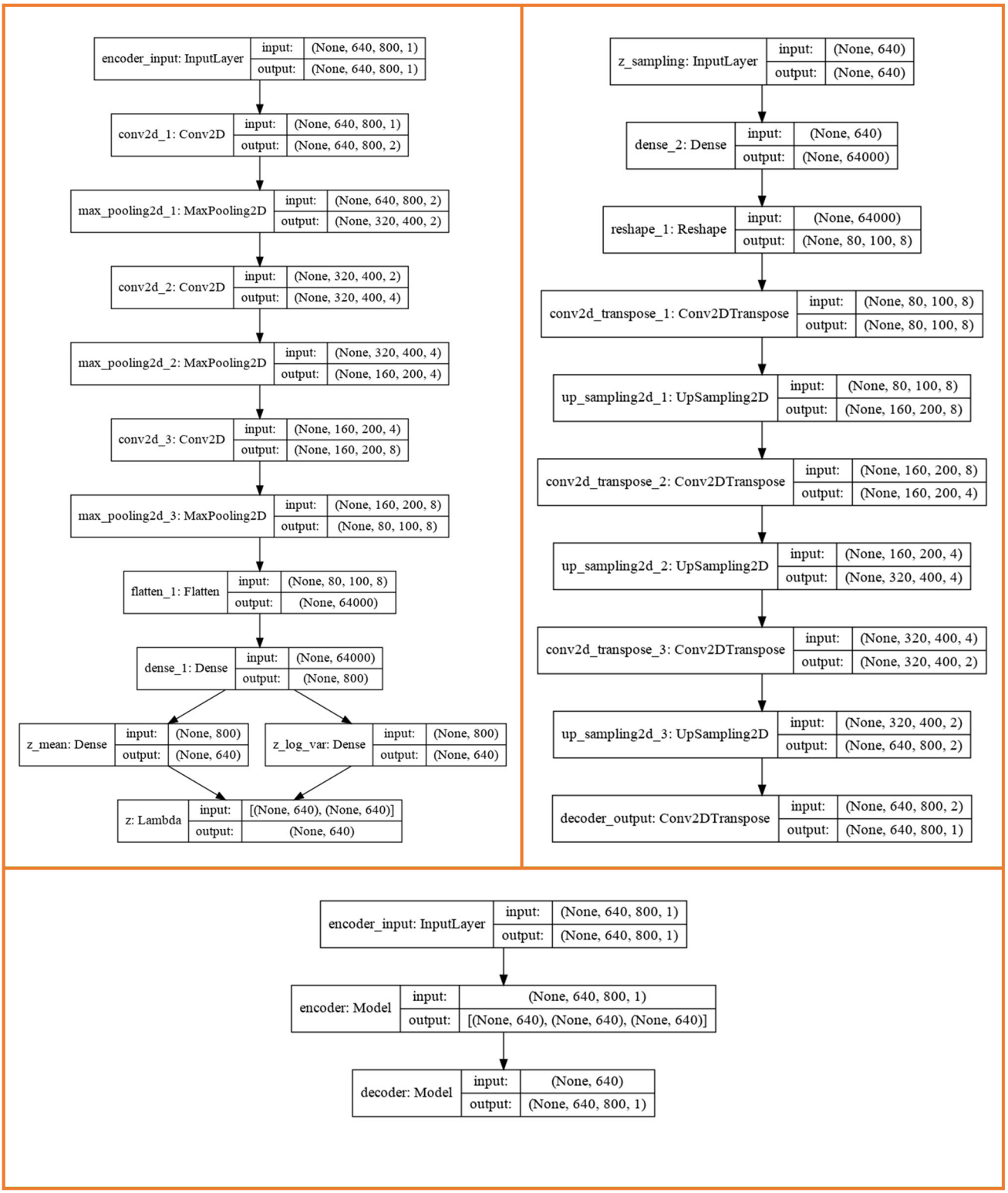

Fig. 2 The detailed encoder and decoder networks of VAE. The encoder network comprises three 2D convolutional layers with max-pooling after each convolution layer. The decoder architecture is the reverse of the encoder, followed by three 2D transposed convolutional layers and 2D upsampling layers

each dataset. Thus, two separate VAE models were developed for the INbreast and MIAS datasets to obtain an encoded space for subsequent use in the classification step.

3.2 Downstream task network with convolutional neural networks

Convolutional Neural Networks (CNN) [51] are generally built up with convolutional, pooling, and fully connected layers. While convolution layers extract features and pooling layers reduce the dimension of the extracted features, the last layer, which is fully connected, is used for classification. According to filtering and downsampling strategies, CNN-based classifier models are frequently used in computer vision tasks, including medical image analyses.

In this study, we used a simple CNN model for dealing with over-fitting problems on limited data for classification and took low dimensional encoded space extracted by the VAE as an input. Thus, the classification model has only one 1D convolutional layer employing a 3-kernel size, ReLU activation, followed by a global average pooling layer and the last fully connected layer for classification. The classifier model considers only the encoded space vector as an input vector in both training and testing to classify data that are in breast density range 0 to 3, benign-malignant, malignant-nonmalignant, and normal-benign-malignant. In summary, after the VAE extracts encode feature spaces in an unsupervised manner, the classifier model tries to model a low-dimensional encoded space vector (640,3) instead of high-dimensional mammography images (640,800,1). The computational complexity of the proposed models for each network is given in Table 1.

Table 1 The computational complexity of the proposed models is evaluated by including the number of depths, FLOPs (Floating-point Operations), and parameters

Network	Subnetwork	Depth	FLOPs(G)	Parameters
Pretext Task Network	Encoder of VAE	10	0.163	52.22M
	Decoder of VAE	9	0.147	41.03M
	VAE	19	0.311	93.25M
Downstream Network	CNN	3	0.000205 (204.95K)	228

4 Experimental study and results

4.1 Dataset

4.1.1 INbreast

This study used the INbreast dataset due to the high quality and limited digital mammogram images published by [38]. The INbreast dataset has 115 cases and 410 digital mammography images. We analyzed the proposed model's performance on two different classification tasks in the INbreast dataset: density classification and malignant-nonmalignant classification. In the density classification, the classification model tries to classify breast density ratios that are fatty (0-25 %), of scattered density (26-50 %), heterogeneously dense (51-75 %), and extremely dense (76-100 %). On the other hand, in the malignant and nonmalignant classification, BI-RADS 1 and 2 (but not BI-RADS 3) were targeted as nonmalignant, and BI-RADS 4, 5, and 6 were targeted as malignant. The number of images belonging to each class in INbreast is given in Table 2.

We used the data augmentation techniques of horizontal flipping and applying preprocessing before training with VAE on mammography images from the INbreast dataset. First, 410 mammography images were increased to 1230 using contrast-limited adaptive histogram equalization (CLAHE). Then, the originals and the CLAHE images were horizontally flipped; consequently, the total number of images increased to 2460. The augmented mammography dataset was combined into one class and split into training and validation sets in ranges of 75 and 25 for the VAE model, respectively. There was no augmentation in classification except with regard to the malignant training set. The number of images in the malignant train set was increased twice using the original and CLAHE versions of images. The training set and test set were split into 80% and 20%, respectively, for each class in the classification process, and 20 percent of the training set was used as a validation in the classifier model for INbreast.

4.1.2 MIAS

The Mammographic Image Analysis Society (MIAS) is a public digital mammography dataset published by [52]. MIAS has 322 digital mammograms, 208 classified as normal, 63 as benign, and 51 as malign. We applied CLAHE

Table 2 The number of images corresponding to breast density, malignant, and nonmalignant class labels in the INbreast dataset

Classes	Number of mammography images
respectively fatty (A)	136
scattered density (B)	147
heterogeneously dense (C)	99
extremely dense (D)	28
nonmalignant	287
malignant	100

to all 322 mammogram images and horizontally flipped the same setting as in INbreast to train VAE. The augmented images were split into training and validation sets, respectively 75% and 25% in the VAE model. In the classification model of MIAS for density, benign-malignant and normal-benign-malignant classification tasks were considered, as those were the annotated labels of the dataset. The number of mammography images belonging to each class in MIAS is given in Table 3. The training and test sets were split into 80% and 20% for the classification model, and 20% of the train data set was used to validate the classifier model for MIAS.

4.2 INbreast dataset transferred to MIAS dataset

Wang et al. [34] reported deep learning models that cannot easily be transferred from one mammogram dataset to other mammogram datasets. Thus, we focus on the transfer model from the INbreast to MIAS datasets to observe its effect on performance. The INbreast dataset is used for training and testing MIAS in different model settings for benign-malignant and normal-benign-malignant classification. While we use all nonmalignant and malignant classes of INbreast for benign-malignant classification, the A, B, and C+D (merged in one class) INbreast density classes are used for normal-benign-malignant classification. When INbreast is set as the training set, MIAS is a test set that includes 51 malignant and 63 benign samples for benign-malignant classification and 51 normal-benign-malignant samples for normal-benign-malignant classification.

4.3 Experimental setup

The VAE and classifier models were built using the Keras library. On the other hand, the models were trained and tested on a GeForce RTX 2080 Ti GPU with the Tensorflow-gpu library. The VAE was set up with 16 batch size, L1= 1e-5 and L2= 1e-4 regularization over mean(μ), var(σ), and a latent layer of VAE, RMSprop optimizer with 0.001 learning rate, and 100 epochs to train in an unsupervised manner on each of the mammography datasets that are INbreast and MIAS. The classifier model was defined with 16 batch size, Adam optimizer with 0.001 learning rate, and 10000 epochs to train with encoded spaces extracted from each dataset of classifying density, benign-malignant, malignant-nonmalignant, and normal-benign-malignant tasks. In addition, 5-fold cross-validation was used for all classification experiments. The classifier model was evaluated by recall, precision, F1, and area under the curve (AUC) evaluation metrics for each dataset calculated with a weighted average.

Table 3 The number of images corresponding to breast density, normal, benign, and malignant class labels in MIAS

Classes	Number of mammography images
fatty	106
glandular	104
dense	112
normal	208
benign	63
malignant	51

4.4 VAE results

The VAE reconstruction results of the INbreast and MIAS datasets are presented in Fig. 3, sorted as original mammograms, implemented CLAHE-resize (640, 800, 1), and reconstructed images. The reconstructed mammograms by the VAE can be described as a rough reconstruction because of some losses, as seen in Fig. 3. While the VAE captures the general and significant framework (e.g., dense lesions in the breast), details are only hinted at (e.g., the capillaries in the breast). The purpose of the pretext network model in this study is not to fully learn the intrinsic representations of mammograms but rather general, representative features. Therefore, the primary objective is to observe whether mapping the images in a lower-dimensional latent space would provide sufficient information to conduct the classification tasks defined. On the other hand, we operate on a small dataset, bringing out a significant overfitting problem. Consequently, missing insignificant details will contribute to the model's generalizing ability and reduce variance on insufficient samples to avoid overfitting. Thus, the classification model tries to correctly classify the extracted encoded features by the VAE. The classification results are presented in the following sections.

4.5 Classification results for INbreast

The mean and standard deviation of the 5-fold cross-validation results on density and malignant-non-malignant classification experiments for the INbreast dataset are presented in Table 4. The results of MIAS are given in Table 5. In addition, ROC and precision-recall curves are given in Fig. 4. The results from various views (image-only, MLO, and CC) have also been obtained and reported. The proposed model with image view for density classification is better than MLO and CC view models when it comes to accuracy, AUC, precision, recall, and F1 scores, with 88.75, 0.94, 82.86, 88.75, and 85.52. MLO and CC models produced results close to these with the image-only model by a difference of 0.95 for MLO and 2.12 for CC. On the other hand, the model's precision results are lower than its other metrics; the proposed model generally assigns an extremely dense class

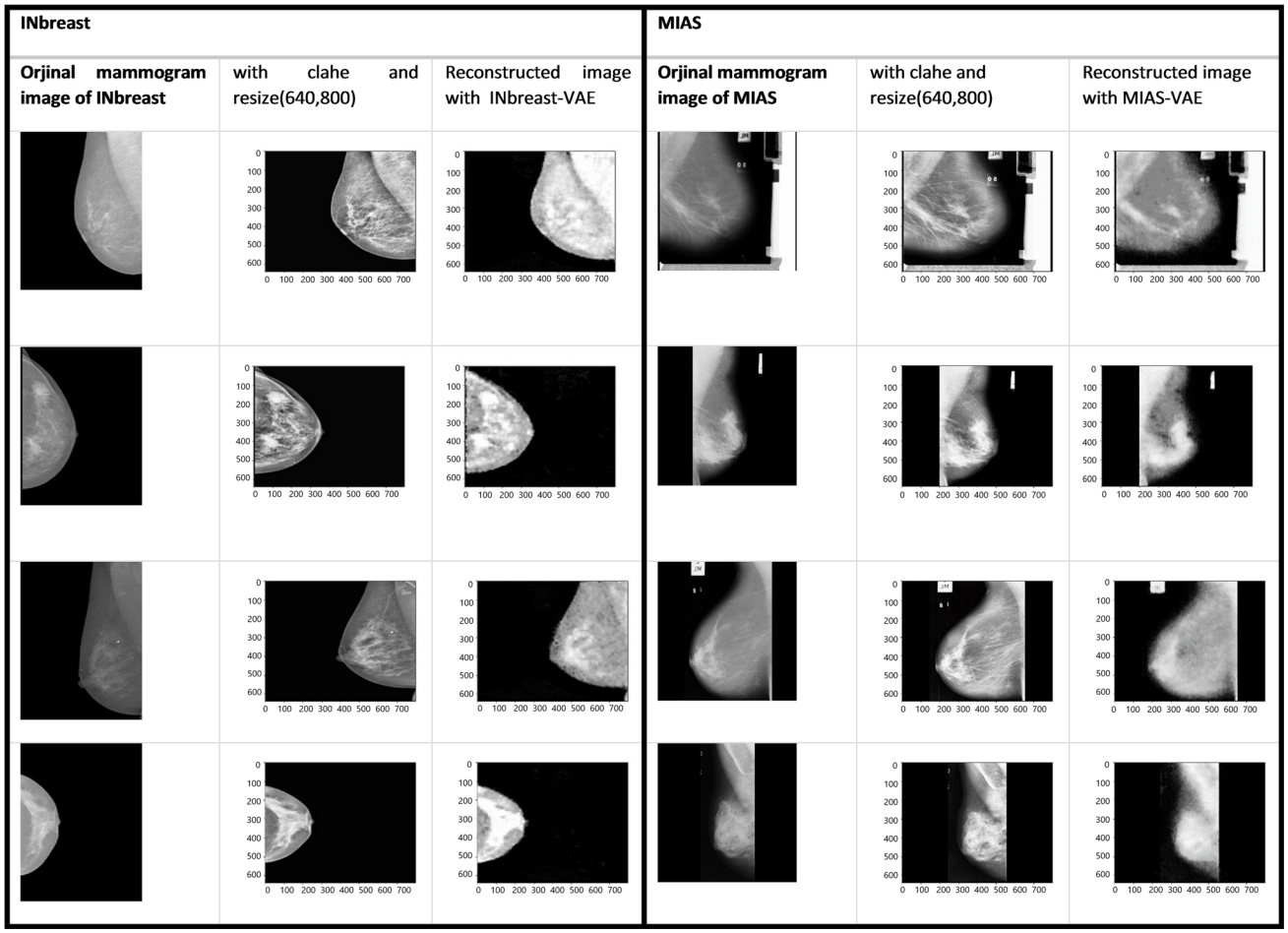

Fig. 3 Reconstruction results of mammograms from both the INbreast and MIAS datasets, involving original, original mammograms processed with CLAHE and resizing, reconstructed mammograms by VAE, respectively

(D) as a heterogeneously dense class (C), as illustrated in Fig. 4. While the AUC of the other classes is over 0.96, the AUC of the highly dense class is 0.59. The low sample of D (28), as well as the imbalance between classes, might have made it impossible to distinguish D from C of the proposed model.

Table 4 The test accuracy, AUC, precision, recall, and F1 with mean and standard deviation(std) results are assessed through 5 cross-validation on the INbreast dataset using the INbreast-VAE model

Objectives	VİEW		accuracy	auc	precision	recall	F1
Density	image	mean	88.75	0.94	82.86	88.75	85.52
classification		std	0.012	0.011	0.014	0.013	0.012
	MLO	mean	87.80	0.92	83.28	87.80	84.91
		std	0.000	0.007	0.000	0.000	0.000
	CC	mean	86.63	0.91	80.41	86.67	83.33
		std	0.011	0.006	0.013	0.011	0.012
Malignant	image	mean	91.79	0.95	93.46	91.79	92.09
nonmalignant		std	0.011	0.003	0.009	0.011	0.011
classification	MLO	mean	90.86	0.99	92.94	90.86	91.11
		std	0.013	0.004	0.007	0.013	0.012
	CC	mean	94.86	0.98	94.90	94.86	94.86
		std	0.013	0.007	0.014	0.013	0.013

The metrics are evaluated for density and malignant-nonmalignant classification, considering image, MLO, and CC views

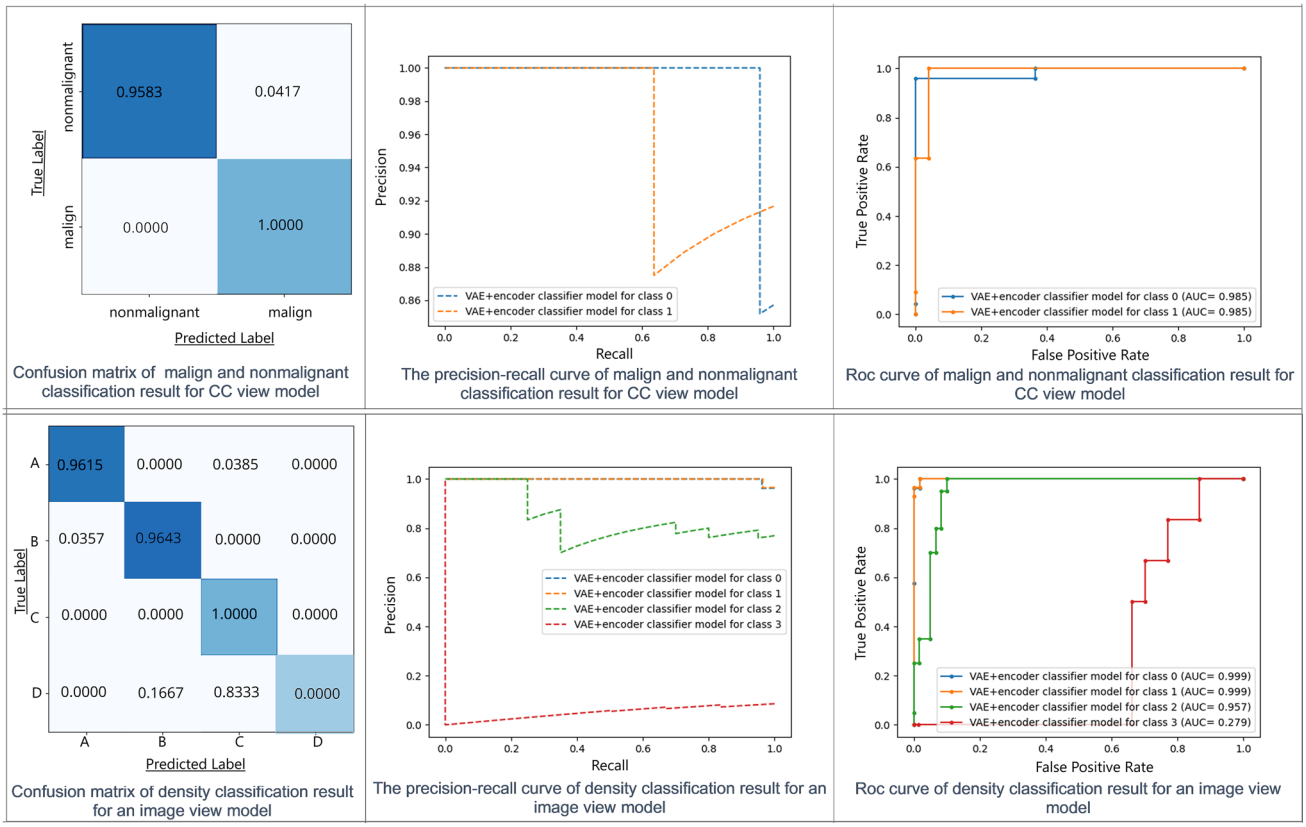

Fig. 4 Confusion Matrix, Precision-Recall Curve, and ROC Curve depicting the best evaluation results on INbreast for density and malignant-nonmalignant classification using the INbreast-VAE model

On the other hand, the best results of malignant and nonmalignant classification were 94.86 accuracy, 94.90 precision, 94.86 recall, and 94.86 F1 with CC view and AUC 0.99 with MLO view. According to the image view, the classification of malignant and nonmalignant results is still high, with 91.79 accuracy, 0.95 AUC, 93.46 precision, 91.79 recall, and 92.09 F1. The MLO and image view results are similar, except an AUC of MLO is 0.99, which is superior to the others. Regarding different tasks, the malignant-nonmalignant classifications are more stable than density classifications for each class, as seen by the ROC and precision-recall curves given in Fig. 4. As the number of classes increases, the classification task becomes more challenging for the proposed SSL model.

The results of previous studies on INbreast classification are shown in Table 6 to facilitate a comparative analysis. The proposed model for density classification on the MLO view performed better than earlier works regarding accuracy, AUC, and F1 results. Although BASCNet (ResNet18) [42] on the MLO view resulted in better performance, with 0.94 AUC, BASCNet (ResNet18) has a low F1 score (68.09) and higher variances than the proposed model. Therefore, the proposed model is more robust against different test sets than BASCNet (ResNet18). Based on the outcomes of malignant-nonmalignant classification, the proposed model for image views, with an AUC of 0.95, surpasses earlier studies [20, 34, 36]. While the proposed model exhibits comparable performance on the image view of mammograms, as seen in prior studies [44–46], it achieves superior performance on the CC view, with an AUC of 0.98 compared to all the previous studies [20, 34, 36, 44–46]. On the other hand, Table 6 presents the self-supervised Bi-channel networks (TSBN)[20] for mammogram classification with Resnet50, RDN, and U-Net based on various pretext tasks; this study is in the same category as ours. The proposed SSL model significantly outperformed the TSBN introduced by [20]. Moreover, Welch's t-test was employed to conduct statistical significance tests comparing AUC scores across studies. The studies of [20] and [42] assessed the F1 score as an exception since AUC scores were unavailable in [20] and there was a substantial difference between the F1 score and AUC in [42]. The p-value represents the statistical significance of the difference between the models, the test statistic (T) measures the difference, and the effect size (d) indicates the magnitude of the difference. While lower p-values ($p \leq 0.05$) suggest more significant differences, the large test statistics and effect sizes in some cases indicate substantial distinctions between the compared models. In Table 6, the proposed model is highlighted in bold,

Table 5 The test accuracy, AUC, precision, recall, and F1 with mean and standard deviation(std) results are assessed through 5 cross-validation on MIAS

Objectives	Train data		accuracy	auc	precision	recall	F1
Density	MIAS	mean	94.60	0.99	94.90	94.60	94.60
Classification		std	0.009	0.005	0.008	0.009	0.009
Benign	MIAS	mean	89.09	0.97	89.93	89.09	88.95
Malignant		std	0.025	0.007	0.022	0.025	0.026
	INbreast	mean	87.23	0.95	88.80	87.23	86.93
		std	0.000	0.002	0.000	0.000	0.000
Normal	MIAS	mean	81.90	0.83	73.02	81.90	76.45
Benign		std	0.009	0.008	0.008	0.009	0.008
Malignant	INbreast	mean	86.54	0.99	88.61	86.54	85.97
		std	0.012	0.002	0.009	0.012	0.013

The assessment is conducted using the MIAS-VAE model for density, benign-malignant, and normal-benign-malignant classification tasks. Additionally, the proposed model, trained on the INbreast dataset, is evaluated on an external dataset (MIAS)

indicating significant improvements over the compared models. The proposed model exhibits significant improvement over Resnet18 [42] with the lowest p-value of 0.000319, the largest test statistic (T) of 9.6774, and effect size (d) of 6.12 in the context of density classification. Similarly, for malignant-nonmalignant classification, the proposed model significantly outperforms AlexNet + CNN [34] with a p-value of 1.681e-27, a test statistic (T) of 4901.5304, and an effect size (d) of 3100. In summary, the proposed model consistently demonstrated significant improvements over the compared models in density and malignant-nonmalignant classification tasks on INbreast, as indicated by low p-values ($p \leq 0.05$), large negative test statistics, and substantial effect sizes. These findings suggest the effectiveness and superiority of the proposed model across different datasets and tasks. Furthermore, our model was assessed in terms of computational complexity compared to previous studies, as shown in Table 8. The proposed model enables a lightweight network characterized by a small number of trainable parameters (228 parameters), low computational cost (204.95K), and a depth of 3, positioning it as a resource-efficient and highly accurate model. Consequently, our proposed model emerges as a robust, reliable, and efficient solution in the context of mammogram classification.

4.6 Classification results for MIAS

The results of 5-fold cross-validation on density, malignant-nonmalignant, and normal-benign-nonmalignant classification experiments for the MIAS dataset are reported in Table 5. In addition, the confusion matrix, ROC curve, and precision-recall curve are given in Fig. 5. The proposed model reached accuracy, precision, recall, and F1 scores with approximately 95 and 0.99 AUC for density classification and approximately 89 and 0.97 AUC for benign-malignant. In normal-benign-malignant classification, the classifier using INbreast data for training and MIAS for testing has better evaluation metrics that are 86.54 accuracy, 0.99 AUC, 88.61 precision, 86.54 recall, and 85.97 F1.

We presented a comprehensive comparison between the proposed SSL model and previous studies [34, 44, 45] in the context of MIAS classification, as outlined in Table 6. While the CNN model proposed by [44] demonstrates high accuracy, recall, and F1 scores, the proposed model reaches a high AUC score of 0.97 for benign and malignant classification. In addition, the proposed model has seen an outstanding performance with 0.97 AUC against the study by [34] for the identification of benign and malignant. Moreover, the study by [34] focused on transfer learning among mammogram datasets. The AlexNet + CNN model [34] is trained on mixed mammogram datasets (INbreast, MIAS, private set), then tested on unseen external mammogram datasets, given in Table 7. They [34] reported inconsistency in deep learning models that have high performance on one mammography dataset and cannot be transferred or generalized to unseen external data sets. In our proposed model, we used the INbreast dataset for training and MIAS for testing, which demonstrates that the encoded space of the INbreast dataset is transferable and performs effectively on the unseen external MIAS dataset. This observation contradicts the findings of [34], as indicated by a p-value of 4.181e-26, a test statistic (T) of 7144.2472, and an effect size (d) of 4518.4 obtained from Welch's t-test, highlighting the transferability issue of deep learning models between different mammogram datasets. The encoded space extracted by VAE proved to be a valuable asset, facilitating model transfer between INbreast and MIAS datasets and enhancing normal-benign-malignant classification results, as reported in Table 7. Regarding density classification, the proposed model exhibits performance similar to that of the previous study conducted by Razali et al. [45], with a p-value of 0.5. This p-value of 0.5 typically indicates that there is no statistically significant difference

Table 6 Comparative results for each classification task model on INbreast and MIAS datasets between state-of-the-art previous methods and the proposed model

Objectives	references	model	dataset	view	accuracy	auc	recall	F1
density	Zhao et al. [42]	ResNet18	INbreast	mlo	70.42±7.98	0.82±1.54	–	48.08±8.51
classification	Zhao et al. [42]	BASCNet	INbreast	mlo	86.27±6.75	0.94±6.49	–	68.09±14.56
	Li et al. [43]	ResNet50+DC+CA	INbreast	multi-view	70.0	0.85	–	63.5
		the proposed model	**INbreast**	**mlo**	**87.80±0.000**	**0.92±0.007**	**87.80±0.000**	**84.91±0.000**
		the proposed model	**INbreast**	**image**	**88.75±0.012**	**0.94±0.011**	**88.75±0.013**	**85.52±0.012**
density	Razali et al. [45]	GoogleNet+SVM	MIAS	image	95.39	0.99	–	–
classification		the proposed model	MIAS	image	94.60±0.009	0.99±0.005	94.60±0.009	94.60±0.009
malignant-	Gong et al. [20]	TSBN	INbreast	image	85.53±1.87	–	84.00±2.24	75.06±2.52
nonmalignant	Shen L. [36]	Resnet-Resnet Resnet-VGG VGG-VGG VGG-Resnet]	INbreast	image classification	–	0.87-0.92	–	–
	Wang et al. [34]	AlexNet+CNN	INbreast	image	–	0.67±0.01	–	–
	El Houby et al. [44]	CNN	INbreast	image	93.04	0.95	94.83	93.22
	Razali et al. [45]	Three-stage PAA	INbreast	image	–	0.96	–	–
	Lou et al. [46]	ECA-Net50	INbreast	image	92.9	0.96	92.8	–
		the proposed model	**INbreast**	**image**	**91.79±0.011**	**0.95±0.003**	**91.79±0.011**	**92.09±0.011**
		the proposed model	**INbreast**	**CC**	**94.86±0.007**	**0.98±0.007**	**94.86±0.013**	**94.86±0.013**
benign-	Wang et al. [34]	AlexNet+CNN	MIAS	image	–	0.58±0.01	–	–
malignant	El Houby et al. [44]	CNN	MIAS	image	93.39	0.95	92.72	93.58
classification		**the proposed model**	**MIAS**	**image**	**89.09±0.025**	**0.97±007**	**89.09±0.025**	**88.95±0.026**

The proposed model is highlighted in bold, signifying significant improvements where $p - value(p) \leq 0.05$ compared to the other models

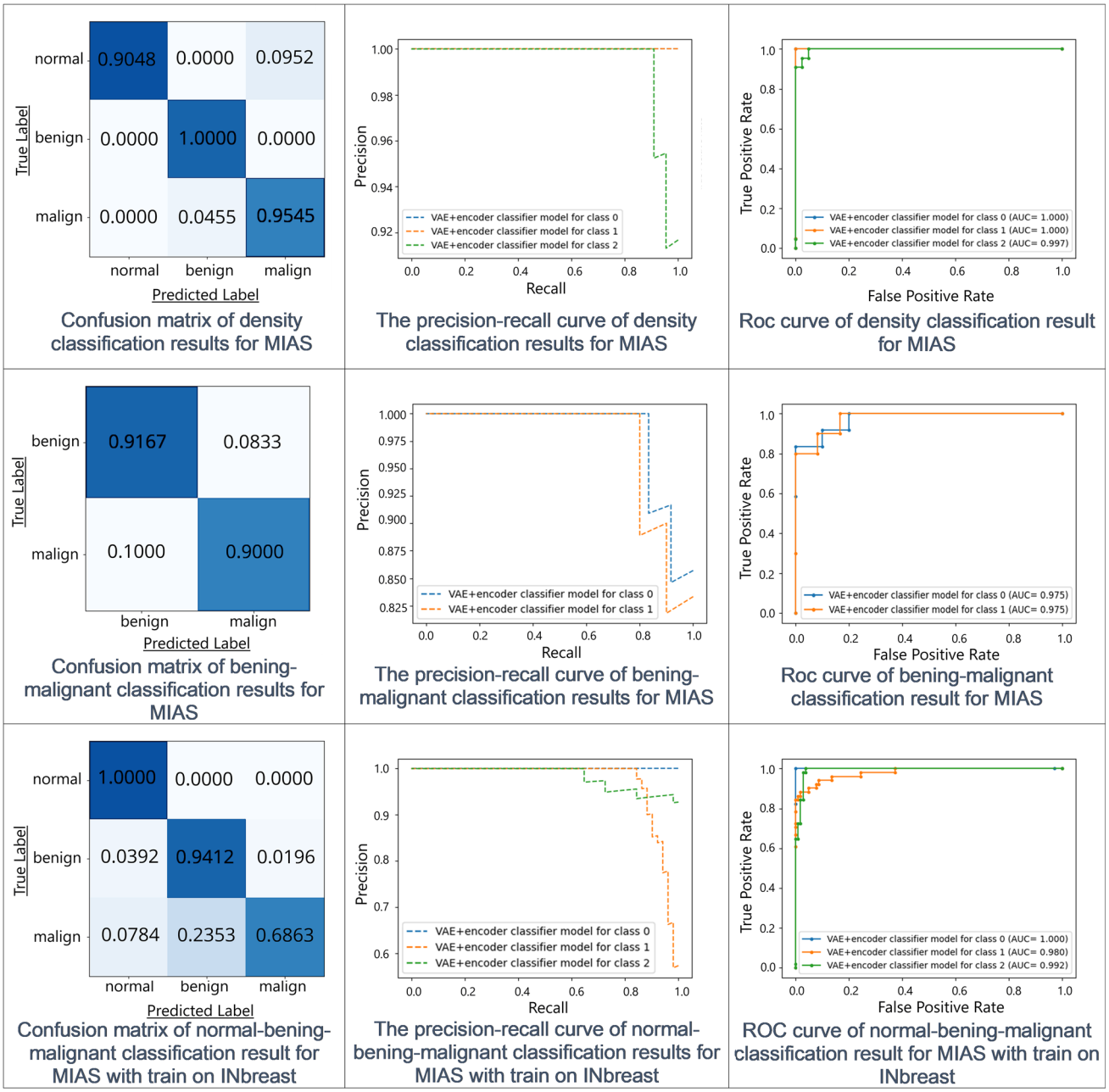

Fig. 5 Confusion Matrix, Precision-Recall Curve, and ROC Curve depicting the best evaluation results on MIAS for density, benign-malignant, and normal-benign-malignant classification tasks using the MIAS-VAE model

between the performance of the proposed model and that of the model from the previous study. Moreover, the proposed model stands out by providing substantially lower computational complexity (Table 8) while still maintaining high performance on the MIAS dataset for both benign-malignant and density classifications. This highlights the efficiency and effectiveness of the proposed model in the context of mammogram classification on the MIAS dataset.

5 Discussion

The proposed SSL model addresses the challenge of limited data availability in mammogram classification. Our proposed model offers a promising solution to the challenges posed by limited mammogram samples. Furthermore, the proposed SSL classification model enables effective and efficient models with lower computational costs. Overall, the performance,

Table 7 Comparative results for MIAS between the proposed SSL model and the study by [34] are presented to illustrate the transferability of the encoded space from the INbreast dataset during training to an unknown external dataset (MIAS)

Objectives	references	model	train data	test data	auc
Benign-malignant classification	Wang et al. [34]	AlexNet+CNN	mixed data	MIAS	0.58±0.01
		proposed	INbreast	MIAS	0.95±0.002
		proposed	MIAS	MIAS	0.97±0.007

The study by [34] utilized mixed mammogram datasets (INbreast, MIAS, private set) for training and MIAS for testing

generalization ability, ease of computation, and transferability of encoded space make it a valuable contribution to the field of mammogram classification, with potential implications for enhancing the accuracy and efficiency of breast cancer diagnosis.

On the other hand, the results demonstrate a notable challenge with the proposed model in distinguishing between an extremely dense class (D) and a heterogeneously dense class (C). The AUC of the highly dense class is notably lower at 0.59 compared to the AUCs of other classes, which are reported to be over 0.96. This discrepancy in AUC values indicates a difficulty in accurately classifying density classification belonging to the highly dense class on INbreast. Furthermore, the visualization of the reconstructed mammograms by the VAE provides insights into the limitations of the model. The reconstructed images appear to undergo rough reconstruction, indicating that the VAE faces challenges in capturing fine details. While the VAE effectively captures general and significant features such as dense lesions in the breast, it falls short in representing intricate details like capillaries. Therefore, the proposed model utilizes only latent space features for classification.

In future studies, our objective is to improve the generation results for mammograms by exploring advanced generative models, such as GAN and diffusion-based networks. These superior models will be employed to enhance the quality of the generated images. Subsequently, these generated images will be used as data augmentation in the classification step. Furthermore, the current model relies on a single mammography image for classification. In future work, we aspire to develop a multi-view model for mammography classification to ensure consistency in identifying mammograms through a patient-based model, rather than relying solely on individual images.

6 Conclusion

This study proposed an SSL model in two stages to overcome the drawback of limited mammogram samples on deep learning models. The proposed SSL model has achieved better performance than robust architectures such as Resnet, VGG, and AlexNet, as reported by previous studies on the public mammogram dataset. Furthermore, we also focused on the performance of the study that used self-supervised strategies for mammogram classification. The proposed self-supervised strategy achieved more accurate results than the previously reported bi-channel self-supervised network on the INbreast dataset. Thus, the proposed self-supervised model has shown promising effectiveness with respect to repeatability, consistency, generalization ability, and transfer among various datasets. As a result, the proposed self-supervised model provides higher performance on limited data without high variance and eases computation of both lower-dimensional representations with encoded space and fewer computational costs, with only 228 trainable parameters, 204.95K FLOPs, and a depth of 3 for mammogram classification. It should be noted that the number of trainable

Table 8 The computational complexity of the proposed models is compared to that of previous studies, considering metrics including the number of layers (depth), FLOPs (Floating-point Operations), and the number of parameters

References	Model	Depth	FLOPs(G)	Parameters
Wang et al. [34]	AlexNet	8	7.27	60.97M
Lee et al. [36]	VGG16	16	154.7	138.36M
Zhao et al. [42]	BASCNet	22	7.32	24.23M
El Houby et al. [44]	CNN	8	–	1.21M
Razali et al. [45]	GoogleNet	22	16.04	6.9M
Lou et al. [46]	ECA-Net50	50	–	23.51M
Gong et al. [20]	TSBN-downstream network (ResNet50)	50	3.8	25.56M
The proposed	The downstream network with CNN	3 (204.95K)	0.000205	228

parameters for various classification tasks in this study was relatively small (228 parameters). Nevertheless, the resulting classifiers were shown to be competitive. This might imply that general image features are fundamental to related problems, and representation learning might find suitable applications in medical imaging diagnostics.

Medical image labeling processing requires time, expertise, and cost; therefore, SSL presents a solution for using many unlabeled data in medicine. Future studies of SSL on medical image tasks could have potential because the practical scenarios often lack sufficient unlabeled data yet present a small number of annotated samples. Furthermore, SSL allows for the use of unlabeled data in unsupervised pretext tasks. In conclusion, SSL has emerged as a remarkable field for overcoming the problem of many limited and unlabeled datasets.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data Availability The data and code of the study are available from the corresponding author upon reasonable request.

Declarations

Conflicts of interest The authors declare no potential conflicts of interest regarding any financial support, research, authorship, and publication of this article.

References

1. Who (2021) Cancer. https://www.who.int/en/news-room/fact-sheets/detail/cancer/
2. Karellas A, Vedantham S (2008) Breast cancer imaging: a perspective for the next decade. Medical Phys 35(11):4878–4897
3. Ueda D, Yamamoto A, Onoda N et al (2022) Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets. PLoS One 17(3):e0265,751
4. Yassin NI, Omran S, El Houby EM et al (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Comput Methods Programs Biomed 156:25–45
5. Ribli D, Horváth A, Unger Z et al (2018) Detecting and classifying lesions in mammograms with deep learning. Sci Report 8(1):1–7
6. Wu N, Phang J, Park J et al (2019) Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 39(4):1184–1194
7. Hu WC, Chen LB, Huang BK et al (2022) A computer vision-based intelligent fish feeding system using deep learning techniques for aquaculture. IEEE Sensors J 22(7):7185–7194
8. Yan X, She D, Xu Y et al (2021) Deep regularized variational autoencoder for intelligent fault diagnosis of rotor-bearing system within entire life-cycle process. Knowl-Based Syst 226(107):142
9. Weng Z, Qin Z, Tao X et al (2023) Deep learning enabled semantic communications with speech recognition and synthesis. IEEE Trans Wireless Commun
10. Rostami M, Oussalah M, Farrahi V (2022) A novel time-aware food recommender-system based on deep learning and graph clustering. IEEE Access 10:52,508–52,524
11. Rostami M, Muhammad U, Forouzandeh S et al (2022) An effective explainable food recommendation using deep image clustering and community detection. Intell Syst Appl 16(200):157
12. Chen X, Yao L, Zhou T et al (2021) Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images. Pattern Recognit 113(107):826
13. Liu P, Du J, Vong CM (2023) A novel sequential structure for lightweight multi-scale feature learning under limited available images. Neural Netw 164:124–134
14. Yan X, She D, Xu Y (2023) Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions. Expert Syst Appl 216(119):479
15. Yan X, Liu Y, Jia M (2020) Health condition identification for rolling bearing using a multi-domain indicator-based optimized stacked denoising autoencoder. Structural Health Monitoring 19(5):1602–1626
16. Goodfellow I, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets. Adv Neural Inform Process Syst 27
17. Yan X, Liu Y, Jia M (2020) Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions. Knowl-Based Syst 193(105):484
18. Yan X, Liu Y, Xu Y et al (2020) Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition. Energy Conversion Manag 225(113):456
19. Altaf F, Islam SM, Akhtar N et al (2019) Going deep in medical image analysis: concepts, methods, challenges, and future directions. IEEE Access 7:99,540–99,572
20. Gong R, Lu Z, Shi J (2021) Task-driven self-supervised bi-channel networks learning for diagnosis of breast cancers with mammography. arXiv:2101.06228
21. Noroozi M, Favaro P (2016) Unsupervised learning of visual representations by solving jigsaw puzzles. In: European conference on computer vision, Springer, pp 69–84
22. Zhang R, Isola P, Efros AA (2016) Colorful image colorization. In: European conference on computer vision, Springer, pp 649–666
23. Pathak D, Krahenbuhl P, Donahue J et al (2016) Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2536–2544
24. Jing L, Tian Y (2020) Self-supervised visual feature learning with deep neural networks: A survey. IEEE Trans Pattern Anal Machine Intell
25. Ledig C, Theis L, Huszár F et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690
26. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507

27. Chee E, Wu Z (2018) Airnet: Self-supervised affine registration for 3d medical images using neural networks. arXiv:1810.02583
28. Chen L, Bentley P, Mori K et al (2019) Self-supervised learning for medical image analysis using image context restoration. Med Image Anal 58(101):539
29. Zhou Z, Sodha V, Siddiquee MMR et al (2019) Models genesis: Generic autodidactic models for 3d medical image analysis. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 384–393
30. Truong T, Mohammadi S, Lenga M (2021) How transferable are self-supervised features in medical image classification tasks? In: Machine learning for health, PMLR, pp 54–74
31. Gildenblat J, Klaiman E (2019) Self-supervised similarity learning for digital pathology. arXiv:1905.08139
32. Taleb A, Lippert C, Klein T et al (2021) Multimodal self-supervised learning for medical image analysis. In: International conference on information processing in medical imaging, Springer, pp 661–673
33. To MS, Sarno IG, Chong C et al (2021) Self-supervised lesion change detection and localisation in longitudinal multiple sclerosis brain imaging. arXiv:2106.00919
34. Wang X, Liang G, Zhang Y et al (2020) Inconsistent performance of deep learning models on mammogram classification. J American College Radiol 17(6):796–803
35. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer, pp 234–241
36. Shen L (2017) End-to-end training for whole image breast cancer diagnosis using an all convolutional design. arXiv:1711.05775
37. Lee RS, Gimenez F, Hoogi A et al (2017) A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific Data 4(1):1–9
38. Moreira I, Amaral I, Domingues I (????) a., & cardoso, js (2012). INbreast: Toward a Full-field Digital Mammographic Database Academic Radiology 19(2):236–248
39. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inform Process Syst 25:1097–1105
40. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
41. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
42. Zhao W, Wang R, Qi Y et al (2021) Bascnet: Bilateral adaptive spatial and channel attention network for breast density classification in the mammogram. Biomed Signal Process Control 70(103):073
43. Li C, Xu J, Liu Q et al (2020) Multi-view mammographic density classification by dilated and attention-guided residual learning. IEEE/ACM Trans Computat Biol Bioinform 18(3):1003–1013
44. El Houby EM, Yassin NI (2021) Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. Biomed Signal Process Control 70(102):954
45. Razali NF, Isa IS, Sulaiman SN et al (2022) Improvement of breast density classifier based on cnn features extraction and svm in mammogram images. Training 7:18
46. Lou Q, Li Y, Qian Y et al (2022) Mammogram classification based on a novel convolutional neural network with efficient channel attention. Comput Biol Med 50(106):082
47. Jiang J, Peng J, Hu C et al (2022) Breast cancer detection and classification in mammogram using a three-stage deep learning framework based on paa algorithm. Artif Intell Med 134(102):419
48. Deng J, Dong W, Socher R et al (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Ieee, pp 248–255
49. Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv:1312.6114
50. Zemouri R (2020) Semi-supervised adversarial variational autoencoder. Mach Learn Knowl Extraction 2(3):361–378
51. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
52. Suckling J, Parker J, Dance D et al (2015) Mammographic image analysis society (mias) database v1. 21. https://www.repository.cam.ac.uk/handle/1810/250394