Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Breast Abnormality Classification in Resource-Constrained Settings Through GAN Generated Synthetic Mammography Patches

InChan Hwang

inchan.hwang@montreat.edu

Montreat College https://orcid.org/0000-0001-9859-7498

Research Article

Keywords: Generative Adversarial Networks, Synthetic medical images, Deep learning, Data scarcity

Posted Date: July 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10293673/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Breast Abnormality Classification in Resource-Constrained Settings Through GAN Generated Synthetic Mammography Patches

InChan Hwang
Department of Cybersecurity
Montreat College
Montreat, NC, USA
inchan.hwang@montreat.edu

Beatrice Brown-Mulry
HITI Lab
Emory University
Atlanta, GA, USA
bbrownmulry@gmail.com

Hari Trivedi
HITI Lab
Emory University
Atlanta, GA, USA
hari.trivedi@emory.edu

Judy Wawira Gichoya
HITI Lab
Emory University
Atlanta, GA, USA
judywawira@emory.ed

Laleh Seyyed-Kalantari
Vector Institute for AI
Toronto, ON, Canada
lsk@yorku.ca

Sanghoon Lee
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
slee297@kennesaw.edu

MinJae Woo
Department of Health Sciences
Clemson University
Clemson, SC, USA
mwoo@clemson.edu

***Abstract*— Developing deep learning (DL) models for medical imaging is often hindered by limited resources because it requires substantial computing resources and sufficient training data. We propose the use of low-resource-consuming Generative Adversarial Network (GAN) to generate synthetic medical image patches to train mammography patch classifier to classify abnormal and normal breast mammography patches. Contrastive Unpaired Translation GAN (CUT-GAN) translated normal mammography patches to abnormal ones. Preprocessing pipelines were suggested to apply CUT-GAN in the generation of synthetic mammography patches. Thus, a synthetic training dataset can be obtained to train a DL model for the binary classification between real normal and abnormal image patches, protecting patients' privacy due to the use of synthetic image patches. Model performance metrics for the synthetic image trained Residual Neural Network (ResNet) 152 showed 0.751 (0.742 - 0.759) in terms of AUC. In contrast, the real image trained ResNet152 demonstrated 0.913 for the AUC score. In conclusion, synthetic mammography patches generated using CUT-GAN demonstrate the potential to serve as substitutes for real mammography patches in training DL models to accurately classify abnormalities in clinical breast images.**

Keywords— Generative Adversarial Networks; Synthetic medical images; Deep learning; Data scarcity;

I. INTRODUCTION

The research on the application of DL models in medical imaging has seen a significant increase [1], [2], [3], and those DL models have demonstrated promising results in classifying malignant and normal cases of patient images after detecting diseases [4]. One of the most active areas of DL application is the diagnosis of breast cancer [5], [6]. Breast cancer is a common disease for women, approximately 12.5% developing it during their lifetimes [7], [8]. Examining breast mammography images of patients is a cost-effective step in the diagnosis of breast cancer. Early detection through mammography raises the chances of early-stage cancer detection, allowing the design of effective treatment plans for patients [9]. However, mammography examination tasks assigned to radiologists are demanding regardless of their work experiences, so their diagnostic outcomes vary significantly [10]. Thus, DL models have been applied to breast cancer diagnosis to reduce diagnostic errors [11], [12].

A critical issue in training DL models is class imbalance, and it occurs when the number of cases in one class is significantly lower than the other cases [13]. Breast cancer is diagnosed with approximately 5 in 1000 screening exams (0.5%) [14]. As a result, breast cancer datasets exhibit imbalanced distributions between normal and abnormal cases. Training DL models on such datasets creates high rates of false positives and false negatives due to the training from inadequate data [15]. A common method to address class imbalance is data augmentation, collecting more samples for an underrepresented class [16].

Recent research has explored the use of GANs to tackle class imbalance for medical image data. They are trained to learn the distribution of class samples and generate synthetic images for underrepresented classes [17]. Image-to-image translation GAN models, such as CUT-GAN, address this issue by transforming normal image patches into abnormal ones. CUT-GAN is notable for low GPU memory consumption, and high-quality synthetic images [18].

Another significant issue associated with medical datasets, regardless of domains, is the protection of patients' identifiable information (PII) [19]. X-ray images, including mammography, sometimes contain identifiable patient information, hindering the release of large public datasets [20]. Consequently, the use of GANs has been explored to enhance privacy protection by replacing real data with synthetic data [21].

This study proposes a pipeline to generate synthetic breast mammography patches using CUT-GAN and assesses DL models trained on synthetic data while validating them on real mammography patches. The proposed pipeline selects optimal source mammography patches to train CUT-GAN model and applies post-processing techniques to match synthetic and real mammography image distributions. This research addresses two key challenges in medical AI: class imbalance and privacy protection in a breast mammography dataset. By demonstrating that DL models trained on synthetic patches achieve competitive

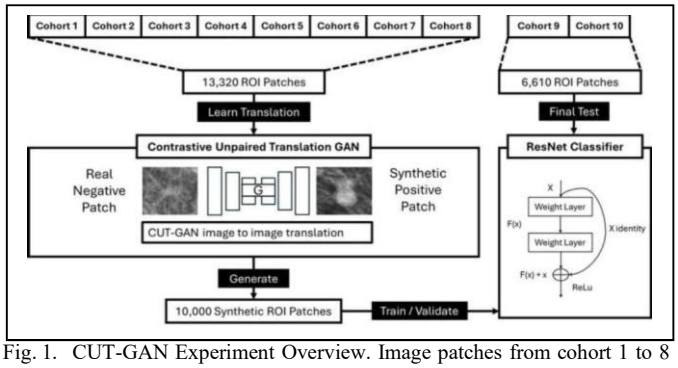


Fig. 1. CUT-GAN Experiment Overview. Image patches from cohort 1 to 8 are selected for CUT-GAN train, and image patches from cohort 9 and 10 are allocated to validate a ResNet

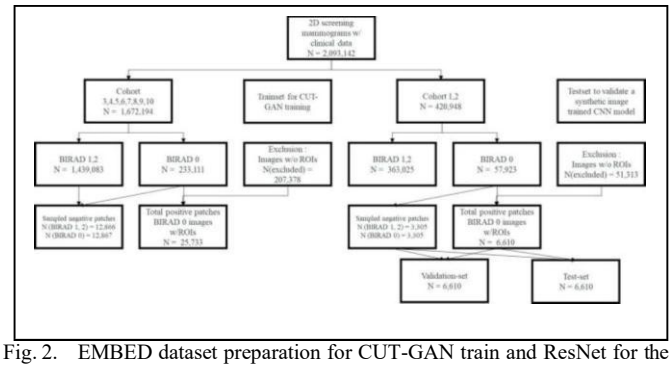


Fig. 2. EMBED dataset preparation for CUT-GAN train and ResNet for the validation of synthetic image trained model.

recall rates when validated on real images, this work supports the feasibility of privacy-preserving, synthetic-data-driven AI models for breast cancer diagnosis. This approach also enables AI-driven synthetic mammography generation in resource-limited settings and supports the creation of reliable breast cancer risk assessment models trained on synthetic data.

II. MATERIAL AND METHODS

A. Experimental design and overview

The experiment workflow is depicted in Figure 1, outlining the process of generating synthetic mammography patches using CUT-GAN and validating their effectiveness in the training of AI-based breast abnormality classification models, ResNet-based models. The study was conducted in two phases: (1) synthetic image generation using CUT-GAN and (2) model training and validation using ResNet models. In the first phase, mammography image patches from cohorts 1 to 8 of the Emory Breast Imaging Dataset (EMBED) were selected for CUT-GAN training. Normal patches were transformed into synthetic abnormal patches following a proposed pipeline, including image selection, window-level adjustment, and hyperparameter tuning. For the generation of high-quality synthetic mammography images, the CUT-GAN model was optimized by searching optimized hyperparameters for the model training. In the second phase, synthetic image-trained DL models were assessed using real mammography patches obtained from cohorts 9 and 10. ResNet models were trained on the generated synthetic dataset.

B. Data acquisition

To ensure generalizability of the research, this study utilizes EMBED, comprising 3.4 million breast mammography images from 383,379 screening and diagnostic exams conducted on 115,931 patients across four Emory University hospitals between 2013 and 2020 [22]. EMBED provides clinical imaging data including detailed demographics, image metadata, and pathologic outcomes, Breast Imaging Reporting and Data System (BI-RADS) scores, and tissue density classifications for both screening and diagnostic mammograms. Additionally, the dataset contains Regions of Interest (ROIs) identified by radiologists.

C. Image data preprocessing for synthetic patch generation

The template is used to format your paper and style the text. All The data preparation workflow for the EMBED is depicted in Figure 2. Positive patches were defined as the selection of BI-RADS 0 images containing radiologist-annotated ROIs, indicating potential abnormalities. Negative patches were sourced from two categories: (1) non-annotated regions within BI-RADS 0 images, and (2) regions extracted from negative screening mammograms classified as BI-RADS 1 or 2. This data preprocessing approach employed in this study aligns with methods used in previous research [23], [24].

All negative and positive image patches of 256 × 256 pixels were extracted from regions containing tissue pixels, ensuring over 70% of pixels contain tissue images, avoiding background air, whether they contained tumor or normal tissue. Additionally, mammography image patches with a color count below the first quartile of the distribution were removed from the training set to reduce likelihood of train failure, as summarized in Table 1. Examples of synthetic images generated by a train failure model are presented in Figure 3(a) and Figure 3(b). These images exhibit unidentified objects, such as black or white circle noises in synthetic image patches, or completely empty pixel patches which are the results of train failure in CUT-GAN training.

D. CUT-GAN configuration

Training of CUT-GAN requires optimal selection of hyperparameters to ensure stability in CUT-GAN training, facilitating the generation of high-quality synthetic mammography patches. Therefore, identifying the key parameters for an effective training phase is essential. Selecting an appropriate optimization strategy is critical to minimizing errors by estimating the weights of the CUT-GAN model. Although several optimization techniques have been developed to minimize loss functions, gradient descent (GD) remains the best method in the deep learning community. In this study, GD was adopted for the training of the CUT-GAN model, and the adaptive moment estimation (ADAM) algorithm—a first-order, gradient-based optimization approach for stochastic objective functions—was employed as the optimizer. It has gained significant attention among GAN-based researchers due to its effectiveness in model training.

The key hyperparameters optimized in CUT-GAN training include the learning rate, batch size, weight decay, β_1, and β_2.

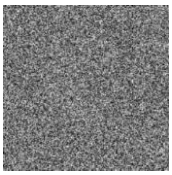


Fig. 3(a). Synthetic mammogram patch generation failed due to train failure.

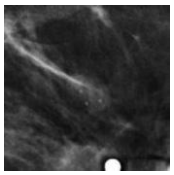


Fig. 3(b). Synthetic mammogram patch generation failed due to train failure.

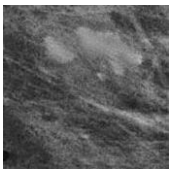


Fig. 3(c). Synthetic mammogram patch generation failed due to insufficient training.

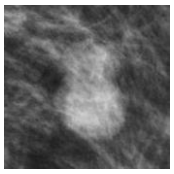


Fig. 3(d). Successful synthetic mammogram patch generation

Fig. 3. Synthetic images created by CUT-GAN model. Synthetic images (a) and (b) created when train failure occurs in the training phase. CUT-GAN generates a synthetic image (c) when CUT-GAN is trained for 200 epochs. Synthetic image (d) is created when CUT-GAN is trained for 400 epochs.

Additionally, batch size and number of epochs influence training performance. Experiments were conducted with a variety of batch sizes, such as 8, 32, and 128 – more GPU memory is taken where larger batch sizes reduce training time but require more GPU memory. The number of epochs was set from 200 to 400 to improve the image quality of synthetic mammography patches as shown in Figure 3(c). This synthetic image quality now resembles those of real mammography patches without noise pixels. The learning rate is a critical hyperparameter for weight updates in GAN training, with this experiment carried out with values in the range from 0.1 to 0.0001. In the ADAM optimization algorithm, the learning rate (α) regulates weight updates at the end of each batch. Meanwhile, the momentum hyperparameter determines which previous updates influence the current weight update optimizing the model training.

While training GANs, selecting a small learning rate (e.g., 0.00001) slows the weight update rate, whereas larger learning rates (e.g., 0.3) can lead to quicker learning. To ensure the generation of well-developed synthetic image patches, it is crucial for researchers to choose hyperparameters to train CUT-GAN model. The suggestion for optimal hyper-parameter settings is depicted in Table 2. Improper selection leads to the synthesis of underdeveloped image patches, as illustrated in Figure 3(d). Training and testing were performed using the TensorFlow library and Python. The computer system configuration used for CUT-GAN training was 2 x NVIDIA RTX3090 GPUs with a total of 24 GB of GPU RAM, and Intel (R) Core i7-7500 CPU 3.40 GHz; RAM of 128 GB do not revise any of the current designations.

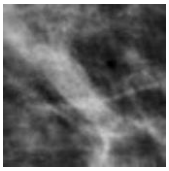


Fig. 4(a). Synthetic mammography patch with abnormality.

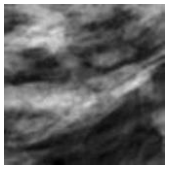


Fig. 4(b). Synthetic mammography patch with abnormality.

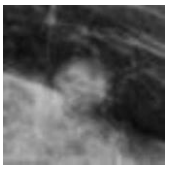


Fig. 4(c). Mammography image.

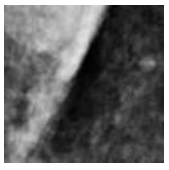


Fig. 4(d). Mammography image.

Fig. 4. Comparison Between Synthetic and Real Mammography Image Patches. Synthetic patch images resemble real mammography patches. The synthetic patches are generated to closely resemble real mammography patches in terms of texture, intensity distribution, and anatomical structures.

E. Postprocessing for synthetic image patches

To ensure generalizability of the research, this study utilizes Synthetic images generated by CUT-GAN contain three Red Green Blue (RGB) channels, whereas real mammography images are grayscale. To resolve this issue, the synthetic image type was converted from RGB to grayscale [25]. The window levels of the synthetic images and the real images demonstrate distinct distributions. This discrepancy negatively affects the performance of classification models trained on synthetic images. Therefore, the brightness and contrast distributions of the synthetic images were adjusted to match those of the real mammography images. This transformation was carried out using the Pillow library, ensuring the window-level distribution of the synthetic images fell within the acceptable window level range of the real images. A Student's t-test was performed to confirm that the adjusted distribution of window levels of the synthetic images and the window levels of the real mammography images showed insignificant differences between the means of those two distributions. This indicates that the distributions are statistically as close as possible with the P-value greater than 0.05 showing statistical insignificance, as demonstrated in Table 3 [26], and their examples of synthetic mammography patches are shown in Figure 4.

F. DL classifier trained on synthetic image patches

To Negative patches were carefully sampled to match positive patches in terms of size, abnormality location, and patient demographics when constructing the validation dataset. The imaging matching process first detected corresponding positive and negative screening images showing the highest image similarity. Referencing the architecture and position of ROI of the positive images, comparable patches were then extracted from the negative images. The patch sizes for the validation dataset were fixed at 256 × 256 pixels. The validation dataset comprises two subsets: a validation set and a test set, each containing 6,610 patches, resulting in a total of 13,220 patches.

TABLE I. PIXEL COLOR COUNT STATISTICS IN THE CUT-GAN TRAINSET

Statistics	Scores
Mean	187.54
Standard Deviation	32.34
Minimum	1.00
25%	167.00
50%	191.00
75%	211.00
Maximum	256.00

TABLE 2. OPTIMAL HYPER-PARAMETERS FOR CUT-GAN TRAINING

Hyper-parameters	Optimal values
Learning rate	0.0002
Beta 1	0.999
Beta 2	0.999
No. of epochs	400
Batch size	8
No. of generator filters	64
No. of discriminator filters	64

TABLE 3. TRANSFORMED IMAGE WINDOW LEVEL DISTRIBUTIONS FOR SYNTHETIC AND REAL MAMMOGRAPHY

	Brightness Real Images	Brightness Synthetic Images	Contrast Real Images	Contrast Synthetic Images
Mean	86.44	86.43	31.89	33.77
Standard Deviation	30.63	26.46	11.33	8.61
Minimum	29.71	30.68	8.51	12.85
25%	65.73	68.12	23.75	27.87
75%	99.64	99.46	37.05	37.84
Maximum	253.51	183.62	110.20	91.57
T-test Statistics	0.018		0.000	
P-value	>0.098		>2.203	

TABLE 4. MODEL PERFORMANCE MEASURES IN SYNTHETIC AND REAL MAMMOGRAPHY TRAINED MODELS

	AUC	Accuracy	Precision	Recall
Source and decompressed 1024x1024 images from 512x512 and 256x256 images				
ResNet50	0.92(0.91 – 0.93)	0.86(0.85 – 0.86)	0.86(0.85 – 0.87)	0.69(0.68 – 0.70)
ResNet101	0.91(0.909 – 0.92)	0.85(0.84 – 0.86)	0.87(0.85 – 0.88)	0.67(0.65 – 0.68)
ResNet152	0.91(0.90 – 0.92)	0.86(0.86 – 0.87)	0.83(0.82 – 0.84)	0.73(0.72 – 0.75)
ResNet models trained by synthetic mammography patches, and tested on real mammography patches				
ResNet50	0.65(0.64 – 0.66)	0.61(0.60 – 0.62)	0.44(0.43 – 0.46)	0.64(0.63 – 0.66)
ResNet101	0.71(0.70 – 0.72)	0.64(0.63 – 0.65)	0.48(0.47 – 0.49)	0.80(0.79 – 0.81)
ResNet152	0.75(0.74 – 0.76)	0.67(0.66 – 0.68)	0.50(0.49 – 0.51)	0.81(0.79 – 0.82)

ResNet models, including ImageNet pre-trained weights were trained on synthetic mammography patches to detect real mammography patches extracted from EMBED cohorts 9 and 10. Numerous studies have demonstrated the effectiveness of ResNet50 in medical image-based diagnosis, particularly for breast cancer classification and other disease diagnosis. This approach leverages the robust feature extraction capabilities of ResNet model variants [27], [28]. Facilitating transfer learning, the pre-trained ImageNet weights were frozen and five additional trainable layers were added to the pre-trained ResNet models as feature extractors. These added layers consist of 2048, 1024, 512, 128, 32, and 1 neuron as used in a previous AI assisted breast cancer research [24].

Additionally, The input layer was configured to process 256 × 256 pixel patches to analyze abnormality signals in synthetic image patches. The Rectified Linear Unit (ReLU) is applied to the hidden layers as its applicability is widespread across diverse domains [29]. Average pooling was also employed as the downsampling strategy for feature maps [30]. For the output layer, a sigmoid activation function was selected, as it is a well-established choice for binary classification tasks in AI medical imaging applications [31]. Binary cross-entropy, a commonly used loss function for binary classification in medical imaging applications, was employed in this study [32]. The optimal learning rate, $\lambda = 0.002$, for AI-assisted breast abnormality classifiers was determined using Bayesian optimization [33].

III. RESULT AND STATISTICAL ANALYSIS

For model performance evaluation, 95% confidence intervals of the assessment metrics were estimated on the test dataset. The confidence intervals were derived using bootstrapping, where 30 real mammography patches from the test dataset per sampling iteration, which had not been seen by the models during training, were randomly sampled and tested by both the synthetic and real image-trained models over 500 iterations.

As presented in Table 3, the brightness distribution of real images had a mean of 86.44 and a standard deviation of 30.63. To ensure compatibility, the brightness of the synthetic images in the train set was adjusted to closely match that of the validation set, resulting in a mean of 86.43 and a standard deviation of 26.46. Similarly, the contrast distribution of real images in the validation set exhibited a mean of 31.89 with a standard deviation of 11.33, so the synthetic images in the train set were adjusted to yield a contrast distribution with a mean of 33.77 and a standard deviation of 8.61, closely resembling the validation set. Student's t-tests revealed no statistically significant differences in brightness and contrast distributions between the real and synthetic images in both the train set and the validation set. Each of the validation and test datasets comprised 6,610 image patches, balanced between positive and negative labels, with each patch with 256 × 256 pixels.

The results in Table 4 indicate that models trained on authentic mammography patches exhibit higher performance across most evaluation metrics compared to those trained on synthetic patches except for recall. Specifically, models trained on real mammography data achieve consistently high area under the receiver operating characteristic curve (AUC) values (0.913–0.92), demonstrating strong binary classification capability. Accuracy is also significantly higher in these models, with values ranging from 0.857 to 0.865, whereas models trained on synthetic patches show substantially lower accuracy (36.8%–59.0%), indicating low performance compared to real mammography images.

Moreover, precision is notably higher in models trained with authentic data (83.5%–87.1%) than in those trained with synthetic data (34.2%–44.5%), suggesting that real-data-trained models produce fewer false negatives. In contrast, recall is markedly elevated in synthetic-data-trained models (93.1%–97.1%), albeit at the cost of precision, implying a tendency toward high sensitivity but poor specificity. This trade-off suggests that synthetic-data-trained models detect most positive cases but may yield an high rate of false positives.

The depth of the ResNet architecture appears to have a significant impact on performance when it is trained on synthetic image patches. For example, ResNet152 shows dramatic improvements in AUC and accuracy trained on synthetic data. However, the overall performance remains suboptimal compared to models trained with authentic patches.

IV. DISCUSSION

To the best of our knowledge, this study is the first research to apply a resource-efficient, low GPU memory GAN model, CUT-GAN, to generate synthetic mammography image patches. We also propose a pre-processing pipeline to train DL models on them to classify abnormality present in the breast mammography patches. The results imply that the DL models trained on synthetic mammography patches achieve performance that is competitive with models trained on real mammography data. Our findings suggest that successful GAN training requires the exclusion of mammography patches with color counts below the 25th percentile of the dataset's distribution and adjustment of GAN training configurations. Moreover, applying image window-level adjustments to synthetic mammography patches enhances their compatibility with real mammography images.

This research also has important implications for data privacy, particularly in contexts where medical imaging datasets are publicly released. The distribution of patient-derived images carries risks related to confidentiality and regulatory compliance. In contrast, synthetic mammography patches generated by GAN models are artificial images that preserve the structural and textural characteristics of real mammography data without revealing patient-identifiable information. As such, they can serve as effective substitutes for private datasets. Moreover, this approach reduces the need for researchers to upload sensitive patient data to external cloud-based generative AI platforms, such as DALL·E 2 or Stable Diffusion. By enabling the local generation of synthetic imaging datasets, our method not only aligns with privacy regulations but also supports the broader adoption of AI-driven imaging analysis in a secure and ethically responsible manner.

Our study found that CUT-GAN can be trained using as little as 3 GB of GPU memory, a significant reduction compared to the 24 GB required by other generative AI models, such as diffusion models. This efficiency enables researchers in resource-limited settings to leverage synthetic imaging techniques without the need for high-performance computing infrastructure.

Despite these advantages, several methodological challenges remain. One challenge pertains to window-level settings in synthetic images, which were not explicitly controlled during CUT-GAN training. Although window-level adjustments were applied post hoc to standardize brightness and contrast distributions, effective data augmentation between synthetic and real image patches could not be achieved to create a mixed training dataset. Better harmonization techniques should be explored to facilitate effective data augmentation between synthetic and real mammography image patches. Additionally, the process of selecting appropriate training images for CUT-GAN remains a critical factor. To prevent train failure during CUT-GAN training, we excluded patches with low color counts, but further research is needed to refine selection criteria to minimize data loss while preserving training stability.

A key finding of this study was that ResNet models trained on synthetic images demonstrated higher recall to detect abnormal patches compared to those trained on real mammograms. These results suggest that synthetic data augmentation may enhance the identification of true positive cases, a finding consistent with prior research in deep learning-based breast cancer detection. However, further validation using an augmented training dataset is necessary to improve the generalizability of the findings. Thus, it is essential to establish well-defined data augmentation pipelines to ensure the compatibility and applicability of the synthetic data with EMBED as well as Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) and the Categorized Digital Database for Contrast-Enhanced Spectral Mammography (CDD-CESM) beginning of a sentence.

V. CONCLUSION

This study highlights the feasibility of using low GPU memory – demanding GAN model, CUT-GAN, generated synthetic mammography images to support deep learning-based abnormality classification models in breast imaging. Synthetic images may serve as a valuable tool to address class imbalance and improve abnormality detection while offering privacy protection. Further research is needed to optimize synthetic image generation, evaluate its clinical utility, and explore its role in medical education and diagnostic support.

REFERENCES

[1] W. T. Tran *et al.*, "Computational radiology in breast cancer screening and diagnosis using artificial intelligence," *Canadian Association of Radiologists Journal*, vol. 72, no. 1, pp. 98–108, 2021.

[2] S. Kaymak, P. Esmaili, and A. Serener, "Deep learning for two-step classification of malignant pigmented skin lesions," presented at the 2018 14th Symposium on Neural Networks and Applications (NEUREL), IEEE, 2018, pp. 1–6.

[3] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Computers in biology and medicine*, vol. 113, p. 103423, 2019.

[4] Y. Ma *et al.*, "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm," presented at the 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2019, pp. 1–4.

[5] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, 2019.

[6] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks," *Journal of Medical Imaging*, vol. 6, no. 3, p. 031409, 2019.

[7] "Breast Cancer Facts and Statistics." [Online]. Available: https://www.breastcancer.org/

[8] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: a cancer journal for clinicians*, vol. 72, no. 1, pp. 7–33, 2022.

[9] O. Ginsburg *et al.*, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, pp. 2379–2393, 2020.

[10] M. C. Spayne, C. C. Gard, J. Skelly, D. L. Miglioretti, P. M. Vacek, and B. M. Geller, "Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study," *The breast journal*, vol. 18, no. 4, pp. 326–333, 2012.

[11] S. Zahoor, U. Shoaib, and I. U. Lali, "Breast cancer mammograms classification using deep neural network and entropy-controlled whale optimization algorithm," *Diagnostics*, vol. 12, no. 2, p. 557, 2022.

[12] A. Ciritsis *et al.*, "Determination of mammographic breast density using a deep convolutional neural network," *The British journal of radiology*, vol. 92, no. 1093, p. 20180691, 2019.

[13] R. Longadge and S. Dongre, "Class imbalance problem in data mining review," *arXiv preprint arXiv:1305.1707*, 2013.

[14] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[15] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem," *Int. J. Advance Soft Compu. Appl*, vol. 5, no. 3, pp. 176–204, 2013.

[16] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: when to warp?," *Problems in Diagnostic Radiology*, vol. 53, no. 3, pp. 346–352, 2024.

[24] I. Hwang *et al.*, "Impact of multi-source data augmentation on performance of convolutional neural networks for abnormality classification in mammography," *Frontiers in Radiology*, vol. 3, p. 1181190, 2023.

[25] A. Clark, "Pillow (pil fork) documentation," *readthedocs*, 2015.

[26] Student, "The probable error of a mean," *Biometrika*, pp. 1–25, 1908.

[27] A. Yala, T. Schuster, R. Miles, R. Barzilay, and C. Lehman, "A deep learning model to triage screening mammograms: a simulation study," *Radiology*, vol. 293, no. 1, pp. 38–46, 2019.

[28] X. Yu, C. Kang, D. S. Guttery, S. Kadry, Y. Chen, and Y.-D. Zhang, "ResNet-SCDA-50 for breast abnormality classification," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 1, pp. 94–102, 2020.

[29] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.

presented at the 2016 international conference on digital image computing: techniques and applications (DICTA), IEEE, 2016, pp. 1–6.

[17] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN -based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[18] T. Park, A. A. Efros, R. Zhang, and J.- Y. Zhu, "Contrastive learning for unpaired image-to-image translation," presented at the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, Springer, 2020, pp. 319–345.

[19] S. M. Moore *et al.*, "De -identification of medical images with retention of scientific research value," *Radiographics*, vol. 35, no. 3, pp. 727–735, 2015.

[20] K. Packhäuser, S. Gündel, N. Münster, C. Syben, V. Christlein, and A. Maier, "Is medical chest X-ray data anonymous?," 2021.

[21] H. Montenegro, W. Silva, and J. S. Cardoso, "Privacy-preserving generative adversarial network for case-based explainability in medical image analysis," *IEEE Access*, vol. 9, pp. 148037–148047, 2021.

[22] J. J. Jeong *et al.*, "The EMory BrEast imaging Dataset (EMBED): A racially diverse, granular dataset of 3.4 million screening and diagnostic mammographic images," *Radiology: Artificial Intelligence*, vol. 5, no. 1, p. e220047, 2023.

[23] V. Nalla *et al*, "Deep learning for computer-aided abnormalities classification in digital mammogram: A data-centric perspective," *Current*

[30] F. Bieder, R. Sandkühler, and P. C. Cattin, "Comparison of methods generalizing max-and average-pooling," *arXiv preprint arXiv:2103.01746*, 2021.

[31] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," presented at the International workshop on artificial neural networks, Springer, 1995, pp. 195–201.

[32] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, and A. Maria Vanegas, "Breast cancer histopathology image classification using an ensemble of deep learning models," *Sensors*, vol. 20, no. 16, p. 4373, 2020.

[33] Z. Douglas and H. Wang, "Bayesian optimization -derived batch size and learning rate scheduling in deep neural network training for head and neck tumor segmentation," presented at the 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), IEEE, 2022, pp. 48–54.