



BraNet: a mobil application for breast image classification based on deep learning algorithms

Yuliana Jiménez-Gaona^{1,2,3} · María José Rodríguez Álvarez² · Darwin Castillo-Malla^{1,2,3} · Santiago García-Jaen¹ · Diana Carrión-Figueroa⁴ · Patricio Corral-Domínguez⁵ · Vasudevan Lakshminarayanan⁶

Received: 14 November 2023 / Accepted: 26 March 2024
© The Author(s) 2024

Abstract

Mobile health apps are widely used for breast cancer detection using artificial intelligence algorithms, providing radiologists with second opinions and reducing false diagnoses. This study aims to develop an open-source mobile app named “BraNet” for 2D breast imaging segmentation and classification using deep learning algorithms. During the phase off-line, an SNGAN model was previously trained for synthetic image generation, and subsequently, these images were used to pre-trained SAM and ResNet18 segmentation and classification models. During phase online, the BraNet app was developed using the react native framework, offering a modular deep-learning pipeline for mammography (DM) and ultrasound (US) breast imaging classification. This application operates on a client–server architecture and was implemented in Python for iOS and Android devices. Then, two diagnostic radiologists were given a reading test of 290 total original RoI images to assign the perceived breast tissue type. The reader’s agreement was assessed using the kappa coefficient. The BraNet App Mobil exhibited the highest accuracy in benign and malignant US images (94.7%/93.6%) classification compared to DM during training I (80.9%/76.9%) and training II (73.7/72.3%). The information contrasts with radiological experts’ accuracy, with DM classification being 29%, concerning US 70% for both readers, because they achieved a higher accuracy in US ROI classification than DM images. The kappa value indicates a fair agreement (0.3) for DM images and moderate agreement (0.4) for US images in both readers. It means that not only the amount of data is essential in training deep learning algorithms. Also, it is vital to consider the variety of abnormalities, especially in the mammography data, where several BI-RADS categories are present (microcalcifications, nodules, mass, asymmetry, and dense breasts) and can affect the API accuracy model.

Keywords Breast cancer · Mobil app · Deep learning · Ultrasound · Mammography

1 Introduction

Today, in the healthcare landscape, artificial intelligence tools hold great promise for clinicians by enhancing breast cancer diagnostics and tailoring treatment strategies to

match the disease’s characteristics [1–3]. However, in the same line, there are some alternatives, such as command-line tools with shell scripts [4] and manual, semi-automated, and fully automated methods for image processing [5]; these options are not user-friendly for specialists and researchers

✉ Yuliana Jiménez-Gaona
ydjimenez@utpl.edu.ec

Diana Carrión-Figueroa
diana.carrionf@iess.gob.ec

Patricio Corral-Domínguez
patriciocorral@hotmail.com

¹ Departamento de Química y Ciencias Exactas, Universidad Técnica Particular de Loja, San Cayetano Alto s/n CP1101608, Loja, Ecuador

² Instituto de Instrumentación para la Imagen Molecular I3M, Universitat Politècnica de Valencia, 46022 Valencia, Spain

³ Theoretical and Experimental Epistemology Lab, School of Opto ΩN2L3G1, Waterloo, Canada

⁴ Hospital-IESS del Sur de Quito, Av. 18 de Septiembre, Quito, Ecuador

⁵ Corporación Médica Monte Sinaí-CIPAM (Centro Integral de Patología Mamaria) Cuenca-Ecuador, Facultad de Ciencias Médicas, Universidad de Cuenca, Cuenca 010203, Ecuador

⁶ Department of Systems Design Engineering, Physics, and Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada

without a background in computer science. Furthermore, the available graphical interface tools are often task-specific [6, 7], focusing on contour delineation, segmentation, or classification.

In this context, radiomics constitutes an emerging field in medical imaging and offers the potential to extract diagnostic and prognostic information from 2D grayscale images by analyzing lesion features [1, 2]. Hence, graphical and mobile tools are elevating the role of radiomics in biomedical research, potentially serving as a second opinion for radiologists in breast lesion detection. Specifically, computer-aided diagnosis (CAD) systems based on deep/machine learning (DL/ML) play a crucial role in addressing various computer vision challenges, such as medical image pre-processing with super-resolution [8–11] and denoising, data augmentation [12–15], medical image segmentation [16–18] (e.g., NiftyNet [6], MIScnn [16], and NiftySeg [17]), image classification [19], computer-assisted interventions [5], image recognition [20], and annotation [5].

In the context of detecting cancer, there are several radiomic projects, CAD based on deep/machine learning (DL/ML) systems, and studies that propose different artificial intelligence techniques that help to provide decision support for many applications in the patient care processes, such as lesion detection, characterization, cancer staging, and treatment planning. The major challenge in this field of research is to build a fully automatic CAD system that can analyze large quantities of images to provide an accurate diagnosis and, at the same time, robust enough to handle the biological variations in humans [21].

The most successful DL algorithms used in the processing of medical images are convolutional neural networks (CNNs), generative adversarial networks (GANs), and recurrent neural networks (RNNs), which play a crucial role in improving healthcare outcomes by providing accurate and efficient analysis in processing medical images, each offering unique capabilities in data augmentation, pattern recognition, and feature extraction [22]. In the early detection of breast cancer, CAD systems have several stages: (i) image collection, (ii) annotation and detection of tumors based on the region of interest (ROI), (iii) segmentation, (iv) classification based on the ROI shape using deep learning models, and (v) performance evaluation of the models [23, 24].

Image collection and annotation are the main challenges in performing large-scale medical image analysis using DL algorithms. Some CNNs-based options to consider as mask segmentation for detected tumors in medical images are you only look once (YOLO) [25], region-based convolutional neural network (R-CNN) [26, 27] and their variants (Mask R-CNN [26] and Faster R-CNN [27]), deep neural networks such as natural language process (NLP), which can help us to automatically identify and extract relevant information from radiology clinical reports and images [28].

Although there is a variety of CAD systems developed concerning breast cancer, it is also important to mention that there are systems deployed in mobile applications for use in the smartphone, e.g., in [29], an automated breast cancer diagnosis system on mobile phones for taking photos of ultrasound reports was implemented. The authors include the automatic extraction of intricate image features by convolutional neural networks (CNNs) and the precise classification of breast masses. It eliminates the need for manual feature engineering and reduces human error. These applications streamline the diagnostic process, increase efficiency, and, most importantly, enhance patient outcomes by providing reliable, consistent, and accessible early breast cancer detection and treatment tools.

2 Related work

In this section, we will briefly introduce NLP and CNN modeling as more recent approaches using neural networks and discuss how several authors have used these models in radiomics and biomedical applications.

One of the main fundamentals of NLP is extracting image information using patterns such as the accession number, series number, and image number. Information about the imaging modality, magnetic resonance imaging (MRI), CT (computed tomography), positron emission tomography (PET), ultrasound (US), and mammography imaging can be relevant, too. It can be extracted from the accession number and image number, where the patient identification number (ID) can be appropriate if the patient's history is of interest.

Linna et al. [30] indicate that NLP tools in radiology and other medical settings have been used for information retrieval and classification. NLP-based algorithms have opened more possibilities for medical image processing, detecting findings, and giving possible diagnoses [31]. Wang et al. [32] suggested that cancers are the most common subject area in NLP-assisted medical research on diseases, with breast cancers (23.30%) and lung cancers (14.56%) with the highest proportion of studies. Also, Luo et al. [33] specified that NLP is useful for creating new automated tools that could improve clinical workflows and unlock unstructured textual information contained in radiology and clinical reports for the development of radiology and clinical artificial intelligence applications.

Prabadevi et al. [34] proposed a machine learning system using WEKA algorithms to detect cancer staging classification. Buckley et al. [35] used NLP to extract clinical information from > 76,000 breast pathology reports, the model of which demonstrated a sensitivity and specificity of 99.1% and 96.5% compared to expert humans. Chen et al. [36] proposed an NLP extraction pipeline system that accepts scanned images of operative and pathology reports. The

system achieved 91.9% (operative) and 95.4% (pathology) accuracy. The pipeline accurately extracted outcomes data pertinent to breast cancer tumor characteristics, prognostic factors, and treatment-related variables. Liu et al. [37] implemented an NLP program to extract index lesions and their corresponding imaging features accurately from the text of breast MRI reports.

The NLP system demonstrated 91% recall and 99.6% precision in correctly identifying and extracting image features from the index lesions. The recall and precision for correctly identifying the BI-RADS categories were 96.6% and 94.8%, respectively. Kirillov et al. [38] created the NLP-based segment anything model (SAM) as a mask extraction and promptable segmentation task. Thus, it can transfer zero-shot [39] to new image distributions.

The NLP system demonstrated 91% recall and 99.6% precision in correctly identifying and extracting image features from the index lesions. The recall and precision for correctly identifying the BI-RADS categories were 96.6% and 94.8%, respectively. Kirillov et al. [38] created the NLP-based segment anything model (SAM) as a mask extraction and promptable segmentation task. Thus, it can transfer zero-shot [39] to new image distributions.

Likewise, in a CAD system, the classification task is an important step after the segmentation process. The most widely used deep learning-based algorithms for image classification are CNN models (ResNet [40], DenseNet [41], NasNet [42, 43], VGG-16 [44], GoogLeNet [45], and Inception-V3 [46]).

Several authors [47–51] have used models for benign and malignant breast mass classification. The CNN used for breast classification is divided into two main categories: (i) novo-trained model (e.g., Scratch) and (ii) transfer learning-based models that exploited previously trained networks (e.g., AlexNet, VGG-Net, GoogLeNet, and ResNet) [47]. In [48], the ResNet model was used as a classification training model using an original and synthetic mammography (DDSM) dataset, obtaining a performance of 67.6 and 72.5%, respectively.

In [49], several CNN models were proposed (GoogLeNet, Visual Geometry Group Network (VGGNet), and ResNet), to classify malignant and benign cells using average pooling classification. The results overcome all the other deep learning architectures in terms of accuracy (97.67%). However, the choice of architecture depends on the specific problem and involves commitments between factors such as model size, computational efficiency, and accuracy.

Despite the extensive availability of medical radiomic tool research and CAD-based deep learning systems [52, 53], this technology has limited support within mobile app infrastructure for 2D breast medical image analysis. Consequently, the BraNet's workflow has two main phases off-line and on-line, to achieve the following aims: (i) to develop a

mobile app based on deep learning models for segmenting and classifying 2D breast images into benign and malignant lesions and (ii) to implement statistical metrics as a prediction performance evaluation tool.

3 Methods

3.1 Data collection

We collected seven open-access breast image databases, including three datasets of breast ultrasound (US) images and four datasets of mammography images.

- (i) Breast Ultrasound Images Dataset (BUSI): this dataset, gathered by [43], comprises 780 images (133 normal, 437 benign, and 210 malignant).
- (ii) Dataset A: collected by Rodrigues et al. [54] available at (<https://data.mendeley.com/datasets/wmy84gzngw/1>), Dataset A contains 250 breast US images (100 benign and 150 malignant).
- (iii) Dataset B: Comprising 163 US images, these data were acquired from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell, Spain [55].
- (iv) CBIS-DDSM: Curated Breast Imaging Subset–Digital Database for Screening Mammography, accessible at (<https://n9.cl/qt148>), this database comprises 2620 cases [56].
- (v) mini-MIAS (Mammographic Image Analysis Society): available at <http://peipa.essex.ac.uk/info/mias.html>, includes 322 (208 normal, 63 benign and 51 malignant images) Medio Lateral Oblique (MLO) mammograms from 161 patients [57, 58].
- (vi) Inbreast: this dataset comprises a total of 115 images and can be found at (<https://biokeanos.com/source/INBreast>) [59].
- (vii) VinDr-Mammo: introduces a large-scale full-field digital mammography dataset of 5,000 four-view exams (<https://physionet.org/content/vindr-mammo/1.0.0/>) [60].

3.2 Pretraining models in phase off-line

3.2.1 Data normalization and automatic ROI annotation

An ROI annotation is needed from a large dataset of US and mammography images from the above public database to improve the previously trained GAN and ResNet models and their computational performance. The breast images vary in size, see Table 1.

Thus, it is necessary to perform transformations and standardize the images taken at different sizes to a single dimension ($128 \times 128 \times 1$ pixel). It was also necessary to

Table 1 US and mammography ROIs

Type	Training	Database	Image size (pixels)	Benignant	Malignant
US	I	BUSI	500×500	427	201
		Dataset B	-	100	150
		BUS (UDIAT)	760×570	109	54
		Total 1041		636	415
DM	I	Mini-MIAS	1024×1024	118	91
		INbreast	3328×4084	2106	144
			256×3328		
	II	CBIS-DDSM	3784×5912	1225	779
		VinDr-Mammo	3518×2800	893	1236
		Total 5892		4342	1550

transform it to a single channel (grayscale pixel) and normalize it in the range [- 1,1] with a mean of 0.5 and a standard deviation of 0.5. The torch-vision (pytorch) library and Jupyter notebook algorithm (crop_vindr_images.ipynb) were used as the image annotation region processes to identify ROIs that may contain lesions.

Figure 1 details the overall process followed in this study.

3.2.2 User ROI extraction and segmentation

As other studies have pointed out [61], to improve the detection accuracy, smaller patches (i.e., ROIs) where all breast masses and micros (e.g., cysts and calcification) are included inside this extracted area are generated from the original mammogram. In most mammogram images, 32 to 56% are background pixels, which do not contribute to breast cancer diagnosis.

In this research, the segment anything model (SAM) [38] (an encoder-decoder architecture based on NLP prompt-based learning) [35] was trained as automatic ROI

segmentation before being implemented in the Module 5 (BraNet application phase on-line). SAM is an open-source software, and the quality of the segmentation masks was rigorously previously evaluated, with automatic masks deemed high quality and effective for training models, leading to the decision to include automatically generated masks.

NLP tasks include sentence boundary detection, tokenization, and problem-specific segmentations, and the *SamAutomaticMaskGenerato* function was used for automatic mask extraction [28]. SAM model is available under a permissive open license (Apache 2.0) at <https://segment-anything.com>.

The SAM predefined hyperparameters used are as follows: points_per_side (32), points_per_batch (64), pred_iou_thresh (0.88), stability_score_thresh (0.95), stability_score_offset (1.0), box_nms_thresh (0.7), crop_n_layers (0), crop_nms_thresh (0.7), crop_overlap_ratio (512/1500), crop_n_points_downscale_factor (1), point_grids (Null), min_mask_region_area (0), and output_mode ('binary_mask'). Also, SAM accuracy ROI segmentation is evaluated

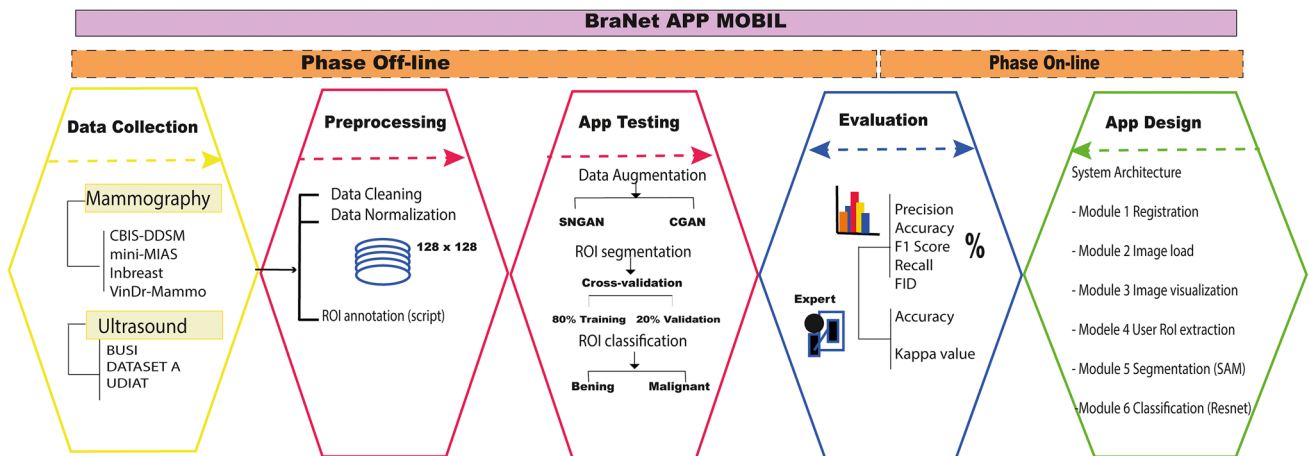


Fig. 1 The BraNet’s workflow has two main phases off-line: (i) breast data collection from public databases and (ii) data preparation and app testing and evaluation. On-line: (iii) system architecture design

and (iv) statistical metrics comparison between the mobile application and human experts

by the intersection-over-union (Jaccard index) metric [18] using *calculate_stability_score* function.

The model was trained on a large and diverse set of masks of mammography and US images. These ROIs were previously extracted from the Mini-MIAS, Inbreast, and VinDr-Mammo databases (corresponding annotated bounding boxes are available in a.csv file). However, it is worth noting that its authors had already classified the ROIs from the CBIS-DDSM dataset; thus, no additional ROI extraction was required. Regarding US images, RoIs were extracted from the BUSI and Dataset B (UDIAT), excluding Dataset A, because it already contained RoIs.

A total of 6592 breast ROI images were used for pre-trained SAM and GAN models, with 4463 mammography images (benign and malignant) and 1041 US images, as shown in Table 1.

3.2.3 Data augmentation

This technique was previously employed in the phase off-line to mitigate the risk of overfitting effectively. To generate new realistic images and improve BraNet’s classification task performance, all ROIs were previously augmented by a GAN using the spectral normalization technique (SNGAN). SNGAN introduces a novel weight normalization technique known as spectral normalization to enhance the training stability of the discriminator network [62, 63], serving as the foundation for synthetic image generation, which use Hinge loss function (see Eq. 1).

$$\begin{aligned}
 L_D &= -E_{(x,y) \sim P_{data}} [\min(0, -1 + D(x, y))] \\
 &\quad - E_{z \sim P_z, y \sim P_{data}} [\min(0, -1 + D(G(z), y))] L_G \\
 &= -E_{z \sim P_z, y \sim P_{data}} D(G(z), y)
 \end{aligned}
 \tag{1}$$

where P_{data} is the real data distribution, $P(z)$ is a prior distribution on noise vector z , $D(x)$ denotes the probability that x comes from the real data rather than generated data, $E_{x \sim P_{data}}$ represents the expectation of x from real data distribution P_{data} , and $E_{z \sim P_z}$ is the expectation of z sampled from noise.

The *clean-fid* library was used to obtain the FID value, using the Tensorflow and PyTorch libraries, some original implementations of the metric were taken from Parmar et al. [64]. The GAN model was trained using a cross-validation technique in Google Colab Pro 1 GPU model V100 with CUDA cores execution; with the hyperparameters detailed in Table 2.

3.2.4 Cross-validation analysis

The technique divides the dataset into multiple folds and trains (DM: training I (4463) and training II (6592) and US: training I (1041)) the model on different subsets while validating the remaining fold can provide a more robust estimate of the model's performance, effectively mitigating the risk

Table 2 Hyperparameter tuning of deep learning models

Hyperparameter	SNGAN	ResNet
	DM/US	DM/US
Batch size	64/32	32/16
Image size	128 × 128	128 × 128
Nro epochs	100	100
Learning rate	2 × 10 ⁻⁴ /1.5 × 10 ⁻⁴	2 × 10 ⁻⁵
Optimizer	Adam	Adam
Activation function	LeakyReLU	ReLU
β1	0.3	0.1
β2	0.999/0.75	0.9
Latent vector	100	-
Loss function	Hinge/BCE	2.190
Optimization function (discriminator)	LeakyReLU	-
Optimization function (generator)	ReLU	-

Table 3 Training and validation datasets of DM 80% (5273) and 20% (1319) and US 80% (832) and 20% (209) breast images

Classes	Training datasets		Validation dataset	
	US	DM	US	DM
Bening	505	3471	131	871
Malignant	327	1802	78	448
Total	832	5273	209	1319

of overfitting. It helps detect overfitting early and tune the model accordingly. A total of 6933 benign and malignant ROIs were split into 80% training and 20% validation, using the Sklearn library from Pytorch, see Table 3.

3.2.5 ROI classification process

Before implementing Module 6 (ROI image classification) in the BraNet mobil interface, the ResNet model was pre-trained on a large set of generated mammography and US ROIs using also cross-validation technique.

ResNet18 training model The ResNet18 CNN-deep learning-based classification model has been widely used in medical image classification, especially in breast lesion diagnosis and detection, and was chosen for its effectiveness in transfer learning, offering reduced training time and the automatic extraction of features [40]. This approach effectively mitigates the issues of vanishing or exploding gradients that can arise from increasing neural network depth, ultimately leading to improved accuracy [65–68].

Thus, to train the ResNet model and distinguish between malignant and benign breast lesions, the datasets were divided in two categories: (i) dataset A (original + synthetic

ROIs) and (ii) dataset B (synthetic ROIs). The model consists of three convolutional layers and two fully connected layers. The kernel size for the first convolutional layer is 5×5 , and, for the rest, 3×3 . The size of the first and the second fully connected layers are 128 and 2 (the number of classes), with a dropout of 0.5. After the flattening and the first fully connected layers, the ReLU activation function for all layers except the output layer, where softmax was used. The model was pre-trained with the PyTorch library using a Google Colab Pro 1 GPU model V100 with CUDA cores execution; the training hyperparameters are outlined in Table 2.

3.3 System architecture in phase on-line

The system architecture consists of two primary components facilitating scalability and system maintenance: (i)

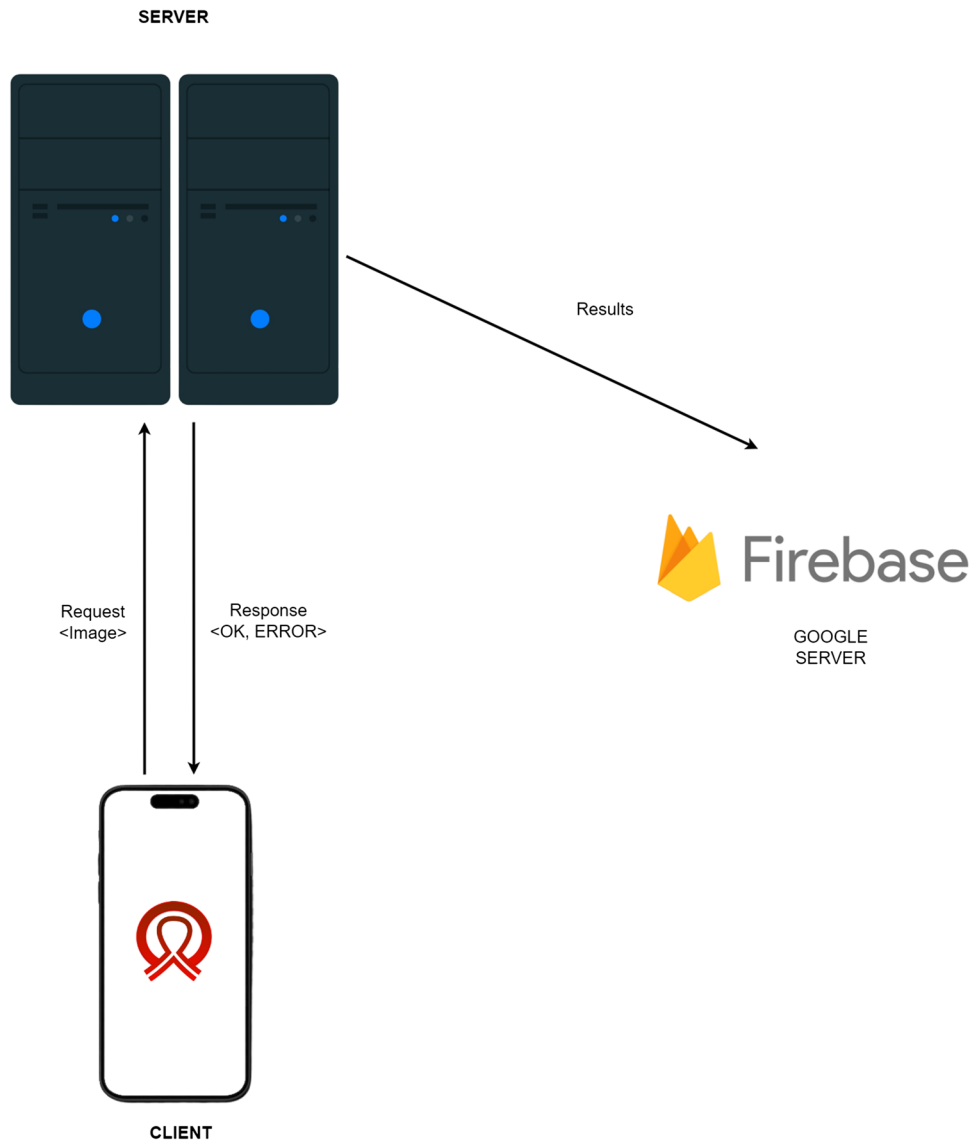
the mobile application and (ii) the backend server, following a client–server architecture.

The mobile application is a client that communicates with the backend server to request services and image analysis. The backend server processes these requests and returns the results to the client for display (see Fig. 2). The backend server was developed using react native and was implemented in the Python programming language.

The mobile application comprises several interrelated components:

- Module 1: Registration, Login Synchronization, and User Profile Information. Data generated by the application, such as images and metadata, is stored in Firebase (a mobile and web application development platform). Firebase is also used for user authentication, mobile application registration, and log in.

Fig. 2 Client–server architecture to BraNet App



- Module 2: Image Import allows users to upload breast ultrasound and mammography images in PNG, JPG, and JPEG formats, with a maximum file size of 10 MB.
- Module 3: Visualization Area (a history of image analysis results, image analysis capabilities, and user assistance).
- Module 4: User ROI extraction.
- Module 5: ROI segmentation (see Section 2.2.2).
- Module 6: ROI Image Classification (see Section 2.2.4).

3.4 Evaluation metrics in phase off-line and on-line

3.4.1 Quality of synthetic image

The FID and KID quantitative feature-based metrics have been applied to evaluate the quality of real and synthesized ROIs generated by GANs models and to compute the distance between the vector representation of the synthesized and authentic images.

Fréchet inception distance (FID): FID compares the distributions of the original and synthetic images to assess how well the generated images represent the training dataset. Lower FID scores indicate better quality images [69], and it is calculated as shown in Eq. 2:

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}) \quad (2)$$

Here, μ_r represents the mean of the feature vector calculated from the real images, μ_g is the mean of the feature vector calculated from the fake images, Σ_r is the covariance of the feature vector from the real images, and Σ_g is the covariance of the feature vector from the fake images.

Kernel inception distance (KID) KID employs the cubic kernel to compare the skewness, mean, and variance [69]. A lower KID value signifies a higher visual similarity between the actual and generated images. The cubic polynomial kernel is defined as shown in Eq. 3):

$$k(x, y) = \left(\frac{1}{d}x^T y + 1\right)^3 \quad (3)$$

where d represents the dimension of the feature space for vectors x and y .

Table 4 Confusion matrix to distinguish between two classes (benign, and malignant)

Actual classes	Classes	Predicted classes		Measures
		Positive	Negative	
	C_1 (benign)	TP	FP	PPV
	C_2 (malignant)	FN	TN	NPV
	Measures	Sen	Spec	Acc

3.4.2 BraNet’s classification performance evaluation

For assessing the *BraNet’s* classifier’s performance, we employed a confusion matrix, *F1* score (Dice), accuracy (Acc), precision (Prec), sensitivity (Sen), recall, and specificity (Spec) [24] metrics (see Tables 4 and 5). The accuracy of the model was calculated using statistical score libraries such as the classification report and confusion matrix from the Python sci-kit-learn module.

3.4.3 Human expert evaluation

Two senior radiologists were asked to assess, annotate, and classify images independently to ensure that BI-RADS categories are correctly assigned. Representative original ROI images for each breast type is available in https://drive.google.com/drive/folders/1HMeqPfl8qL58hAqwVpZupH6uq4W_kHrI.

A comparison between the images tested by human experts and those annotated in public databases was conducted using an independent test set of 212 mammography images (47 malignant and 165 non-malignant) and 78 US images (24 malignant and 54 non-malignant).

Two diagnostic radiologists (reader 1 with 20 years of experience and reader 2 with 13 years of experience) were given a reading test consisting of 290 total original ROI images to assign the perceived breast tissue type. The readers rated each image as (1) benign or (0) malignant.

Kappa coefficient and overall accuracy Furthermore, the agreement between the two readers’ answers (considering all elements of error matrix) was assessed by determining

Table 5 Validation assessment metrics

Model	Equation
Accuracy	$Acc = \left(\frac{TP+TN}{TP+TN+FP+FN}\right)$
Sensitivity	$Sen = \left(\frac{TP}{TP+FN}\right)$
Specificity	$Spec = \left(\frac{TN}{TN+FP}\right)$
Precision	$Prec = \left(\frac{TP}{TP+FP}\right)$
<i>F1</i> score	$F1 \text{ score} = 2 \times \left(\frac{Prec \times recall}{prec + recall}\right)$

the kappa coefficient (K), using the ranges between 0 (when there is no agreement) and 1 (when there is substantial agreement), and is calculated using the Eq. 4. The error matrix was calculated by comparing the two readers' choices from five possibilities and was interpreted as follows: <0.2 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 high; and 0.81–1.0 almost perfect [70].

$$k = \frac{P_o - P_e}{1 - P_e} \tag{4}$$

P_o is the correctly allocated samples (agreement cases), and P_e is the hypothetical random agreement.

The overall accuracy (Eq. 5) allows the description of model performance and is calculated by dividing the total number of correctly classified samples by the total number of samples.

$$Acc = \frac{C_s}{N_s} \tag{5}$$

C_s is the number of correct samples classified, and N_s is the total number of samples.

4 Results

The main results are categorized into two phases, off-line and on-line. First, we introduced the preprocessing and pre-training models' section with data augmentation (GAN), segmentation (SAM), and classification (ResNet) algorithms. Then, we presented the on-line phase with a practical utility of BraNet's user interface and its modules.

4.1 Preprocessing and pretrained models

4.1.1 Synthetic data to feed the classification network

A significantly number of synthetic RoIs (10,000 (training I) and 2000 (training II) mammography RoIs and 4000 US RoIs

(training I)) were generated by SNGAN to feed the classifier. The loss function and accuracy of the generator and discriminator play a crucial role in assessing the training stability and performance of GANs. A stable GAN is characterized by a discriminator loss around 0.5 or higher than 0.7, while the generator loss typically ranges from 1.0 to higher values like 1.5, 2.0, or even more. The accuracy of the discriminator, both on real and synthetic images, is expected to hover around 70 to 80%. Appendix Table 9 presents the accuracy plot to SNGAN.

The average FID and KID values in SNGAN are 58.80/0.052 and 52.34/0.051 for mammography training I and training II, respectively, and 116.85/0.06 for the training I in US (see Appendix Fig. 6). The lowest values indicate that SNGAN-generated synthetic images closely resemble to the original mammography and US images in clinical characteristics, suggesting their potential utility in clinical data augmentation and training, particularly for enhancing diagnostic skills in breast imaging.

4.1.2 ResNet training model

The model shows the highest accuracy in US image classification (see Table 6) concerning the mammography dataset. Although the network received more mammography images (6592) as input (Mini-MIAS, Inbreast, CBIS-DDSM, and VinDr-Mammo) with respect to the small number (1041) of US data (BUSI, UDIAT, and DATASET A). It means that not only the amount of the data is important to train deep learning algorithms. Also, it is important to consider the variety of abnormalities especially in the mammography data, such as microcalcifications, nodules, mass, asymmetry, and dense breasts, because it can improve the accuracy of the ResNet training model.

Therefore, it is essential to monitor the evolution and performance of the models using training and validation datasets, see Fig. 3a–d. Figure 3a, c displays the loss and accuracy values concerning each epoch during the ResNet training and validation model using mammography and US images, while Fig. 3b, d shows the accuracy, $F1$ score, recall, and precision by each epoch in both image types. Appendix Table 10 shows the details of the network training and validation dataset.

Table 6 ResNet statistical performance evaluation in US and DM image classification

RESNET18	Image modality					
	Training I (1041)		Training I (4463)		Training II (6592)	
Image type	US (%)		DM (%)		DMb(%)	
Classes	Benignant	Malignant	Benignant	Malignant	Benignant	Malignant
Accuracy	94.7	93.6	80.9	76.9	73.7	72.3
Precision	97	89	92	56	84	59
Recall	93	95	81	77	74	72
$F1$ score	95	92	86	65	78	65

The BraNet App Mobil exhibited the highest accuracy in benign and malignant US images (94.7%/93.6%) classification compared to DM during training I (80.9%/76.9%) and training II (73.7/72.3%). And the ResNet model does not improve the accuracy of benign and malignant ROI lesion classification during training II compared to the previous training I.

4.2 The BraNet and its graphical user interface (GUI)

The mobile app’s user interface was developed using Python v3.11 and React Native as a JavaScript framework for creating native mobile applications compatible with iOS and Android platforms. The interface is composed of several modules, each serving distinct purposes:

- **Module 1:** Registration, Login Synchronization, and User Profile Information: This module handles user registration and login functionalities, synchronizing user data and providing access to user profile information.
- **Module 2:** Image Import: Users can import images in standard picture formats, such as JPG, JPEG, and PNG, with a maximum size limit of 10 MB.
- **Module 3:** Visualization Area: This area displays loaded images. The selected image is displayed in grayscale, preserving the original image's aspect ratio (see Fig. 4a).
- **Module 4:** Manual ROIs Extraction: This module allows users to manually or semi-automatically create masks

and define ROIs within the selected image. Masks are represented as binary matrices with the same dimensions as the loaded mammography image, where true values indicate the ROIs. Users can customize the size and sampling method for ROIs.

- **Module 5:** ROI Segmentation: Users can segment a subset or the entire set of features from the segmentation section (as shown in Fig. 4b). Before performing calculations on the image, the user must add at least one ROI in the “Regions and Masks.”
- **Module 6:** ROI Image Classification: This module employs the ResNet18 model to classify ROI images into benign and malignant classes. Example output classes are provided in Fig. 4b.

The BraNet’s graphical user interface enhances the user experience by providing intuitive image analysis and classification tools, making it a valuable resource for medical professionals and researchers. The practical usage of the tool can be accessed via the following link: <https://drive.google.com/file/d/1d1vnjQ6LqOd0fdz65eaVg791d7cFRPWO/view>

4.2.1 Comparison of the BraNet with human experts’ evaluation

The accuracy percentages of correct rates between benign and malign images classification for readers 1 and 2 were

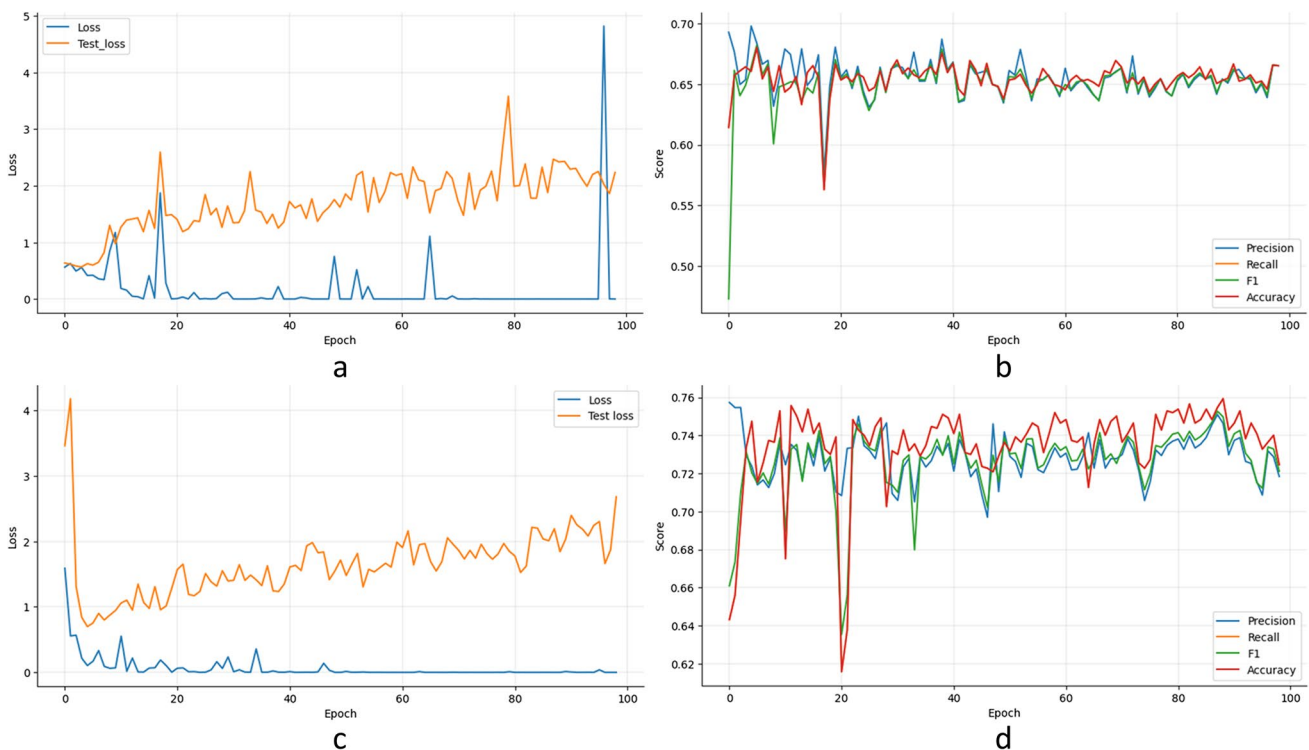


Fig. 3 Training II and testing plots for mammography images: **a** loss vs. acc (real ROI data), **b** loss vs. acc (real + data augmentation), **c** evaluation metrics (real ROI data), and **d** evaluation metrics (real + data augmentation)

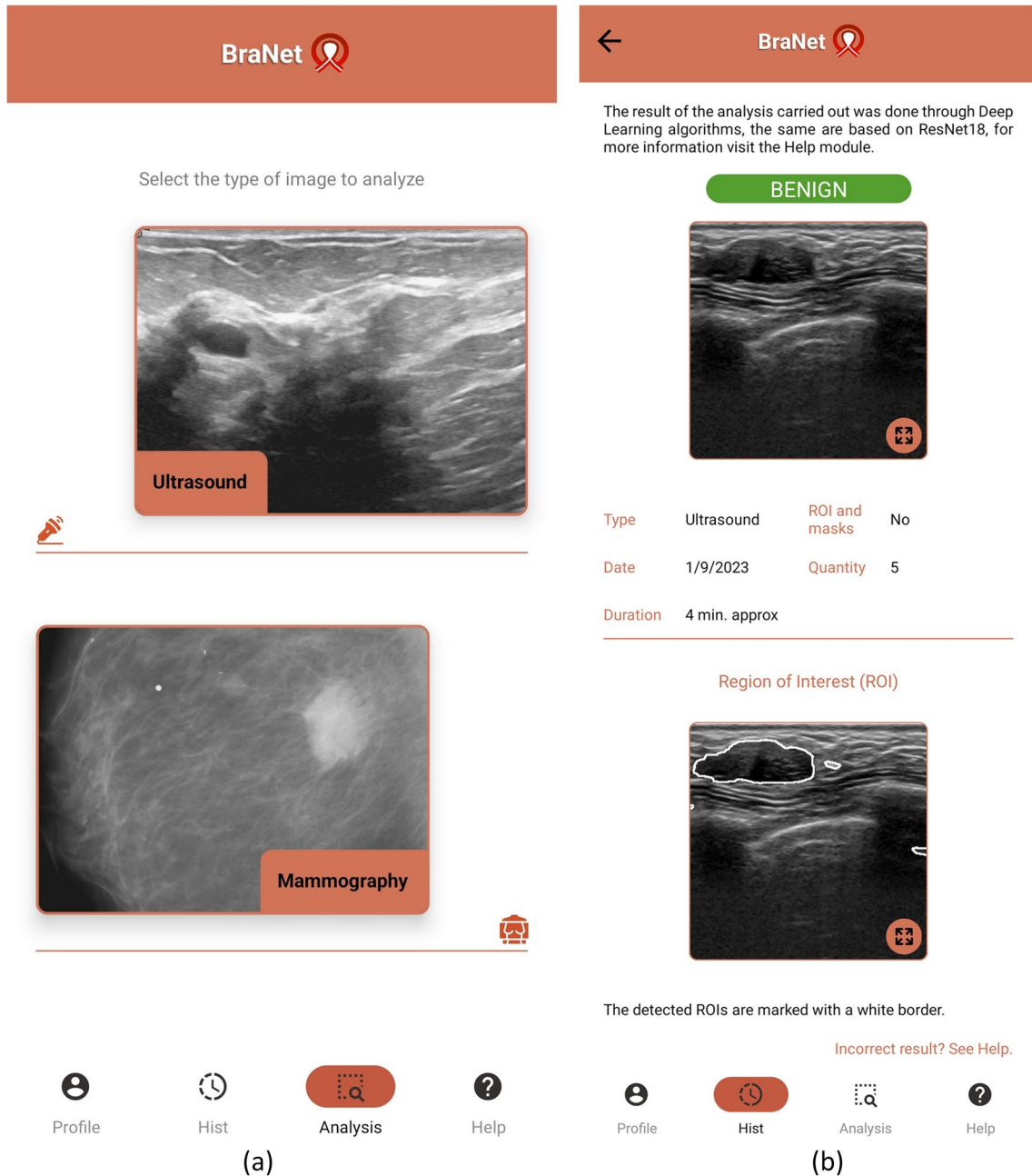


Fig. 4 BraNet user interface within the toolbox. **a** Upload the breast image type. **b** US breast ROI selection and classification as benign class

29% and 42%, respectively. The reader agreement was assessed using the kappa coefficient, which values are 70% and 71% in mammography and US classification, respectively. Table 7 indicates a fair agreement (0.3) for mammography images and moderate agreement (0.4) for ultrasound images in both readers, with a change in prevalence from the lowest in US images to the highest value in mammography images, resulting in a corresponding change in sensitivity (19.2/60) to specificity (51/84.4) percentage points. This effect was statistically significant ($p < 0.05$) for either sensitivity or specificity in both image types.

5 Discussion

The pressing need to transition automatic medical image classification by CAD systems from research laboratories into practical clinical applications is evident. BraNet’s aims to provide an API for setting up a breast image classification pipeline with ROI mask extraction and segmentation capabilities. The tool offers an open-source solution for processing US and mammography images, complete with statistical metrics for evaluating model performance.

There have been many published examples of AI algorithms that demonstrate excellent performance in cancer

Table 7 Interrater reliability, Cohen's kappa, and statistical values for 2 raters in both classes

<i>Method</i>	<i>DM</i>	<i>US</i>
<i>Subjects</i>	212	78
<i>Agreement %</i>	70	71
<i>Kappa</i>	0.294	0.426
<i>p-value</i>	< .001	< .001
<i>Sensitivity</i>	19.2%	60.0%
<i>Specificity</i>	51.6%	84.8%
<i>Prevalence</i>	69.5%	57.7%
<i>Accuracy</i>	29.0%	70.5%
<i>PPV</i>	47.5%	84.4%
<i>NPV</i>	21.9%	60.9%

detection for screening mammography. These include several algorithms trained and evaluated on private and public data sets. Table 8 compares BraNet's performance against other state-of-the-art medical image classification applications.

However, there is a significant gap in understanding how these AI applications will perform with multimodal images in the real world when radiologists use them in clinical practice [75, 76].

The BraNet Mobile App is an open interface for classifying specific 2D breast image types using deep learning models. It is believed that this is the first system for breast cancer diagnosis deployed on mobile phones to both types of images. The API's development comprises two main phases: (i) off-line to pre-train the deep algorithms and (ii) on-line to release the app, which includes several modules, including model selection, model extraction (by a human expert), segmentation (SAM model), model classification (ResNet18 model), and model evaluation.

During the off-line phase, the pre-trained GAN algorithm was implemented as synthetic image generation, and the image quality was evaluated by two feature-based metrics FID and KID. It is widely acknowledged that the preprocessing images, quality, and diversity of the training dataset greatly impact the training of GAN and CNN deep learning models [77–79]. The lower FID and KID values mean a higher visual similarity between the real and generated images. The results (Appendix Fig. 6) indicate that the SNGAN model is suitable for mammography and US synthetic data generation with average values of FID = 52.4/ KID = 0.051 for mammography and FID = 116.85/ KID = 0.06 for US.

With these datasets Dataset A (original + synthetic ROIs) and Dataset B (real ROIs), the classification model was trained. Table 6 and Fig. 3 show the accuracy results averaged in BraNet ROI classification are as follows: (i) training I in US (94.7 (B)/93.6 (M)) and DM (80.9 (B)/76.9 (M)) and (ii) training II in DM (73.7 (B)/72.3 (M)). The result demonstrated that ResNet model during training II

with original + synthetic images (where the VirDrMammo database was added) did not improve the accuracy (73.7 / 72.3%) concerning Training I (80.9/76.9). In comparison, with radiological experts, accuracy in DM was 29% concerning with 70% in DM for both readers. These results show that both API and Readers obtained a better percentage of accuracy in classifying the ROIs of mammography images than US images.

A final comparison between BraNet and radiological experts' evaluation demonstrates that for the all-breast image types, reader accuracy was higher with US images (75%) than with original ROI images from public databases. The reader agreement was 70% and 71% in mammography and US classification, respectively. The kappa value indicates a fair agreement (0.3) for mammography images and moderate agreement (0.4) for ultrasound images in both readers (Table 7). This can be contrasted with BraNet classification accuracy (Table 8), where the API shows the highest accuracy in US image classification (Table 6) concerning the mammography dataset. Although the network received more mammography images (5892) with respect to US (1041). It means that not only the amount of the data is important to train deep learning algorithms. Also, it is important to consider the variety of abnormalities especially in the mammography data, where several BI-RADS categories are present (microcalcifications, nodules, mass, asymmetry, and dense breasts), and can be affect the accuracy in the ResNet training model.

According to the previous results, some limitations in implementing BraNet must be addressed in future work. One is the need to classify and characterize images based on different abnormalities, such as architectural distortion, asymmetry, mass, and microcalcification. BraNet no was trained using different breast tissue types and variations in mammography and US imaging techniques; the ROI classification process was performed only using two classes 1 (benignant according to BI-RADS 1–3) and 0 (malignant according to BI-RADS 4–6) categories. Oyelade et al. [80] indicate that is better to focus on previously classified and characterizing abnormalities into architectural distortion, asymmetry, mass, and microcalcification so that training distinctively learns the features of each abnormality. It generates sufficient images for each category before training a GAN model.

Thus, in future work, we plan to annotate the datasets with more fine-grained classes to get more targeted training in GAN and CNN models. Moving forward, we should consider pre-processing with denoising, super-resolution, improving the overall image quality and reducing blur and artifacts. Also, previous breast tissue types of classification are needed to obtain a diverse range of synthetic data, resulting in a more accurate image generation and classification process using GAN and convolutional algorithms.

Table 8 Comparison of the BraNet's performance against other deep learning applications

Author	Application name	Description	Acc/Sen/Spec/Prec /AUC (%)
Gibson et al. [6]	NiftyNet	A DL open-source platform used in three medical image analysis applications (MRI, CT and US), including a conditional GAN model as ultrasound image generation	88/7.5/9.1/-/-
Pang et al. [71]	TripleGAN	Method to perform data augmentation in breast US images and feed a CNN mode to classify breast masses	90.41/87.94/ 85.86/-
Al-Dhabyani et al. [43]	AlexNet + GAN (CNN)	US breast classification with data augmentation. The model examines two different methods: a CNN approach and a transfer learning (TL). The results confirm an overall enhancement using augmentation methods with TL classification methods	78/-/-/-
	-BUSI		80/-/-/-
	-Dataset B		84/-/-/-
	-BUSI+DatasetB		94/-/-/-
	TL		92/-/-/-
	-BUSI		99/-/-/-
	-Dataset B		99/-/-/-
-BUSI+DatasetB			
Jiménez et al. [72]	Radiomic tool	Colposcopy image classification combining UNET + SVM as segmentation and classification cervix abnormalities	80/70/48.8/-
Dihge et al. [73]	NILS	A web-based tool for noninvasive lymph node staging in breast cancer	-/90/34/-/71
To T. et al. [74]	DUV-WSI	DUV-WSI Deep ultraviolet (DUV) fluorescence scanning microscopy provides rapid whole-surface imaging (WSI) of breast tissues. Images are split into small patches (512×512), and features are extracted using a pre-trained ResNet 50 as patch classification	81.7/91.7/66.7/-/-
Qi et al. [29]	Deep-CAD system	The breast cancer system is deployed on mobile phones, takes a photo of the US as input, and performs diagnosis on each image. Then, the system to classify images into malignant and non-malignant using CNNs	89.34/87.31/87.49
Ours	BraNet	A deep learning tool for breast regions classification using mammography and US images DM (TRAINING I)	Acc 94.7/93.6 Prec 97/89 Recall 93/95 F1 score 95/92
		DM (training II)	Acc 93.7/ 72.3 Prec 84/ 59 Recall 74/72 F1 score 78/65
		US (training I)	Acc 80.9/76.9 Prec 92/56 Recall 81/77 F1 score 86/ 65

We must also compare our image classification with other TL models, such as Nasnet and DenseNet, to ensure we use the most effective techniques.

An updated version of the BraNet application and prospectively explore the real AI/human interaction could be implemented, which can recognize full 2D images and not only resized images of 128×128 pixels. The app could be used for performance and load testing to assess how the application processes many images simultaneously. It simulates an increasing number of users or requests to see how the application performs under progressively higher loads.

We implement the app as a web server and realize scalability testing; incrementally increase resources (like CPU, GPU, and memory) available to the application and measure performance improvement to determine how efficiently the application scales; make full use of available CPU/GPU cores to process images in parallel, enhancing throughput; and utilize image compression techniques to reduce the size of high-resolution breast images without losing critical details necessary for analysis.

Finally, the use of IA in medical diagnosis brings about a range of ethical considerations that must be carefully navigated to ensure that the integration of these technologies

benefits patients, healthcare providers, and the broader healthcare system responsibly and equitably. It is essential to highlight ethical considerations regarding using artificial intelligence in developing CAD systems in healthcare.

The patient’s well-being is paramount, necessitating a comprehensive approach to protecting their data privacy and confidentiality [81, 82]. This project ensures patient privacy through the anonymization and coding of training image databases during the application’s first and second modules, which are also publicly available.

Another ethical consideration is the fairness of AI models [83], which requires providing equitable healthcare outcomes across various patient demographics. Thus, the developed application aims to contribute to medical service equity, particularly by facilitating pathology diagnosis in rural groups and sectors typically deprioritized in healthcare, especially in developing countries.

Finally, transparency regarding the capabilities and limitations of CAD systems is fundamental [84], ensuring that medical staff and patients know that decisions and outcomes adhere to ethical standards. In this context, the developed application is merely a test prototype that aspires to achieve the necessary maturity for use in a real healthcare setting, ensuring the requisite medical reliability.

6 Conclusions

In this paper, we have introduced BraNet, a mobile app for breast image classification based on deep learning algorithms. The API enables the rapid construction of breast image classification workflows, encompassing data input/output, ROI mask extraction, segmentation, and evaluation metrics. The client–server architecture, coupled with its open interface, empowers users to customize the pipeline and swiftly establish comprehensive medical image classification setups using Python libraries and the react native framework for creating native mobile applications on iOS and Android. We have demonstrated the functionality of the BraNet app by conducting automatic cross-validation on data augmentation, ROI segmentation, and classification using public ultrasound and mammography datasets, resulting in a preclinical tool. After implementing some improvements and future updates, BraNet will facilitate the migration of medical image segmentation and classification from research laboratories to practical applications. Also, ensuring that the App complies with all regulations and standards governing data privacy and security in healthcare is essential. It is only a preclinical testing phase; thus, there is still work to be done in this area. BraNet currently offers a pipeline for breast image segmentation and classification, and it will continue to receive regular updates and extensions in the future. This data must be rigorously analyzed, reported, and often published in scientific journals to ensure its accuracy and reliability.

Appendix 1

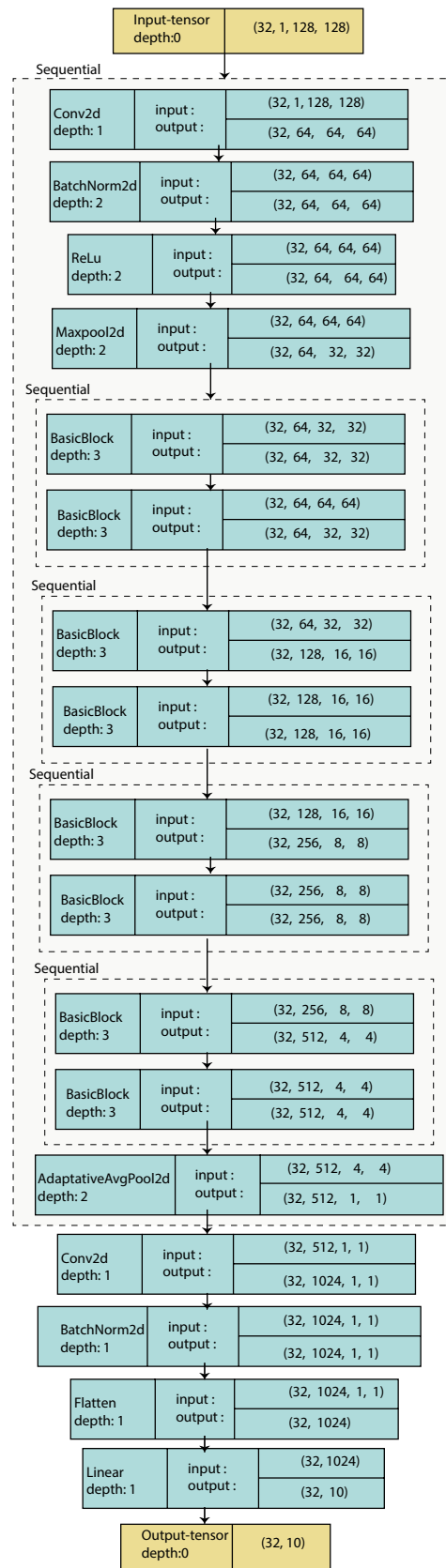
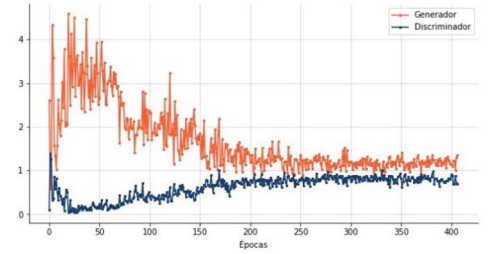
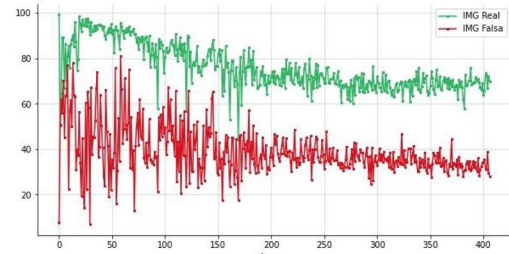
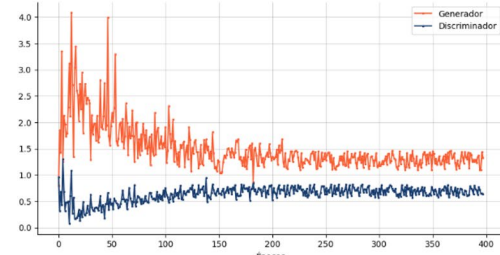
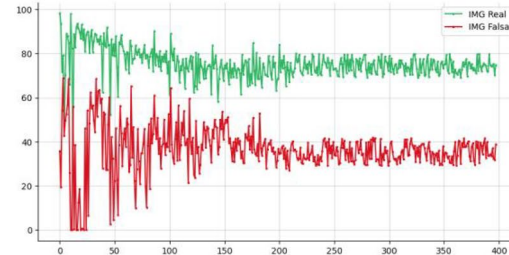
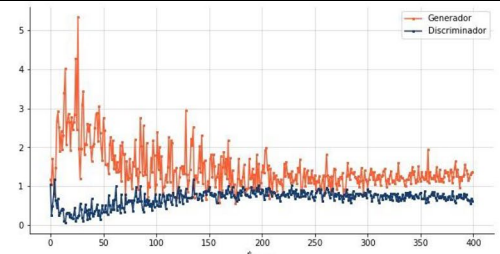
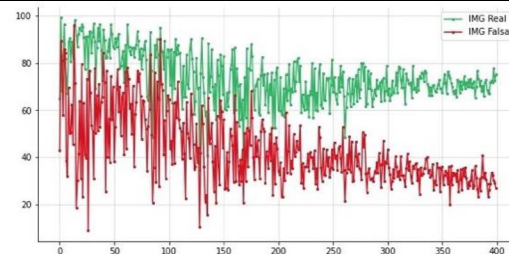


Fig. 5 ResNet 18 classification model

Appendix 2

Table 9 The GAN discriminator’s accuracy on real (green) and fake (red) images during training. Figures illustrate the loss and accuracy trends for SNGAN

GAN Model	Image Type	Generator and Discriminator GAN performance	
SN GAN	DM Training I		
	DM Training II		
	US		

Appendix 3

Training	Image Type	FID/KID
I	DM	
II	DM	
I	US	

Fig. 6 The FID and KID values a. SNGAN-FID values using mammography images during training I and II. b. SNGAN-FID values taken from US images

Appendix 4

Table 10 ResNet 18 evaluation model on training and validation datasets

Image Type	Databases	Nro images Training /validation	Datasets	Hyperparameters	Graphic Loss and accuracy training and validation	Graphic Metric vs Score
DM Taining I	Mini MIAS INBREAST CBIS-DDSM		Real + synthetic ROI images	num_epochs = 100 batch_size = 16 image_size = 128 lr = 0.0002 beta1 = 0.1 beta2 = 0.9		
US Taining I	BUSI Dataset B UDIAT		Real + synthetic ROI images	num_epochs = 100 batch_size = 16 image_size = 128 lr = 0.0002 beta1 = 0.1 beta2 = 0.9		
DM Taining II	VinDr-Mammo		Real ROI images	num_epochs = 100 batch_size = 16 image_size = 128 lr = 0.0002 beta1 = 0.1 beta2 = 0.9 dropout = 0.5		
			Real + synthetic ROI images	num_epochs = 100 batch_size = 32 image_size = 128 lr = 0.0001 beta1 = 0.1 beta2 = 0.999 dropout = 0.5		

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11517-024-03084-1>.

Author contributions Conceptualization, Y.J.G., M.J.R.-Á and V.L.; study conception and design: Y.J.G, M.J.R.-Á, D.C.M, VL; formal analysis, Y.J.G., M.J.R.-Á. and V.L.; investigation, Y.J.G.; Data Collection Y.J.G and S.G.J; analysis and interpretation of results: Y.J.G, S.G.J;D.C.F and P.C.D writing—original draft preparation, Y.J.G, M.J.R.A, D.C.M, VL writing—review and editing, Y.J.G, M.J.R.A, D.C.M, D.C.F, P.C.D and VL visualization, Y.J.G.; supervision, M.J.R.-Á. and V.L.; project administration, M.J.R.-Á.; funding acquisition M.J.R.-Á. All authors reviewed the results and approved the final version of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Funding was obtained from the Universidad Técnica Particular de Loja, PROY_INV_QU_2022_3576. CRUE-UNIVERSITAT POLITÈCNICA DE VALÈNCIA.

Data availability All codes are available as Mendeley Data: <https://data.mendeley.com/preview/jh9trvbjbv?a=57b040ca-ae6d-4ebb-bc04-ac8c27deae59> [85].

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

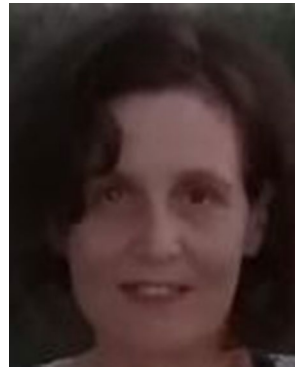
- Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, Walsh S (2017) Radiomics: the bridge between

- medical imaging and person-alized medicine. *Nat Rev Clin Oncol* 14(12):749–762
2. Papadimitroulas P, Brocki L, Chung NC, Marchadour W, Vermet F, Gaubert L, ..., Hatt M (2021) Artificial intelligence: deep learning in oncological radiomics and challenges of interpretability and data har-mo-ni-zation. *Phys Medica* 83:108–121
 3. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, ..., Sardanelli F (2021) AI applications to medical images: from machine learning to deep learning. *Phys Medica* 83:9–24
 4. Wollny G, Kellman P, Ledesma-Carbayo MJ, Skinner MM, Hublin JJ, Hierl T (2013) MIA-A free and open source software for gray scale medical image analysis. *Source Code Biol Med* 8(1):1–20
 5. Philbrick KA, Weston AD, Akkus Z, Kline TL, Korfiatis P, Sakinis T, ..., Erickson BJ (2019) RIL-contour: a medical imaging dataset annotation tool for and with deep learning. *J Digital Imaging* 32:571–581
 6. Gibson E, Li W, Sudre C, Fidon L, Shakir DI, Wang G, ..., Vercauteren T (2018) NiftyNet: a deep-learning platform for medical imaging. *Comput Methods Progr Biomed* 158:113–122
 7. Papademetris X, Jackowski MP, Rajeevan N, DiStasio M, Okuda H, Constable RT, Staib LH (2006) BioImage Suite: an integrated medical image analysis suite: an update. *Insight J* 2006:209
 8. Zhang L, Dai H, Sang Y (2022) Med-SRNet: GAN-based medical image super-resolution via high-resolution repre-sentation learning. *Comput Intell Neurosci* 20(2022):1744969. <https://doi.org/10.1155/2022/1744969>
 9. Zhu Y, Zhou Z, Liao G, Yuan K (2020) CsrGAN: medical image super-resolution using a generative adversarial network. 2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops), Iowa City, IA, USA, pp. 1–4. <https://doi.org/10.1109/ISBIWorkshops50223.2020.9153436>
 10. Zhang K, Hu H, Philbrick K, Conte GM, Sobek JD, Rouzrokh P, Erickson BJ (2022) SOUP-GAN: su-per-resolution MRI using generative adversarial networks. *Tomography* 8:905–919. <https://doi.org/10.3390/tomography8020073>
 11. Ahmad W, Ali H, Shah Z et al (2022) A new generative adversarial network for medical images super resolution. *Sci Rep* 12:9533. <https://doi.org/10.1038/s41598-022-13658-4>
 12. Bargsten L, Schlaefel A (2020) SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. *Int J CARS* 15:1427–1436. <https://doi.org/10.1007/s11548-020-02203-1>
 13. Haque A (2021) EC-GAN: low-sample classification using semi-supervised algorithms and GANs (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence* 35(18):15797–15798
 14. Sun Y, Yuan P, Sun Y (2020) MM-GAN: 3D MRI data augmentation for medical image segmentation via gen-erative adversarial networks. 2020 IEEE International Conference on Knowledge Graph (ICKG), Nanjing, China, pp. 227–234. <https://doi.org/10.1109/ICKG50248.2020.00041>
 15. Bashar MA, Nayak R (2020) TAnoGAN: time series anomaly detection with generative adversarial net-works. In: 2020 IEEE symposium series on computational intelligence (SSCI). Canberra, ACT, Australia, pp 1778–1785. <https://doi.org/10.1109/SSCI47803.2020.9308512>
 16. Müller D, Kramer F (2021) MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. *BMC Med Imaging* 21:12. <https://doi.org/10.1186/s12880-020-00543-7>
 17. Cardoso M, Clarkson M, Modat M, Ourselin S (2012) NiftySeg: open-source software for medical image segmentation, label fusion and cortical thickness estimation. In: *IEEE international symposium on biomedical imaging*, Barcelona, Spain
 18. Jiménez Gaona Y, Castillo Malla D, Vega Crespo B, Vicuña MJ, Neira VA, Dávila S, Verhoeven V (2022) Ra-diomics diagnostic tool based on deep learning for col-poscopy image classification. *Diagnostics* 12:1694. <https://doi.org/10.3390/diagnostics12071694>
 19. Chen Y, Zhang Q, Wu Y, Liu B, Wang M, Lin Y (2018) Fine-tuning ResNet for breast cancer classification from mammography. In *The International Conference on Healthcare Science and Engineering* (pp. 83–96). Springer, Singapore
 20. Dourado CMJM, da Silva SPP, da Nobrega RVM, Rebouças Filho PP, Muhammad K, de Albuquerque VHC (2021) An open IoHT-based deep learning framework for online medical image recognition. *IEEE J Select Areas Commun* 39(2):541–548. <https://doi.org/10.1109/JSAC.2020.3020598>
 21. Lee H, Chen YPP (2015) Image based computer aided diagnosis system for cancer detection. *Expert Syst Appl* 42(12):5356–5365
 22. Chowdhury D, Das A, Dey A, Sarkar S, Dwivedi AD, Rao Mukkamala R, Murmu L (2022) ABCanDroid: a cloud integrated android app for noninvasive early breast cancer detection using transfer learning. *Sensors* 22(3):832
 23. Jiménez-Gaona Y, Rodríguez-Álvarez MJ, Lakshminarayanan V (2020) Deep-learning-based computer-aided systems for breast cancer imaging: a critical review. *Appl Sci* 10(22):8298
 24. Guo Z, Xie J, Wan Y, Zhang M, Qiao L, Yu J, ..., Yao Y (2022) A review of the current state of the computer-aided diagnosis (CAD) systems for breast cancer diagnosis. *Open Life Sci* 17(1):1600–1611
 25. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788
 26. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2961–2969
 27. Bharati P, Pramanik A (2020) Deep learning techniques—R-CNN to mask R-CNN: a survey. *Computational intelligence in pattern recognition: proceedings of CIPR 2019*, pp 657–668
 28. Shin HC, Lu L, Summers RM (2017) Natural language processing for large-scale medical image analysis using deep learning. *Deep learning for medical image analysis*, pp 405–421. <https://doi.org/10.1016/B978-0-12-810408-8.00023-7>
 29. Qi X, Yi F, Zhang L, Chen Y, Pi Y, Chen Y, ..., Yi Z (2022) Computer-aided diagnosis of breast cancer in ultrasonography images by deep learning. *Neurocomputing* 472:152–165
 30. Linna N, Kahn CE Jr (2022) Applications of natural language processing in radiology: a systematic review. *Int J Med Inform* 163:104779
 31. Yin C, Qian B, Wei J et al (2019) Automatic generation of medical imaging diagnostic report with hierarchical recurrent neural network. In: Wang J, Shim K, Wu X (eds) *Proceedings IEEE international conference on data mining ICDM*. Vol 2019-November. Institute of electrical and electronics engineers inc, pp 728–737. <https://doi.org/10.1109/ICDM.2019.00083>
 32. Wang J, Deng H, Liu B, Hu A, Liang J, Fan et al (2020) Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: bibliometric study on PubMed. *J Med Int Res* 22(1):e16816
 33. Luo JW, Chong JJ (2020) Review of natural language processing in radiology. *Neuroimaging Clinics* 30(4):447–458
 34. Prabadevi B, Deepa N, Krithika LB, Vinod V (2020) Analysis of machine learning algorithms on cancer dataset. In: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE), Vellore, India, pp 1–10. *Intelligence algorithms*. *Cancers* 14(14):3442. <https://doi.org/10.1109/ic-ETITE47903.2020.36>
 35. Buckley JM, Coopey SB, Sharko J, Polubriaginof F, Drohan et al (2012) The feasibility of using natural language processing to extract clinical information from breast pathology reports. *J Pathol Inform* 3(1):23
 36. Chen Y, Hao L, Zou VZ, Hollander Z, Ng RT, Isaac KV (2022) Automated medical chart review for breast cancer outcomes research: a novel natural language processing extraction system. *BMC Med Res Methodol* 22(1):136
 37. Liu Y, Liu Q, Han C, Zhang X, Wang X (2019) The implementation of natural language processing to extract index lesions from

- breast magnetic resonance imaging reports. *BMC Med Inform Decis Mak* 19(1):1–10
38. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, ..., Girshick R (2023) Segment anything. In: proceedings of the IEEE/CVF international conference on computer vision, pp 4015–4026
 39. Keshari R, Singh R, Vatsa M (2020) Generalized zero-shot learning via over-complete distribution. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 13300–13308
 40. Wu N, Phang J, Park J, Shen Y, Huang Z, Zorin M, ..., Geras KJ (2019) Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Trans Med Imaging* 39(4):1184–1194
 41. Huang Y, Han L, Dou H, Luo H, Yuan Z, Liu Q, ..., Yin G (2019) Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. *Biomed Eng Online* 18(1):1–18
 42. Weng Y, Zhou T, Li Y, Qiu X (2019) Nas-unet: neural architecture search for medical image seg-men-tation. *IEEE Access* 7:44247–44257
 43. Al-Dhabyani W, Gomaa M, Khaled H, Aly F (2019) Deep learning ap-proaches for data augmentation and classification of breast masses using ultrasound images. *Int J Adv Comput Sci Appl* 10(5):1–11
 44. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015–Conference Track Proceedings. 2015, 1–14
 45. Mahmoud HAH, Alharbi AH, Khafga DS (2021) Breast cancer classification using deep convolution neural network with transfer learning. *Intell Autom Soft Comput* 29(3):803–814. <https://doi.org/10.32604/iasc.2021.0186>
 46. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
 47. Abhisheka B, Biswas SK, Purkayastha B (2023) A comprehensive review on breast cancer detection, classification and segmentation using deep learning. *Arch Comput Methods Eng* 30(8):5023–5052
 48. Li Y, Wu J, Wu Q (2019) Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. *Ieee Access* 7:21400–21408
 49. Murtaza G, Shuib L, Abdul Wahab AW, Mujtaba G, Mujtaba G, Nweke HF, ..., Azmi NA (2020) Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artif Intell Rev* 53:1655–1720
 50. Muramatsu C, Nishio M, Goto T, Oiwa M, Morita T, Yakami M, ..., Fujita H (2020) Improving breast mass classification by shared data with domain transformation using a generative adversarial network. *Comput Biol Med* 119:103698
 51. Khan S, Islam N, Jan Z, Din IU, Rodrigues JJC (2019) A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recogn Lett* 125:1–6
 52. Jiménez-Gaona Y, Rodríguez-Álvarez MJ, Carrión-Figueroa D, Castillo-Malla D, Lakshminarayanan V (2024) Breast mass regions classification from mammograms using convolutional neural networks and transfer learning. *J Modern Optics TMOP*. <https://doi.org/10.1080/09500340.2024.2313724>
 53. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B (2020) Radiomics in medical imaging “how-to” guide and critical reflection. *Insights Imaging* 11(1):1–16
 54. Rodrigues PS (2017) Breast ultrasound image. *Mendeley Data, V1*. <https://doi.org/10.17632/wmy84gzngw.1>
 55. Yap MH, Pons G, Marti J et al (2018) Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Heal Informatics* 22(4):1218–1226. <https://doi.org/10.1109/JBHI.2017.2731873>
 56. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K et al (1998) Current status of the digital database for screening mammography. *En Digital mammography*. Springer, Dordrecht
 57. Suckling J, Parker J, Dance D et al (2015) Mammographic image analysis society (MIAS) database v1.21. [Dataset]. Apollo - University of Cambridge Repository. <https://www.repository.cam.ac.uk/handle/1810/250394>
 58. Li S, Hatanaka Y, Fujita H, Hara T, Endo T (1999) Automated detection of mammographic masses in MIAS Database. *Med Imaging Technol* 17:427–428
 59. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) INbreast. *Acad Radiol* 19(2):236–248. <https://doi.org/10.1016/j.acra.2011.09.014>
 60. Pham HH, Nguyen Trung H, Nguyen HQ (2022) VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography (version 1.0.0). *PhysioNet*. 10.13026/br2v-7517
 61. Ibromhimov B, Kang JY (2022) Two-stage deep learning method for breast cancer detection using high-resolution mammogram images. *Appl Sci* 12(9):4616
 62. Shao S, Wang P, Yan R (2019) Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput Ind* 106:85–93
 63. Woldehellasse H, Tesfamariam S (2023) Data augmentation using conditional generative adversarial network (cGAN): application for prediction of corrosion pit depth and testing using neural network. *J Pipeline Sci Eng* 3(1):100091
 64. Parmar G, Zhang R, Zhu J-Y (2022) On aliased resizing and surprising subtleties in GAN evaluation. 11400–11410. <https://doi.org/10.1109/cvpr52688.2022.01112>
 65. Gao M, Song P, Wang F, Liu J, Mandelis A, Qi D (2021) A novel deep convolutional neural network based on ResNet-18 and transfer learning for detection of wood knot defects. *Journal of Sensors* 2021:1–16
 66. Aljuaid H, Alturki N, Alsubaie N, Cavallaro L, Liotta A (2022) Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Comput Methods Programs Biomed* 223:106951
 67. Swarnambiga Ayyachamy, Varghese Alex, Mahendra Khened, and Ganapathy Krishnamurthi (2019) “Medical image retrieval using Resnet-18”. *Proc. SPIE 10954, Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, 1095410. <https://doi.org/10.1117/12.2515588>
 68. Guo M, Du Y (2019) “Classification of thyroid ultrasound standard plane images using ResNet-18 networks,” 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, China, pp. 324–328. <https://doi.org/10.1109/ICASID.2019.8925267>
 69. Borji A (2019) Pros and cons of GAN evaluation measures. *Comput Vis Image Underst* 179:41–65. <https://doi.org/10.1016/j.cviu.2018.10.009>
 70. Zama S, Fujioka T, Yamaga E, Kubota K, Mori M, Katsuta L, ..., Tateishi U (2023) Clinical utility of breast ultrasound images synthesized by a generative adversarial network. *Medicina*, 60(1):14
 71. Pang T, Wong JHD, Ng WL, Chan CS (2021) Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification. *Comput Methods Programs Biomed* 203:106018
 72. Jiménez Gaona Y, Castillo Malla D, Vega Crespo B, Vicuña MJ, Neira VA, Dávila S, Verhoeven V (2022) Radiomics diagnostic tool based on deep learning for colposcopy image classification. *Diagnostics* 12:1694. <https://doi.org/10.3390/diagnostics12071694>
 73. Dihge L, Bendahl PO, Skarpping I, Hjartström M, Ohlsson M, Rydén L (2023) The implementation of NILS: a web-based artificial neural network decision support tool for noninvasive lymph node staging in breast cancer. *Front Oncol* 13:1102254
 74. To T, Lu T, Jorns JM, Patton M, Schmidt TG, Yen T, ..., Ye DH (2023) Deep learning classification of deep ultraviolet fluorescence images toward intra-operative margin assessment in breast cancer. *Front Oncol* 13:1179025

75. Taylor CR, Monga N, Johnson C, Hawley JR, Patel M (2023) Artificial intelligence applications in breast imaging: current status and future directions. *Diagnostics* 13(12):2041
76. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ (2019) Artificial intelligence in breast imaging. *Clin Radiol* 74(5):357–366
77. Huynh HN, Tran AT, Tran TN (2023) Region-of-interest optimization for deep-learning-based breast cancer detection in mammograms. *Appl Sci* 13(12):6894
78. Afrin H, Larson NB, Fatemi M, Alizad A (2023) Deep learning in different ultrasound methods for breast cancer, from diagnosis to prognosis: current trends, challenges, and an analysis. *Cancers* 15(12):3139
79. Prodan M, Paraschiv E, Stanciu A (2023) Applying deep learning methods for mammography analysis and breast cancer detection. *Appl Sci* 13(7):4272
80. Oyelade ON, Ezugwu AE, Almutairi MS, Saha AK, Abualigah L, Chiroma H (2022) A generative adversarial network for synthesis of regions of interest based on digital mammograms. *Sci Rep* 12(1):6166
81. Herington J, McCradden MD, Creel K, Boellaard R, Jones EC, Jha AK, ..., Saboury B (2023) Ethical considerations for artificial intelligence in medical imaging: data collection, development, and evaluation. *Journal of Nuclear Medicine*, 64(12), 1848–1854
82. Boellaard R, Jones EC, Jha AK, ..., Saboury B (2023) Ethical considerations for artificial intelligence in medical imaging: deployment and governance. *J Nuclear Med* 64(10):1509–1515
83. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, ..., Naganawa S (2024) Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol* 42(1):3–15
84. Drabiak K, Kyzer S, Nemov V, El Naqa I (2023) AI and machine learning ethics, law, diversity, and global impact. *Br J Radiol* 96:20220934
85. Jimenez Y, Rodriguez-Alvarez MJ, Castillo-Malla D, Garcia S, Carrión-Figueroa D, Lakshminarayanan V (2024) BraNet: a mobil application for breast image classification based on deep learning algorithms. Mendeley Data, V1. <https://doi.org/10.17632/jh9trvbjbv>

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



María José Rodríguez Álvarez she received her PhD in Physics from the University of Valencia, Spain. She is currently a professor in the computer science degree at the Universitat Politècnica de Valencia in Spain. Her research interests include medical image reconstruction, image processing, and deep learning. Since 2009, she is a member of the Physics Medicine Group at the i3M (Instituto de Instrumentación para la Imagen Molecular).



Darwin Castillo-Malla he received his Master's degree in Biomedical Engineering, Universidad Politécnica de Madrid, Spain. Bachelor in Electronics and Telecommunications Engineering, Universidad Técnica Particular de Loja, Ecuador. He is currently attending an PhD program in Mathematics at the Universitat Politècnica de València, Spain. Associate Professor at Universidad Técnica Particular de Loja, Ecuador. SPIE and IEE EMBS Chapter member.



Santiago García-Jaén he received his BSc degree in Computer Science Engineering at the Universidad Técnica Particular de Loja (UTPL) at Ecuador in 2023. His areas of interest are Biomedical, Implementation of neural networks through Deep Learning algorithms and development applied to web and mobile technologies.



Yuliana Jiménez-Gaona she received her MSc degree in Bioinformatics from University of Bologna at Italy, and her BSc degree in Computer Science Engineering from Universidad Técnica Particular de Loja (UTPL) in Ecuador. She is currently attending a PhD program in Health Technologies at Universitat Politècnica de València. Associate Professor at Universidad Técnica Particular de Loja, Ecuador. Her areas of interest are computer vision, biomedical, bioinformatic, Math and educational innovation projects. She is a member of SPIE.



Diana Carrión Figueroa she received her degree in medicine from the Universidad Técnica Particular de Loja (UTPL) in Ecuador. She is Obstetrics and Gynecology specialist in mastology at the Pontifical Catholic University of Ecuador. She received a university master's degree in Oncological Gynecology and Multidisciplinary Breast Pathology at Universitat of Barcelona in Spain. She provides diagnosis and treatment in benign and malignant breast pathology at Hospital Carlos Andrade Marín – IESS, in Quito-Ecuador.

tional innovation projects. She is a member of SPIE.



Patricio Corral-Domínguez medical Oncologist Surgeon. Master's Degree in Breast Pathology and Senology. President of CIPAM (CUENCA COMPREHENSIVE BREAST PATHOLOGY CENTER – ECUADOR). Member of the Ecuadorian Society of Oncology Past President of the Ecuadorian Society of Oncology chapter of Azuay Member of SEPAM (Ecuadorian Society of Breast Pathology). Member of the SESPAM (Spanish Society of Senology and Breast Pathology). Member of ESMO (European

Society for Medical Oncology). Member of the Latin American Thyroid Society (LATS). Member of the Erobreast Network Physician of the Surgical Staff of the Monte Sinai Hospital in Cuenca – Ecuador Professor at the Faculty of Medicine of the University of Cuenca of the Postgraduate Gynecology with the chair of Breast Pathology and Senology.



Vasudevan Lakshminarayanan he received his Ph.D. from the University of California at Berkeley and is currently a professor of optometry and vision science, physics, ECE and systems design engineering at the university of Waterloo. His research ranges from optical science and engineering, applied math, biomedical engineering, neuroscience, cognitive science, clinical ophthalmology and optometry, and history of science. He is a member of SPIE.