



Digital mammography dataset for breast cancer diagnosis research (DMID) with breast mass segmentation analysis

Parita Oza¹ · Urvi Oza² · Rajiv Oza³ · Paawan Sharma⁴ · Samir Patel⁴ · Pankaj Kumar¹ · Bakul Gohel²

Received: 24 February 2023 / Revised: 8 November 2023 / Accepted: 28 November 2023 / Published online: 21 December 2023
© Korean Society of Medical and Biological Engineering 2023

Abstract

Purpose: In the last two decades, computer-aided detection and diagnosis (CAD) systems have been created to help radiologists discover and diagnose lesions observed on breast imaging tests. These systems can serve as a second opinion tool for the radiologist. However, developing algorithms for identifying and diagnosing breast lesions relies heavily on mammographic datasets. Many existing databases do not consider all the needs necessary for research and study, such as mammographic masks, radiology reports, breast composition, etc. This paper aims to introduce and describe a new mammographic database.

Methods: The proposed dataset comprises mammograms with several lesions, such as masses, calcifications, architectural distortions, and asymmetries. In addition, a radiologist report is provided, describing the details of the breast, such as breast density, description of abnormality present, condition of the skin, nipple and pectoral muscles, etc., for each mammogram.

Results: We present results of commonly used segmentation framework trained on our proposed dataset. We used information regarding the class of abnormalities (benign or malignant) and breast tissue density provided with each mammogram to analyze the segmentation model's performance concerning these parameters. **Conclusion:** The presented dataset provides diverse mammogram images to develop and train models for breast cancer diagnosis applications.

Keywords Mammogram · Dataset · Deep learning · Breast mass segmentation

1 Introduction

Breast cancer is one of the most frequent illnesses among women worldwide. Screening examinations or the start of clinical signs are the most common ways to identify this disease [1]. It is the most common disease, affecting 2.1 million people yearly and accounting for most cancer fatalities. About 500,000 individuals each year die from breast cancer, which accounts for 15% of cancer mortality [2]. Early identification and diagnosis of breast cancer are critical for lowering cancer-related death rates. As a result, the medical community advises regular screening for breast

tumours. Digital mammography is the most effective imaging method to discover and diagnose breast cancer in its earliest stages. For each breast, mammography records two views: the craniocaudal (CC) view (top-to-bottom view) and the mediolateral oblique (MLO) view (side view) [3]. The most common mammogram findings are calcifications, masses or lesions, asymmetries, and architectural distortion. Identifying worrisome findings by the expert radiologist is time-consuming and exhausting, with a 10–30% prevalence of undiscovered lesions [4]. Writing medical reports with descriptive information about symptoms and interpretation of findings is a critical task radiologists perform. Deep learning (DL) based computer-aided detection and diagnosis (CAD) technologies have been developed in the last two decades to assist radiologists in interpreting mammogram images and help reduce this rate [4]. DL is an improvement of artificial neural networks, consisting of more layers that permit higher levels of abstraction and improved predictions from data. In particular, Deep convolutional neural networks (DCNNs) are powerful tools for various computer vision tasks. They learn the mass's appearance by utilizing thousands of images throughout training. Scientific research

✉ Parita Oza
parita.prajapati@nirmauni.ac.in

¹ Nirma University, Ahmedabad, India

² Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India

³ Rad Imaging, X-Ray and Sonography Clinic, Ahmedabad, India

⁴ Pandit Deendayal Energy University, Gandhinagar, India

shows that the training set's size significantly impacts the effectiveness of supervised DL algorithms [5–7]. Moreover, these data must be accurately annotated by specialized radiologists. The most prevalent issue in medical imaging is the need for large, balanced, and well-annotated datasets since manual annotation of medical images is time-consuming and tedious [8, 9]. Hence, the contribution of a new dataset is always essential and valuable for the research community of this domain. Additionally, it enables researchers to evaluate their models using diverse datasets. A large number of CAD systems are developed for different breast imaging modalities. Still, mammogram-based CAD systems are well suited for diagnosing the disease as mammography is considered an international gold standard for the early detection of breast cancer [10–12]. In this paper, we present a new mammogram dataset. We give a detailed description of our proposed mammogram dataset in the corresponding sections of the paper.

1.1 Motivation and research contribution

Researchers need datasets to develop, test, and evaluate mammogram-based CAD systems for breast cancer diagnosis. Most mammography datasets are private to the organization, and few are publicly available. These mammogram datasets are also crucial for comparing the outcomes of different investigations. Additionally, these datasets may also be used to teach and train students in the field of medicine [4]. This paper aims to introduce a new mammographic research database and overcome some limitations of existing databases. We provide mammography images and other essential details such as pixel level annotation, ROI mask for each abnormal mammogram, BI-RADS, and breast composition as per ACR-BIRADS. Additionally, mammograms with radiology reports provided in this dataset can be used to explore and design DL architectures to generate diagnostic reports of the mammogram automatically. To the best of our knowledge, this would be the first dataset in India supported by a radiology report and other essential details needed to train and test the DL models for breast cancer diagnosis. We also analyze the breast mass segmentation model on our proposed model. The goal of designing an AI-based breast cancer diagnosis model is to detect and segment the mass/abnormality boundaries and classify them as benign or malignant in an end-to-end framework. As breast mass is one of the most distinctive symptoms of cancer, segmentation of breast mass to define its shape and location is crucial for diagnosis. Researchers propose multiple methods for breast mass segmentation, including patch-based [13, 14] and whole mammogram segmentation [15, 16]. Here we have presented segmentation results obtained on our proposed dataset by training an end-to-end breast

mass segmentation model. We have analyzed mammogram segmentation results concerning varying tissue density and mass classes.

1.2 Paper organization

The rest of the paper is organized as follows: Sect. 2 discusses some existing mammogram repositories. We give detailed descriptions of our proposed dataset in Sect. 3. We present some experiments and result-analysis of the segmentation model on a proposed dataset in Sect. 4. We finally end with the discussion and conclusion in Sects. 5 and 6, respectively.

2 Some existing mammogram repositories

Researchers in the field of breast cancer utilize a variety of mammogram databases, some of which are open to the public and others limited to specific organizations. This section briefly discusses a few commonly cited mammogram datasets.

- *The Mammographic Image Analysis Society Digital Mammogram Database (MIAS)*: MIAS [17] is an ancient breast cancer dataset that is still widely utilized. This dataset is a collection of 322 images; all are reduced in resolution (1024×1024). The dataset comprises all possible findings, such as malignant and benign abnormalities, along with normal mammograms. The dataset has less number of instances of benign and malignant findings as compared to normal.
- *INbreast*: INbreast [4] is a mammography collection that includes both screening and diagnostic mammograms. Images from both MLO and CC views are included in this collection. There are 410 images in the collection, which were created from 115 patient cases.
- *The Digital Database for Screening Mammography (DDSM)*: Another most used and old mammogram dataset is DDSM [18]. Standard views for the left and right breasts are provided in each instance, including the MLO and the CC view. In addition, the dataset has images with various findings, such as normal, benign, and malignant masses.
- *SureMaPP*: SureMapp [19] is another mammogram dataset with total of 343 images. The dataset includes a separate file with ground truth and other information. The images in the datasets have different resolutions.
- *King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD)*: Another recently published dataset is KAU-BCMD [20], the first in Saudi Arabia to include a substantial number of mammography

images. There are 1416 instances in all. The dataset also contains 205 ultrasound cases that match a portion of the mammography cases.

- *Other:* IRMA [21], BCDR [22], BancoWeb LAPIMO [23] are other publicly available mammogram datasets but not frequently used in the literature. IRMA is an amalgamation of various other mammogram datasets. BCDR is made up of two datasets, BCDR-FM and BCDR-DM. BCDR-DM is currently in the early stages of development. Table 1 presents a comparative analysis of our dataset with the various other mammography datasets utilized in the literature.

Limitations of Existing Mammogram Datasets [24–26]: The MIAS dataset is small and unbalanced; there are fewer positive (abnormal) samples than normal ones. This dataset comprises an aberrant mammogram area's centre point and radius, which need further processing for ground truth. Mammograms from DDSM are compressed using the nonstandard JPEG format. Moreover, some images need to be adequately annotated, necessitating the intervention of a skilled radiologist to provide the proper ground truth. Similar to MIAS, the SureMaPP dataset also offers image coordinates for the centre of the abnormality and an approximation of the radius of a circle containing the anomaly, necessitating additional processing for developing a ground truth. INbreast dataset is also smaller in size, and hence additional methods are needed to increase the number of image samples, such as data augmentation. The INbreast dataset and our proposed dataset also share some characteristics, such as pixel-level annotation, masks, breast density and radiology report. Even though Inbreast offers medical reports, they are in Portuguese, so additional language conversion is required. Inbreast dataset offers segmented masks, but they are in XML format, so additional processing is required to utilize them. For ease of implementation, we provide all images (mammograms and related masks) in our dataset in a simple image format. Additionally, compared to the INBreast dataset, our dataset has more images in total (510) and more abnormal images (300 plus). Annotations in the dataset IRMA are illustrated with several indicators such as arrows, text, and other symbols; this can affect the textural feature extraction. There are also not many additional datasets in the literature that include segmented masks for abnormal mammograms and radiology reports that include other crucial information like BIRADS, ACR-BIRADS, and other pathological details. All of the aforementioned information is included in our proposed dataset and will contribute to the ongoing developments in DL-based mammography systems in the future. The ground-truth annotation using pixel level boundary and corresponding segmented mask

Table 1 Summary of Mammogram Datasets

Dataset	MIAS	DDSM	Inbreast	SureMaPP	KAU-BCMD	BCDR	BancoWeb LAPIMO	IRMA	Truetra	DMID
Origin	UK	USA	Portugal	Saudi Arabia	Brazil	Germany	Spain	India		
Year	1994	1999	2011	2020	2012	2010	2008	2008	2023	
Total images	322	10,480	410	343	5662	7315	1473	10,509	320	510
Image type	PGM	LJPEG	DICOM, XML	DICOM	DICOM	TIFF	TIFF	Several	DICOM	DICOM, TIFF
Multiview	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES
BI-RADS	NO	YES	YES	NO	YES	—	YES	YES	YES	YES
Lesion Type	All kind	Mass, Calcification	All kind	All kind	All kind	—	All kind	All kind	All kind	All kind
Radiologist Report	NO	NO	YES (Portuguese)	NO	NO	NO	NO	NO	NO	YES (English)
Pixel level annotation	NO	YES	YES	NO	NO	YES	NO	NO	NO	YES
Supporting Modality	NO	NO	NO	Yes (US)	NO	NO	NO	NO	NO	NO
Public	YES	YES	YES	YES	YES	YES	YES	NO	NO	YES

and radiology reports are the most distinguishing features of this dataset. We summarize the main features of our proposed dataset as follows:

- Radiologist report with various pathology
- ACR breast density categories
- Pixel-level annotation and corresponding ROI masks
- Breast Imaging Reporting and Data System categories
- Identical distribution of normal and abnormal images
- Segmented mask corresponding to each abnormal mammogram

3 Dataset description

The proposed mammography dataset was collected from Samved Hospital, Ahmedabad, India. The imaging technology MAMMOMAT 3000 Nova by Siemens was utilized for screening. The device produces high-resolution images. All the images have variable resolutions, but common

resolutions are around 4743×6000 and 3520×4784 . Further processing is performed on the dataset to eliminate floating background artefacts. Images of the dataset contain findings such as masses, calcification, architectural distortion, asymmetries, and images with multiple findings. Figure 1 shows mammograms from our collection that include these abnormalities. We also show some examples of normal mammograms and annotated mammograms with multiple findings in Fig. 3 (first row).

The dataset includes normal, benign, and malignant cases. In addition, BI-RADS [27, 28] categories and other pathological details are also provided as a separate medical report. The American College of Radiology established a well-defined risk assessment and quality assurance technique called BI-RADS (Breast Imaging-Reporting and Data System). In BI-RADS, descriptors like shape and margin are used for the diagnosis. Breast imaging studies are classified into one of seven BI-RADS evaluation categories, as indicated in Table 2.

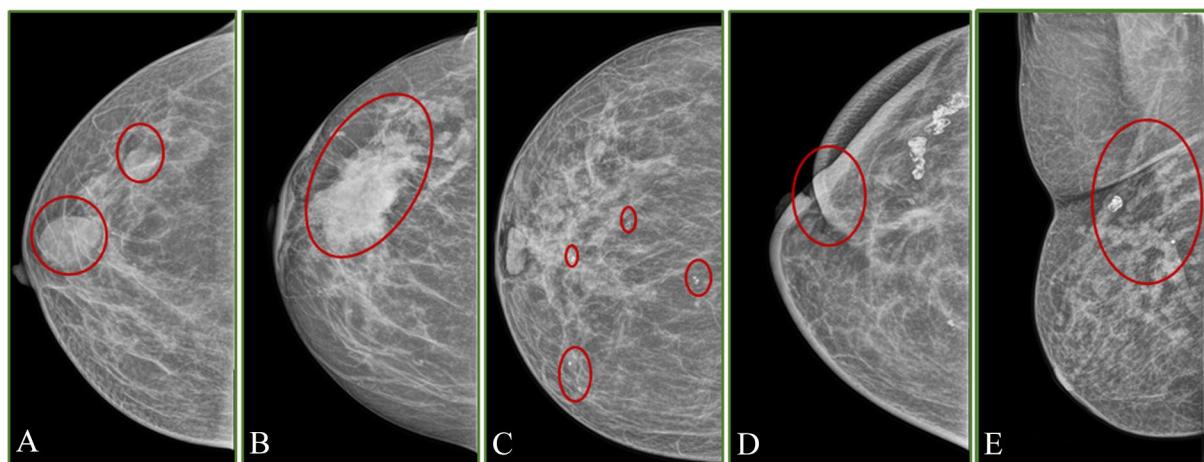


Fig. 1 Breast abnormalities. **A** Mass-Benign, **B** Mass-Malignant, **C** Calcifications, **D** Architectural distortion, **E** Asymmetry

Table 2 BI-RADS Assessment Categories [27, 28]

Category	Assessment	Descriptors	Follow-up
0	Incomplete assessment, Need additional imaging evaluation	–	Need further assistance
1	Normal	–	Suggested annual screening mammography (if age >40)
2	Benign finding (Non-cancerous lesion)	-	Suggested annual screening mammography (if age >40)
3	Probably benign	Circumscribed mass/Obscured mass	6 months follow-up mammogram
4	Suspicious abnormality	Microlubulated mass	May requires biopsy
5	High probability of malignancy	Indistinct and Spiculated mass	Requires biopsy
6	Proven malignancy	-	Biopsy-proven malignancy, Suggested to check the extent and presence in the opposite breast

We provide TIFF as well as DICOM format for images in the dataset. All DICOM images are converted into TIFF format using the "Master View" application. A separate CSV file is provided with various details in eight separate columns, such as reference number, Laterality (CCRT, CCLT, MLORT, MLOLT), Character of background tissue (F: Fatty, G: Fatty-glandular, D: Dense-glandular), Types of abnormality (Calcification, Well-defined / circumscribed masses, Spiculated masses, Ill-defined masses, architectural distortion, a and Normal), Class of abnormality (benign, malignant, no Defined), (x,y) image-coordinates of the centre of abnormality and approximate radius (in pixels) of a circle enclosing the abnormality. Figure 2 presents a sample of metadata that is stored in a CSV file. We briefly describe our dataset in Table 3.

3.1 Validation and ground truth by radiologists

Three expert radiologists have contributed to the annotation and validation of the images. Out of which one radiologist has provided pixel-level annotation on abnormal mass region and the other two have validated and confirmed the same. The images were segmented by manual annotation on suspicious areas of mammograms. The colour green is used to highlight benign abnormalities, while red colour is used

to show malignant abnormalities. These pixel-level annotations are then used in Python script to generate segmented masks. For each abnormal mammogram, a corresponding ROI mask(s) is/are provided (see Fig. 3 third row). With the feature of the app "Efilm," (x y) the centre of abnormality and approximate radius in the pixel are exported into a CSV file.

3.2 Radiology report

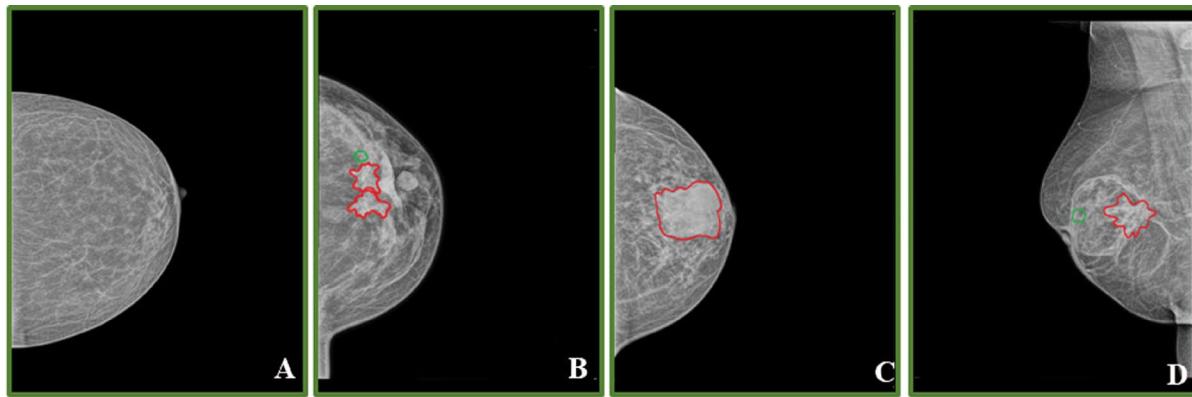
Each mammography image in the dataset is provided with a medical report. The report contains the findings from mammograms, breast composition, BI-RADS scale, presence of microcalcification, skin thickening, nipple, pectoral muscle, etc. In addition, the report presents breast composition in four categories as per ACR BI-RADS; A, B, C, and D. Breasts can be entirely fatty (ACR-A), with scattered areas of dense fibro glandular breast tissue (ACR-B), with multiple areas of glandular and connective tissue (ACR-C), or can be dense in the extreme (ACR-D). Figure 3 (second row) presents the images pertaining to each of these cases from the dataset. In addition, the report provides a BIRADS score for each mammogram image, as shown in Table 2. Based on the imaging test, each number corresponds to a classification

Image reference number	Breast view	Character of background tissue	Type of abnormality present	Class of abnormality	X,Y image-coordinates of centre of abnormality		Approximate radius of a circle enclosing the abnormality
IMG069	CCLT	G	CIRC	B	1516	2512	113
IMG070	CCRT	G	SPIC	M	2538	2072	550
IMG071	MLORT	G	CIRC	B	1666	3541	101
IMG072	MLORT	G	SPIC	M	3314	3021	529

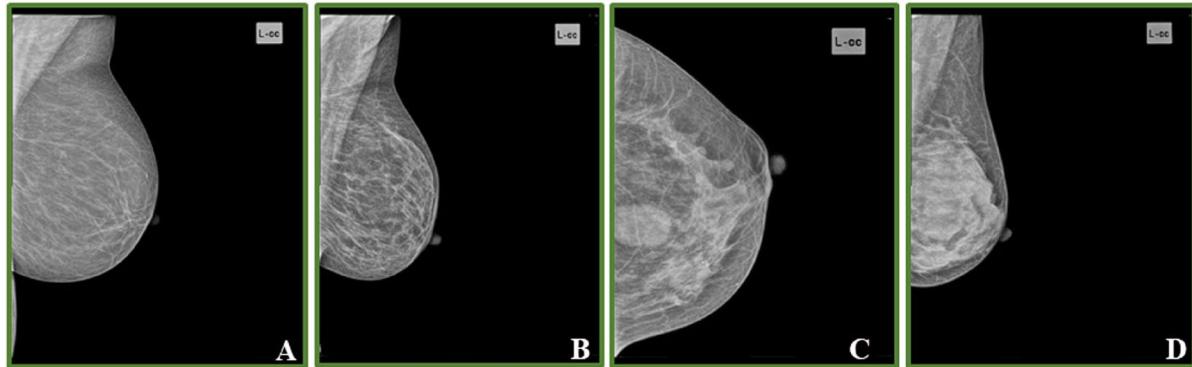
Fig. 2 A sample of the metadata, stored in CSV file

Table 3 Description of dataset

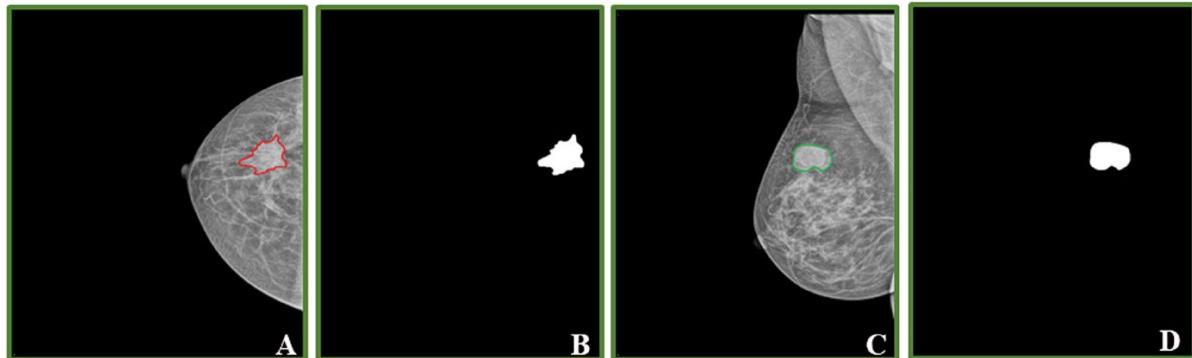
Imaging modality	Digital mammography
Image type	DICOM / TIFF
Data acquisition device	MAMMOMAT 300 Nova, Siemens
Image resolution	Variable but most common are around 4743×6000 and 3520×4784
Breast View	MLO, CC
Assessment	BI-RADS
Annotation	Pixel level boundaries on abnormalities, Supporting file with parameters such as x and y coordinates of abnormality and radius
Additional feature	Patient medical report with other pathological indicators and BI-RADS, ROI masks
Source	Samved Hospital, Ahmedabad, India
Target application	Machine learning / Deep learning based Breast cancer detection and classifications
More specific subject matter	Breast cancer diagnosis using BI-RADS assessment, Automatic radiology report generation, Breast Composition assessment



Examples of normal mammograms and mammograms with multiple findings: (A) Normal (B) Spiculated mass, circumscribed mass, Architectural distortion (C) ill-defined mass and calcification (D) Architectural distortion and califications



The four breast density categories are shown in this image. (A) ACR-A (B) ACR-B (C) ACR-C (D) ACR-D



Mammograms & corresponding mask: (A) Malignant Mammogram (B) Malignant Mask (C) Benign Mammogram (D) Benign Mask

Fig. 3 First row: Examples of mammograms with various findings. Second row: Examples of mammograms with ACR categories. Third row: Abnormal mammograms and corresponding mask

that evaluates your breast cancer risk. For example, in our dataset, BIRADS-4 (defined as a suspicious lesion) is further classified into BIRADS 4a (low level of suspicion for malignancy), BIRADS 4b (moderate level of suspicion for malignancy), and BIRADS 4c (high level of suspicion for malignancy). Figure 4 shows a sample radiologist report and a corresponding mammogram.

3.3 Statistics of proposed dataset

In this section, we present various statistics of our datasets concerning BI-RADS categories, class distribution, number of images per finding, and breast composition. Breast density assessment is one of the most important tasks in identifying people at risk for breast cancer. Each image in the collection is assigned one of the four density categories (decided by the Breast Imaging and Reporting Data Systems) by the experts. Figure 5A shows breast

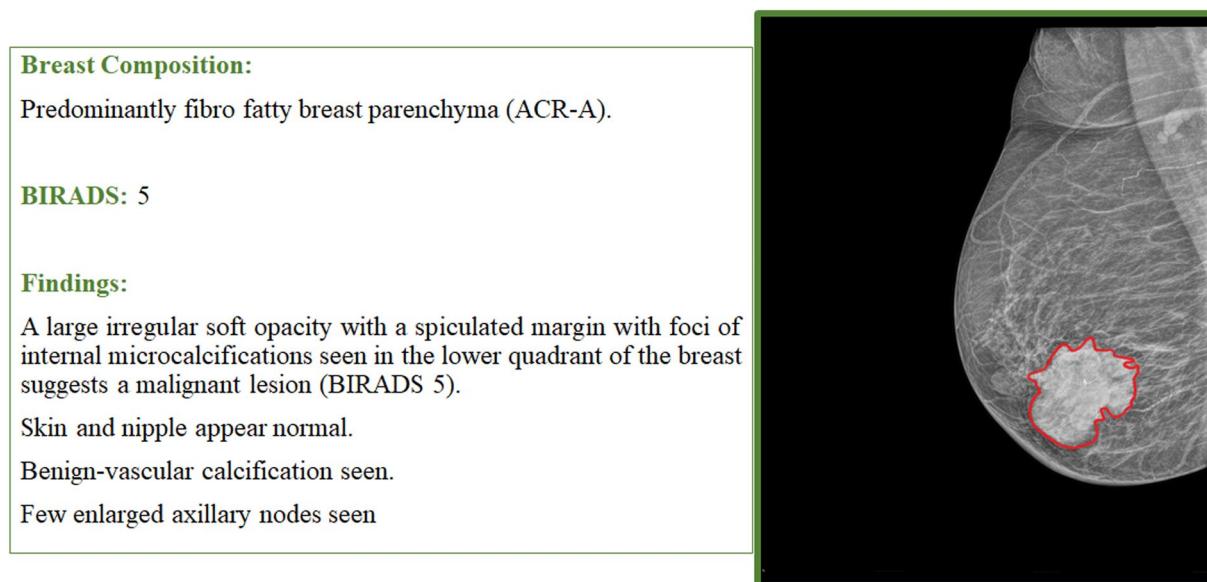


Fig. 4 A mammogram with corresponding radiologist report

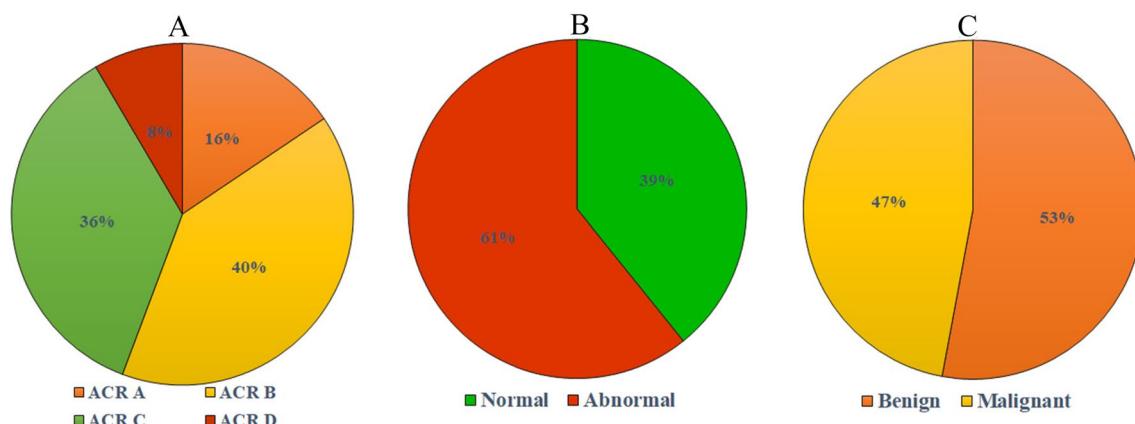


Fig. 5 A Breast composition as per ACR categories. B Normal/Abnormal class distribution. C Benign/Malignant class distribution

composition categorization of all the dataset images. The graph shows that ACR-B is assigned to the highest number of images (around 40%), followed by ACR-C (around 36%). On the other hand, only 8% of images are assigned ACR-D rating, which is highly dense. This is obvious because breast cancer is more commonly diagnosed in older women (age > 40).

Figure 5 also depicts an overall class distribution. There are around 39% normal class images and approximately 61% aberrant class (abnormal class) images in the dataset (see Fig. 5B). In the data collection, there are a handful of images with both anomalies (benign and malignant). Such images, according to doctors, should be classified as cancerous only. Around 36 abnormal images are tagged as "not defined" out of 310 abnormal images. "Not defined" are those mammogram images that need further examination to confirm the

diagnosis. So from the remaining 274 images, finally, we have around 47% malignant class images and 53% benign class images in our data collection, which is a nearly equal class distribution (see Fig. 5C).

The dataset collection includes images with various findings, including benign/malignant masses, calcifications, asymmetries, architectural deformation, and multiple findings. Figure 6 (left) depicts how circumscribed is prominent in our database. This reflects the real world, where the most common mammographic findings are most likely benign. BIRADS 1 (41%) accounts for the majority of our images. Approximately 23% of cases are classified as BIRADS 3, which is most likely benign. BIRADS 2 and 5 accounted for about 5% of the total number of images. Only one image in the dataset is classified as BIRADS 0, indicating that the evaluation is incomplete; hence, further assistance is

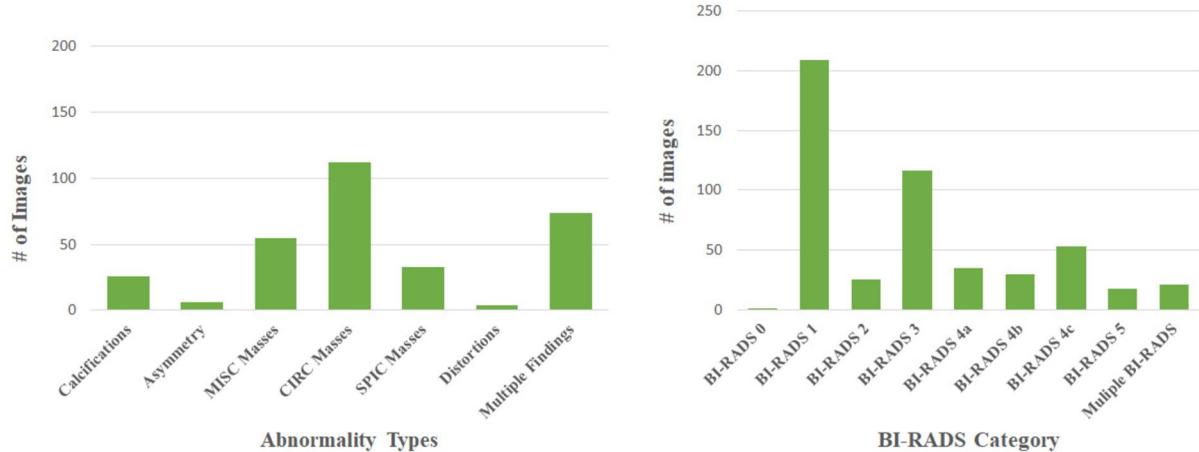


Fig. 6 Left: Findings in the dataset. Right: BI-RADS categories in the dataset

required. Around 23% of scans have a BI-RADS 4 rating, indicating a possible abnormality. Figure 6 (right) presents an overall BI-RADS classification of the dataset.

We also have a nearly equal number of benign and malignant classes. There are about 274 abnormal images in the collection out of 510 total, accounting for over 58%. For every abnormal image of the dataset, we also provide mask images for each abnormal mammogram (see Fig. 3 third row).

3.4 Dataset availability

The dataset can be found at https://figshare.com/authors/Parita_Oza/17353984.

The data is organized into five folders. All DCM files can be found in the "DICOM Images" folder while corresponding tiff files can be found in the "TIFF Images" folder. The folder "Reports" contains the radiologist's reports. We provide pixel-level annotation for each anomalous image; these images can be accessed in the folder "Pixel-level annotation." We also produced segmented masks for each anomalous image, which are found in the folder "ROI Masks." Finally, a CSV file called "Metadata" contains other important details of the complete data set.

4 Segmentation results on proposed dataset

This section outlines the key steps taken in data preprocessing, model training, and evaluation of segmentation models trained on our dataset.

4.1 Preparing data

We used a self-crafted preprocessing method to eliminate artefacts.

4.1.1 Preprocessing method

- Removing border artifacts - We removed border artifacts/bright white borders/corners by cropping out boundary pixels.
- Removing background artifacts - Detecting the largest contour in the mammogram (i.e., breast region) and cropping the rectangle box enclosing it to remove extra background artifacts and pixels.
- Flipping operation - As the dataset has left and right breast mammograms having different breast directions, We apply the flipping operation to orient all the mammograms in the left order.
- Padding - Once all the mammograms have the same orientation, we pad the mammograms with background pixels. Most pretrained DL models take a square image of the same width and height as input. Thus, we are padding the mammogram to have the same width and height. Another reason for padding is that when we apply augmentation like rotation or translation, it sometimes crops some part of the mammogram. Thus, while padding the image, we add pixels on the left side of the breast to have the breast region in the center so that any geometric transformation doesn't cut the breast region.
- Normalization - We applied min-max normalization on mammograms to have all the pixel values range between 0–1.
- Augmentation - We augmented the image with rotation, translation, zooming, and shearing operations.

All the preprocessed images and masks are resized to 512*512 to train the model.

4.1.2 Dataset distribution

We used a five-fold cross-validation method to train and test the segmentation models on our dataset. We divided the entire dataset into five parts, each with an almost equivalent number of normal mammograms and mammograms with benign, malignant, and both kinds of masses. Moreover, we ensured that each fold would have mammograms of all types of tissue density, i.e., Fatty (F), Glandular fatty (G), and Dense (D). While training, four parts of the dataset were used as training, and the remaining one part was used as a test dataset. Thus, we trained five segmentation models with different train and test datasets. From the training dataset, 30% of mammograms are chosen randomly as validation dataset. Table 4 shows the number of mammograms belonging to each class of mammogram and each type of tissue density in the training and testing data of every fold. We trained and analyzed two kinds of segmentation models. The first one was an end-to-end model trained and tested on both normal and abnormal mammograms available in the dataset (model A). Following most of the studies available in the literature, the second model was trained with only abnormal mammograms (model B).

4.2 Training segmentation model

We trained commonly used segmentation frameworks on our proposed dataset to analyze their performance on our proposed dataset. There are two types of commonly used segmentation frameworks available in the literature—(1) Whole mammogram segmentation models [15, 16, 29] and (2) Integrated mass detection and mass patch segmentation frameworks [30–32].

Whole mammogram segmentation models are DL algorithms designed to perform segmentation on a whole mammogram image, with a primary focus on identifying breast masses or abnormalities. The encoder-decoder architectures like Unet models are the most popular whole mammogram segmentation models used in the literature. The Unet architecture was designed for biomedical image

Table 4 Train-test split of images in each fold of segmentation model training

	Training	Testing
Benign	114/115	28/29
Malignant	84	21
Benign and Malignant	16	4
Normal	160	40
F - Fatty	63/64	12/13
G - Fatty glandular	388	77
D - Densoglandular	15/16	4/5

segmentation tasks [33]. The traditional Unet model has been modified by integrating attention gates and dense connections or adding attention-guided dense-up sampling blocks to get more effective whole mammogram segmentation results. In this paper, we used the Unet and Attention-Unet model [34], which was proposed to learn to focus on target structures of varying shapes and sizes for medical image segmentation tasks. The architecture of the Attention-Unet model is presented in Fig. 7. It can be seen from the figure that it is a modified version of the traditional Unet model. The input image is progressively filtered and downsampled by a factor of 2 at each scale in the encoding part of the network. Attention gates (AGs) filter the features propagated through the skip connections. These features are then upsampled in the decoding part of the network to predict the binary segmentation mask as an output. A Schematic of the AGs is shown in the bottom part of Fig. 7. We trained both Unet and Attention Unet models with dice loss as a loss function. An Adam optimizer with a learning rate of $1e^{-4}$ was used to train the models.

The Integrated detection and segmentation frameworks perform segmentation by combining two essential tasks: identifying potential mass regions using detection models and precisely outlining their boundaries using patch segmentation models. We trained the YOLOv8 detection model to detect the breast mass. Based on the location of the detected mass, a mammogram patch of size 1024*1024 is extracted, centering the detected mass. We trained a simple Unet segmentation model to segment the detected mammogram mass patches. We evaluate the performance of these methods on our proposed dataset to determine which one is better suited to the dataset's characteristics.

4.3 Results

Following are the details of evaluation metrics used to assess the performance of our trained models.

- Dice Similarity Coefficient (DSC)/F1 score - The DSC is the most widely used evaluation metric for semantic segmentation. It ranges between 0 and 1. The formula for DSC is given by:

$$DSC = \frac{2 * |GT \cap PR|}{|GT| + |PR|}$$

where GT is the ground truth image, and PR is the predicted segmentation output. $|GT|$ and $|PR|$ is the number of pixels in the ground truth and predicted output, respectively, and $|GT \cap PR|$ is the number of positive pixels predicted correctly.

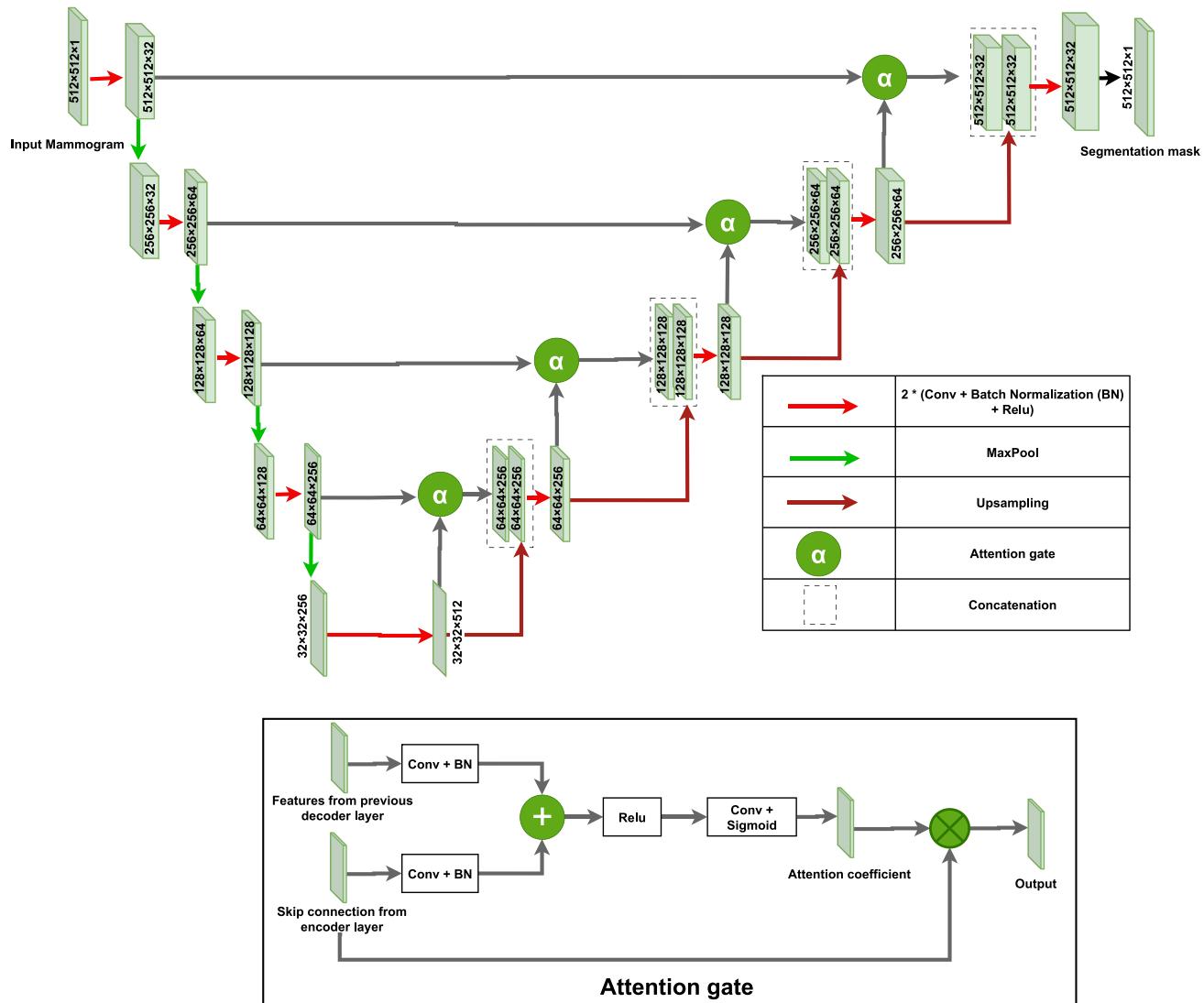


Fig. 7 DL model architecture used as Segmentation Model

- Recall - It is the ratio of the sum of correctly identified mass pixels (True Positives - TP) and actual mass pixels. Here, actual mass pixels are represented by the sum of correctly identified mass pixels and missed mass pixels. The TPR score represents the proportion of mass pixels in the ground truth that the model correctly segmented.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- Precision - It is the ratio of the sum of correctly predicted mass pixels to the total mass pixels. It is a measure of the quality of the predictions/segmentation by the model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Table 5 Results analysis of existing segmentation frameworks over proposed dataset

Model	Mean DSC	Mean recall	Mean precision
Unet	0.6076	0.6578	0.8009
Attention unet	0.6400	0.5533	0.7858
YOLOv8 based detection and segmentation	0.6096	0.6807	0.6243

The performance of different segmentation frameworks on our proposed dataset is shown in Table 5. For our dataset, Attention Unet seems to have comparatively better results than the other two models. Thus, we used the Attention Unet model to perform further analysis of segmentation

Table 6 Results analysis of segmentation model over proposed dataset

Model	Mean DSC	Mean recall	Mean precision
Model trained on Normal and abnormal cases	0.5580	0.5269	0.6635
Model trained on Only abnormal cases	0.6400	0.5533	0.7858

performance when the model is trained with both normal and abnormal mammograms.

The results of both Attention Unet segmentation models (model A and model B) on the test dataset are shown in Table 6. It can be seen that the model trained with only abnormal mammograms achieved better DSC, recall, and precision scores than the model trained with both normal and abnormal mammograms. However, the recall score of model B is only slightly better than the model A. To get a detailed insight into how accurately both models segment breast masses, we calculated and compared mean DSC, precision, and recall scores for abnormal mammograms only. Segmentation results of mammograms with benign, malignant, and when both kinds of masses are present are shown in Table 7. It can be seen that the class of masses also affects the segmentation performance. Results show that both models better segment malignant masses than benign masses. Table 8 shows how the breast's tissue density affects the masses' segmentation. We divided the test mammograms into one of the three categories of breast tissue density - 'F', 'G', and 'D', and calculated mean DSC, precision, and recall scores for each category. It can be seen from the table that mammogram masses with dense breast tissue are more brutal to segment for both models. Both models have the

highest segmentation results on mammograms with fatty tissue density. Model A shows similar performance segmenting masses in breast tissue density of types G and D. On the contrary, model B had low performance while segmenting dense tissue breasts. Figure 8 show visual results for varying abnormality classes and breast densities from both models.

5 Discussion

Our proposed dataset provides both normal and mammograms with mass abnormalities and its segmentation masks to train an end-to-end breast segmentation model. We present the results of commonly used segmentation frameworks such as Unet, Attention Unet, and YOLOv8-based mass patch detection and segmentation framework on our proposed dataset. Attention mechanisms in the Unet model help the model selectively focus on the regions of interest (mass regions), allowing the network to highlight mass boundaries and characteristics, leading to more accurate segmentation.

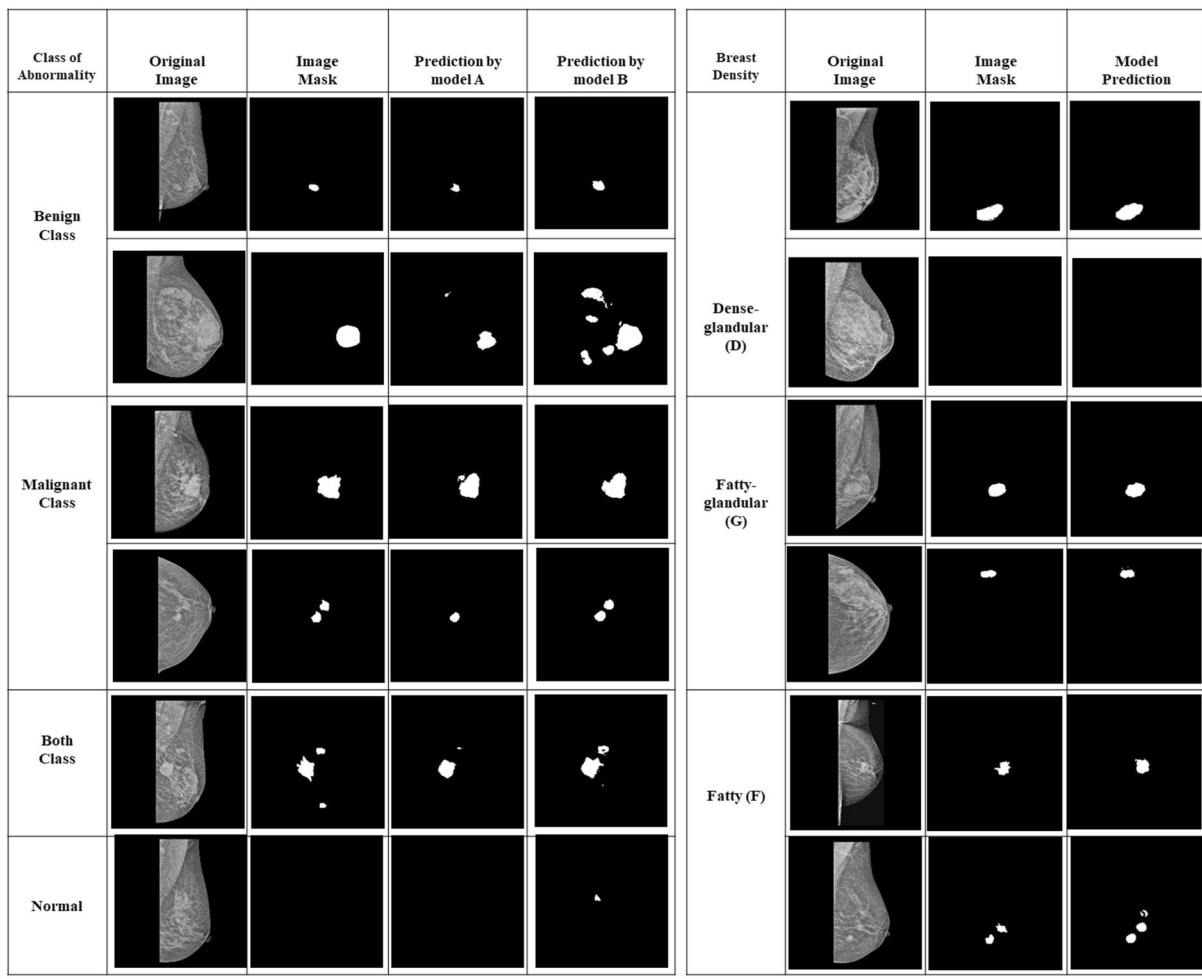
Utilizing normal and mass mammograms provided by our dataset, we analyze Attention Unet's performance further. Our results suggest that training segmentation models with only abnormal mammograms performed better than training with both normal and abnormal mammograms. However, it predicts more FPs when trained with only abnormal mammograms. Figure 8 shows the results of both models when tested on normal mammograms. Provided normal mammograms can be used to analyze the FPs while training and testing the end-to-end segmentation models. We have examined the segmentation results of benign and malignant masses by our trained models. The tissue characteristics of the benign masses are sometimes similar to normal breast tissue, which could be one of the reasons for lower dice scores than malignant

Table 7 Results analysis of proposed end-to-end model segmentation model over mammograms with different breast tissue characteristics

Background tissue	Model trained on Normal and abnormal cases			Model trained on Only abnormal cases		
	Mean DSC	Mean recall	Mean precision	Mean DSC	Mean recall	Mean precision
Fatty	0.6539	0.7462	0.9487	0.8187	0.7477	0.8956
Fatty glandular	0.4940	0.4834	0.5984	0.5751	0.5222	0.6782
Dense glandular	0.3299	0.4250	0.2715	0.5551	0.5177	0.3906

Table 8 Analysis of segmentation model's results for different types of mammogram masses

Present Breast mass classes	Model trained on Normal and abnormal cases			Model trained on Only abnormal cases		
	Mean DSC	Mean recall	Mean precision	Mean DSC	Mean recall	Mean precision
Benign	0.5148	0.4413	0.6347	0.5259	0.4835	0.6170
Malignant	0.6620	0.5593	0.8959	0.6851	0.5677	0.8940
Both	0.6247	0.5173	0.8380	0.6644	0.5548	0.8515

**Fig. 8** Model predictions

masses. Our further analysis includes examining the effect of breast density over breast mass segmentation. Dense breasts have more fibroglandular breast tissues reflecting as bright areas in mammograms resembling a mass. Thus, breasts with dense tissues have a low contrast between tissue and mass, leading to lower dice scores than mammograms with fatty or fatty glandular tissue. Model B is explicitly trained on abnormal mammograms, so it predicts more pixels as mass for dense mammograms than model A.

We surveyed and compared scores of commonly used Unet models on popular public datasets—Inbreast and BCDR. We compared our scores with the cross-validation scores achieved on the INbreast and BCDR datasets. As provided in Table 9, scores achieved on our dataset are comparable to the INbreast dataset when tested on the same ground.

Table 9 Comparision with other benchmark dataset

Method	Dataset	Dice scores
Unet [7]	INbreast	0.62
Fusion Net [8]	INbreast	0.62
FCDenseNet103[9]	INbreast	0.42
AUnet [10]	INbreast	0.64
Unet++Xception [11]	BCDR	0.58
U-Xception [11]	BCDR	0.58
U-ResNet50 [11]	BCDR	0.48
Attention Unet [12]	DMID (proposed dataset)	0.64
Unet [7]	DMID (proposed dataset)	0.60

We can experiment with more advanced segmentation models to analyze and improve DSC, precision, and recall scores. Breast density labels and mass class information from our dataset can be used to design a segmentation model, considering breast tissue and mass characteristics.

6 Conclusion

In the field of mammographic imaging, we have developed a new dataset for the training and assessment of ML and DL networks. Our proposed data collection includes valuable mammography scans with ground truth provided by subject-matter experts. This dataset can be utilized for various tasks in the breast cancer research domain. The ground-truth annotation and radiology reports are the most distinguishing features of this work. Additionally, we have nearly equal numbers of benign and malignant classes, resulting in balanced datasets. We trained and analyzed the segmentation model using breast mass masks provided for each mammogram. Our motive was not just to train the segmentation models but to present a method to examine if the segmentation model can learn the different characteristics of breast mass, such as density, shape, etc. In order to build end-to-end CAD tools for breast cancer diagnosis, DL architectures may be explored and designed using the radiology reports supplied in this dataset.

Acknowledgements We would like to thank Samved Hospital, Ahmedabad, India. Additionally, we would like to thank Dr. Dinesh Patel and Dr. Trupti Patel for their support.

Funding Funding.

Data availability The dataset can be used to train ML or DL models for mammogram classification, BI-RADS classification, as well as breast composition classification. Moreover, it can be used to train segmentation models to segment breast lesions. Additionally, the radiology reports can be utilized to train report generation models. This dataset will be available for research purposes only on the link provided in the article.

Declarations

Conflict of interest None

Ethical approval The use of the dataset for the purpose of AI research has been approved by the Ethical Review Board.

Consent to participate The hospital has provided anonymized images for the dataset.

Consent to publish The authors affirm to publish this dataset with permission.

References

1. American Cancer Society, Breast cancer facts & figures 2019–2020. Am Cancer Soc, pp 1–44 (2019)
2. Breast cancer landscape in asia-pacific. https://novotech-cro.com/sites/default/files/2021-02/Breast20Cancer20Landscape20in/20Asia-Pacific_2021.pdf. Accessed 2022-03-10 (2021).
3. Fenton JJ, Zhu W, Balch S, Smith-Bindman R, Fishman P, Hubbard RA. Distinguishing screening from diagnostic mammograms using Medicare claims data. *Med Care.* 2014;52(7):244. <https://doi.org/10.1097/MLR.0b013e318269e0f5>.
4. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. Inbreast: toward a full-field digital mammographic database. *Acad Radiol.* 2012;19(2):236–48. <https://doi.org/10.1016/j.acra.2011.09.014>.
5. Oza P, Sharma P, Patel S. A transfer representation learning approach for breast cancer diagnosis from mammograms using efficientnet models. *Scalable Comput Practice Exp.* 2022;23(2):51–8. <https://doi.org/10.12694/scpe.v23i2.1975>.
6. Oza P, Sharma P, Patel S. Transfer learning assisted classification of artefacts removed and contrast improved digital mammograms. *Scalable Comput Practice Exp.* 2022;23(3):115–27. <https://doi.org/10.12694/scpe.v23i2.1975>.
7. Oza P, Sharma P, Patel S. A drive through computer-aided diagnosis of breast cancer: a comprehensive study of clinical and technical aspects. In: Recent innovations in computing: proceedings of ICRIC 2021, Vol 1, pp 233–249 (2022c). 10.1007/978-981-16-8248-3_19
8. Oza P, Sharma P, Patel S. Breast lesion classification from mammograms using deep neural network and test-time augmentation. *Neural Comput Appl.* 2023. <https://doi.org/10.1007/s00521-023-09165-w>.
9. Oza P. AI in breast imaging: Applications, challenges, and future research. In: Computational intelligence and modelling techniques for disease detection in mammogram images. 2023.
10. Oza P, Sharma P, Patel S, Kumar P. Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey. *Neural Comput Appl.* 2022;34:1815–36. <https://doi.org/10.1007/s00521-021-06804-y>.
11. Oza P, Sharma P, Patel S, Adedoyin F, Bruno A. Image augmentation techniques for mammogram analysis. *J Imaging.* 2022;8(5):141. <https://doi.org/10.3390/jimaging8050141>.
12. Oza P, Sharma P, Patel S. Deep ensemble transfer learning-based framework for mammographic image classification. *J Supercomput.* 2022. <https://doi.org/10.1007/s11227-022-04992-5>.
13. Li H, Chen D, Nailon WH, Davies ME, Laurenson DI. Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography. *IEEE Trans Med Imaging.* 2021;41(1):3–13. <https://doi.org/10.1109/TMI.2021.3102622>.
14. Baccouche A, Garcia-Zapirain B, Olea CC, Elmaghriby AS. Connected-unets: a deep learning architecture for breast mass segmentation. *NPJ Breast Cancer.* 2021;7(1):1–12. <https://doi.org/10.1038/s41523-021-00358-x>.
15. Abdelhafiz D, Bi J, Ammar R, Yang C, Nabavi S. Convolutional neural network for automated mass segmentation in mammography. *BMC Bioinform.* 2020;21(1):1–19. <https://doi.org/10.1186/s12859-020-3521-y>.
16. Sun H, Li C, Liu B, Liu Z, Wang M, Zheng H, Feng DD, Wang S. Aunet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. *Phys Med Biol.* 2020;65(5):055005. <https://doi.org/10.1088/1361-6560/ab5745>.
17. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I, Stamatakis E, Cerneaz N, Kok S, et al. Mammographic image analysis society (mias) database v1. 21. (2015)

18. Michael Heath, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, Munishkumaran S. Current status of the digital database for screening mammography. In Digital mammography, pp 457–460. Springer (1998). https://doi.org/10.1007/978-94-011-5318-8_75
19. Bruno A, Ardizzone E, Vitabile S, Midiri M. A novel solution based on scale invariant feature transform descriptors and deep learning for the detection of suspicious regions in mammogram images. *J Med Signals Sens.* 2020;10(3):158. https://doi.org/10.4103/jmss.JMSS_31_19.
20. Alsolami AS, Shalash W, Alsaggaf W, Ashoor S, Refaat H, Elmogy M. King abdulaziz university breast cancer mammogram dataset (kau-bcmd). *Data.* 2021;6(11):111. <https://doi.org/10.3390/data6110111>.
21. Oliveira JEE et al. Toward a standard reference database for computer-aided mammography. In: Medical imaging 2008: computer-aided diagnosis, vol 6915, pp 606–614. SPIE (2008). <https://doi.org/10.1117/12.770325>.
22. Lopez MG, Posada N, Moura DC, Pollán RR, Valiente JMF, Ortega CS, Solar M, Diaz-Herrero M, Ramos IMAP, Loureiro J, et al. Bcdr: a breast cancer digital repository. In 15th international conference on experimental mechanics, vol 1215 (2012)
23. Matheus BRN, Schiabel H. Online mammographic images database for development and comparison of cad schemes. *J Digit Imaging.* 2011;24(3):500–6. <https://doi.org/10.1007/s10278-010-9297-2>.
24. Oza P, Sharma P, Patel S, Kumar P. Computer-aided breast cancer diagnosis: comparative analysis of breast imaging modalities and mammogram repositories. *Current Med Imaging.* 2022;18:1–13. <https://doi.org/10.2174/1573405618666220621123156>.
25. Tariq M, Iqbal S, Ayesha H, Abbas I, Ahmad KT, Niazi MFK. Medical image based breast cancer diagnosis: state of the art and future directions. *Expert Syst Appl.* 2021;167:114095. <https://doi.org/10.1016/j.eswa.2020.114095>.
26. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data.* 2017;4(1):1–9. <https://doi.org/10.1038/sdata.2017.177>.
27. D'Orsi CJ. The American college of radiology mammography lexicon: an initial attempt to standardize terminology. *AJR Am J Roentgenol.* 1996;166(4):779–80. <https://doi.org/10.2214/ajr.166.4.8610548>.
28. Weerakkody Y, Niknejad M. Breast imaging-reporting and data system (bi-rads). <https://radiopaedia.org/articles/10003>(2022). Accessed: 10 May 2022
29. Li S, Dong M, Guangming D, Xiaomin M. Attention dense-u-net for automatic breast mass segmentation in digital mammogram. *IEEE Access.* 2019;7:59037–47. <https://doi.org/10.1109/ACCESS.2019.2914873>.
30. Al-Antari MA, Al-Masni MA, Choi M-T, Han S-M, Kim T-S. A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *Int J Med Inform.* 2018;117:44–54. <https://doi.org/10.1016/j.ijmedinf.2018.06.003>.
31. Baccouche A, Garcia-Zapirain B, Castillo Olea C, Elmaghraby AS. Connected-unets: a deep learning architecture for breast mass segmentation. *NPJ Breast Cancer.* 2021;7(1):1–12. <https://doi.org/10.1038/s41523-021-00358-x>.
32. Dhungel N, Carneiro G, Bradley AP. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Anal.* 2017;37:114–28. <https://doi.org/10.1016/j.media.2017.01.009>.
33. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, pp 234–241. Springer (2015) . <https://doi.org/10.48550/arXiv.1505.04597>.
34. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, et al. Attention u-net: learning where to look for the pancreas. [arXiv:1804.03999](https://arxiv.org/abs/1804.03999), <https://doi.org/10.48550/arXiv.1804.03999> (2018)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.