EAI Endorsed Transactions
on AI and Robotics
Research Article EAI.EU

Leveraging synthetic mammograms to enhance deep-learning performance for breast cancer classification using EfficientNetV2L architecture

R. Sutjiadi[1,2], S. Sendari[1,*], H.W. Herwanto[1] and Y. Kristian[3]

[1] Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang, Malang, East Java, Indonesia
[2] Department of Informatics, Faculty of Information Technology, Institut Informatika Indonesia Surabaya, Surabaya, East Java, Indonesia
[3] Department of Informatics, Faculty of Science and Technology, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, East Java, Indonesia

Abstract

INTRODUCTION: To improve survival rates for breast cancer, a leading cause of female mortality globally, early detection is essential. This study presents a deep learning framework for classifying mammogram images as normal or abnormal.
OBJECTIVES: This research aims to enhance the performance of a deep learning model for breast cancer classification by augmenting a real mammogram dataset with synthetic images. The study evaluates the impact of progressively increasing the number of synthetic mammograms on the model's accuracy, precision, recall, and F1-score.
METHODS: The approach utilizes the EfficientNetV2L model for classification. Data augmentation was performed by generating synthetic mammograms using Denoising Diffusion Probabilistic Models (DDPM). A baseline dataset of 410 real mammograms from the INbreast public dataset was augmented with an increasing number of synthetic images across four experimental scenarios.
RESULTS: The model demonstrated substantial performance gains directly linked to the use of synthetic data. The best performance was achieved when 500 synthetic images were used, resulting in all evaluation metrics exceeding a score of 0.90. The results confirm that incorporating more synthetic images is a key factor in achieving both higher classification accuracy and more stable training convergence.
CONCLUSION: These findings highlight the significant potential of synthetic image augmentation to address data scarcity, class imbalance, and model generalisation in medical image analysis. This method provides a scalable and privacy-preserving solution for breast cancer screening systems.

Keywords: breast cancer detection, synthetic mammograms, efficientnetv2l, denoising diffusion probabilistic models, deep learning in medical imaging.

Received on 18 July2025, accepted on 30 August 2025, published on 01 September 2025

doi: 10.4108/airo.9749

1. Introduction

Breast cancer is one of the most common causes of death in women worldwide. In 2022, 2.3 million new cases of female breast cancer and 670,000 deaths from female breast cancer occurred worldwide, with annual rates increasing by 1-5% [1]. By 2050, the number of new cases is expected to rise by 38%, while deaths will increase by 68%, with low Human Development Index (HDI) countries being disproportionately affected [1]. In breast cancer cases, the mortality rate is inversely proportional to the speed of diagnosis and early treatment. Early detection, accompanied by regular periodic

*Corresponding author. Email: siti.sendari.ft@um.ac.id

treatment or therapy, can effectively improve the survival rate of female breast cancer patients [2].

One of the commonly used breast cancer detection methods is mammography. Mammography is a medical imaging tool based on X-rays, used to observe soft tissue in the breast to detect the presence of abnormal tissue that may indicate cancer [3]. Mammography is relatively safer, more comfortable, and more cost-effective while maintaining a reasonably good level of accuracy compared to other methods such as Ultrasound, Magnetic Resonance Imaging (MRI), Magnetic Resonance Spectroscopy, and Positron Emission Tomography Conjugated with Computed Tomography (PET-CT) [4]. Screening using mammography with an adequately targeted population can effectively reduce breast cancer mortality rates [5].

Although mammography is still the gold standard method for detecting breast cancer as early as possible, it has a significant tendency to produce high rates of false positive results in detecting cancerous tissue, especially for women with dense breast tissue [6]. This is due to mammography's inability to differentiate between benign, malignant, and healthy tumours in the soft tissue of the breast. Doctors or radiologists sometimes require additional examinations to establish a diagnosis, such as other screening methods or a biopsy procedure, where a small sample of suspected cancerous breast tissue is taken for laboratory examination. This procedure results in both physical discomfort and increased costs for the patient.

With advancements in artificial intelligence, several studies have been conducted to develop a Computer-Aided Diagnosis (CAD) system that uses deep-learning methods to analyse mammography images and detect abnormal tissue in the female breast [7], [8], [9], [10]. The results of these studies indicate that CAD systems utilising deep-learning methods can reliably and accurately detect abnormal breast tissue from mammography images. Several factors determine the accuracy of deep-learning methods, including dataset diversity (variance) [11], dataset balance [12], and dataset quantity [13] used during the training process to develop the model. An ideal dataset should have sufficient diversity to represent all expected prediction classes, balance across all prediction classes, and an adequate quantity of data [14]. It is known that the generalisation and robustness of deep-learning models can be significantly improved with large-scale datasets [15].

In contrast, data scarcity in the medical domain is still a significant problem. Medical data is complicated to obtain because it is classified as confidential personal data protected by local laws and regulations. In accordance with the World Medical Association's Declaration of Helsinki, ethical clearance is also necessary before biomedical research involving human subjects can be conducted [16]. Therefore, there are challenges in researching the application of deep-learning methods to medical data, especially medical imaging data. Medical image datasets tend to have relatively low diversity, are often imbalanced, and are limited in quantity, which creates immense problems for researchers.

Several publicly available mammography datasets, including INbreast [17], MIAS [18], and CBIS-DDSM [19] are commonly used by deep-learning researchers. Each dataset has different image characteristics and class classifications. Furthermore, using these datasets necessitates a number of conventional image augmentation techniques, like zooming, flipping, and rotating, to improve variability and maximise the accuracy of deep learning systems [20]. However, traditional image augmentation techniques only modify the representation of images but do not create new variations of mammogram objects.

Generating synthetic data is one promising approach to address data scarcity in medical image analysis that has gained attention in recent literature. Generative models, including Generative Adversarial Networks (GANs) [21], Denoising Diffusion Probabilistic Models (DDPMs) [22], and Variational Autoencoders (VAEs) [23] have shown promising results in generating realistic medical images similar to actual patient data. Several studies have explored deep learning for generating synthetic medical images, reaching a promising stage where computers can entirely generate and process medical images.

Yi et al. (2019) conducted research investigating the use of Generative Adversarial Networks (GANs) in medical imaging [24]. GANs are a type of neural network model consisting of two networks: one focused on image generation during training and the other on discrimination [25]. This research aims to serve as a dataset generator in medical radiology, addressing data scarcity while protecting patient privacy in research dataset collection. Nevertheless, when the generator only generates a small range of samples and is unable to fully represent the range of variability found in the training data, the GAN tends to experience mode collapse [26].

Rais et al (2024) studied the exploration of the generation of medical imaging using Variational Autoencoders (VAEs) [27]. VAEs are a promising method for augmenting medical imaging and addressing data scarcity. Also, the use of VAEs to augment images could improve machine learning performance by minimising overfitting and enhancing data variety. On the other hand, this research also highlights the fact that images generated by VAEs are not as realistic as those generated by GANs.

Muller-Franzes et al. (2023) also studied another synthetic medical image generation method called Medfusion [28]. Medfusion is a conditional latent Denoising Diffusion Probabilistic Model (DDPM) used as a synthetic dataset generator for medical imaging. DDPM consists of two processes: forward diffusion (noise addition), which gradually adds noise using a Gaussian distribution, and reverse diffusion (denoising and generation), which learns to reverse the noise process, progressively reconstructing the image. It aims to enhance the performance of deep-learning-based health anomaly detection systems, including glaucoma in the eye, cardiomegaly in the lungs, and colorectal cancer.

This research aims to utilise a DDPM-based generative artificial intelligence method to produce synthetic mammograms and assess their impact on the accuracy of deep-learning systems for mammogram classification. This would address the scarcity of real mammograms as a dataset source for deep-learning systems without being restricted by patient data confidentiality. Synthetic mammograms enhance dataset diversity, improve class balance, and increase the dataset size, thereby addressing key challenges in deep-learning research.

In this paper, our main contribution is as follows:

(i) We propose a DDPM-based method to generate synthetic mammograms for two output classes: normal and abnormal. This approach is used as data augmentation to increase the size and diversity of the real dataset (in this case, we use the public INbreast dataset).
(ii) We examine the effect of adding specific quantities of synthetic mammograms to the real dataset. We run several scenarios to evaluate the impact of these synthetic images on the performance of a machine learning model for abnormality detection.
(iii) We explore the potential of the DDPM-based method to complement traditional image augmentation techniques. Furthermore, we see this approach as a potential solution for researchers to improve medical datasets in terms of size, diversity, and class balance, without violating patient data privacy.

2. Methods

2.1. Synthetic Mammograms Generator

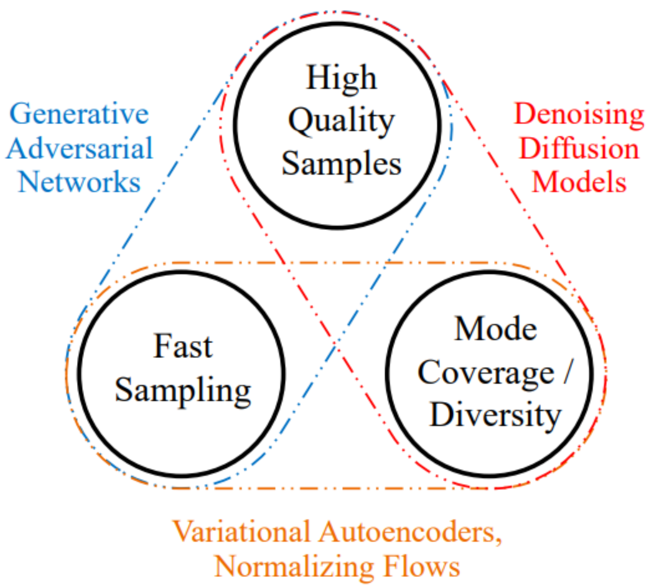


Figure 1. Generative Learning Trilemma [29]

This research utilises the DDPM system as a framework for generating synthetic mammograms [30], a decision justified by its strategic navigation of the generative learning trilemma [29] as shown in Figure 1. For a high-stakes application like mammography, the exceptional sample quality and diversity offered by DDPMs are paramount. The framework excels at producing high-fidelity images that capture critical diagnostic details. It also avoids the mode collapse common in other models like GANs, ensuring the generated dataset represents the full spectrum of anatomical and pathological variations. While DDPMs take longer to generate images, this is an acceptable drawback for our purposes. Since we are creating a dataset offline rather than needing real-time results, the exceptional quality and diversity offered by DDPMs make them the best option over faster methods.

Figure 2 illustrates the DDPM system block diagram, which consists of two primary subsystems: forward diffusion and reverse diffusion [22]. Forward diffusion takes an input in the form of a real mammogram image (x0). The system then gradually adds Gaussian noise over multistep (T), producing intermediate noisy images (xt). After enough steps, the image becomes pure Gaussian noise. This pure Gaussian noise then reverts to an image using a noise predictor. The noise predictor is trained using a U-Net deep-learning model to predict and remove noise. The model learns the mapping between the noisy image (Xt) and the original image (x0). Reducing the discrepancy between actual and predicted noise is the aim of noise predictor training. After training the noise predictor, the system starts the reverse diffusion process. The trained model removes noise from pure Gaussian noise step by step, reconstructing an image. The final output is a generated synthetic mammogram.

We use the publicly available INbreast mammogram dataset to train the DDPM model. The selection of INbreast is motivated by its use of digital mammograms, which provide higher image resolution as they do not undergo film scanning. Furthermore, the standardised nature of the images reduces potential variability, thereby minimising factors that could interfere with the training process.

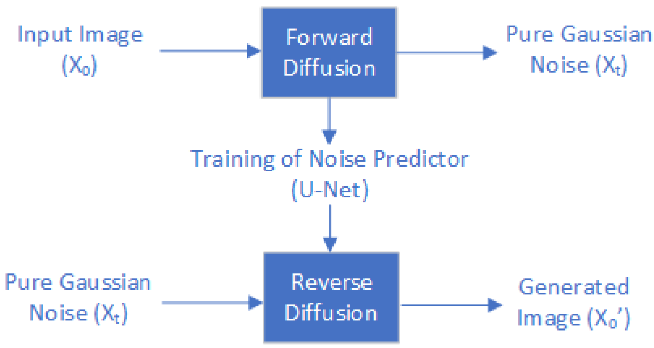


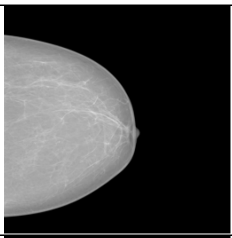
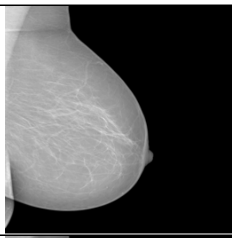
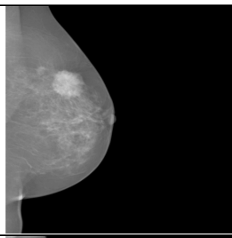
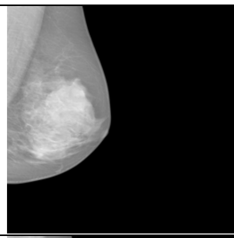
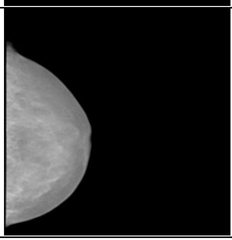
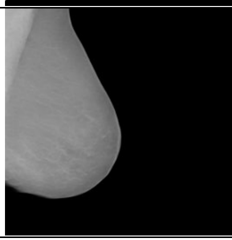
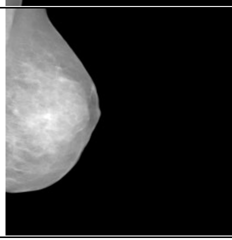
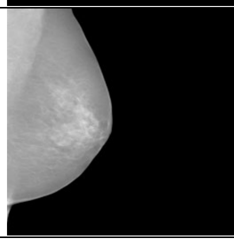
Figure 2. Block Diagram of DDPM System

This dataset comprises 410 mammogram images in Mediolateral Oblique (MLO) and Craniocaudal (CC) projections that are separated into eight BI-RADS (Breast Imaging-Reporting and Data System) categories. To simplify image generation, the eight BI-RADS categories were classified into 'normal' and 'abnormal' labels, and 250 synthetic mammograms were generated for each label. Since the dataset between normal and abnormal is imbalanced, we use the class weights for DDPM training, where the smaller number of normal class is assigned higher weights than the abnormal class. These 500 synthetic mammogram images, combined with 410 real INbreast mammograms, are utilised as the training dataset for a deep-learning system to detect and classify abnormalities in female breast imaging. Detailed dataset specifications are shown in Table 1. Samples of real and generated synthetic mammograms for each label are shown in Table 2.

Table 1. INbreast Dataset Specifications

INbreast Real Dataset		**Synthetic Dataset**	
Category	**Subtotal**	**Label**	**Generated Synthetic Images**
BI-RADS 1	67	Normal	250
BI-RADS 2	220	Abnormal	250
BI-RADS 3	23		
BI-RADS 4a	13		
BI-RADS 4b	8		
BI-RADS 4c	22		
BI-RADS 5	49		
BI-RADS 6	8		
Grand Total	**410**		**500**

Table 2. Samples of Real and Synthetic Mammograms

	Normal		**Abnormal**	
Real Mammogram				
Synthetic Mammogram				

2.2. Deep-Learning Method for Mammogram Classification

Mammogram classification is critical in breast cancer detection, requiring highly accurate and efficient deep-learning models [31]. In image classification applications, conventional Convolutional Neural Networks (CNNs) have proven to be effective. However, they often suffer from high computational costs, making them less feasible for large-scale applications. Researchers have explored efficient architectures that balance performance and computational efficiency to address this issue, such as the EfficientNet family.

EfficientNetV2 is a deep-learning architecture model designed to improve accuracy and efficiency [32]. It outperforms earlier state-of-the-art models in terms of parameter efficiency and training speed. It builds upon the original EfficientNet architecture by introducing several enhancements. EfficientNetV2 utilises a progressive learning strategy, where training begins with lower-resolution images and gradually increases to higher resolutions. This approach helps stabilise training while improving model generalisation.

Another key innovation in EfficientNetV2 is the introduction of fused-MBConv layers. Unlike standard depthwise separable convolutions used in EfficientNetV1, fused convolutions improve training speed and efficiency by combining depthwise and pointwise convolutions into a single operation. This modification results in faster convergence and reduced memory requirements, making the model more suitable for large-scale vision tasks, such as medical imaging.

EfficientNetV2 comes in four main variants: EfficientNetV2-S, M, L, and XL, each designed to balance accuracy and computational efficiency for different use cases. EfficientNetV2S is lightweight and ideal for mobile and edge devices. EfficientNetV2M balances speed and accuracy, making it suitable for general-purpose image classification. EfficientNetV2L provides higher accuracy and is used for large-scale datasets. EfficientNetV2XL is the most powerful variant, designed for research and high-performance computing. These variants leverage Fused-MBConv layers, progressive learning, and optimised scaling to achieve superior accuracy while significantly reducing training time and computational costs compared to previous models. Table 3 summarises the comparison among EfficientNetV2 variants.

This study will employ EfficientNetV2L as a robust and reliable deep-learning model for detecting and classifying abnormalities in mammogram images. While the model exhibits high complexity in extracting detailed structural features from mammogram images, it remains computationally efficient, making it well-suited for large-scale medical imaging applications. The EfficientNetV2L network architecture is structured into eight distinct stages, as illustrated in Figure 3 below.

EfficientNetV2L consists of the following key blocks:

(i) **Fused-MBConv (Fused-Mobile Inverted Bottleneck Convolution)**
It is used in the early layers of the network, combining standard convolution and depthwise separable convolution into a single operation. This layer improves training efficiency by reducing memory access overhead.

(ii) **MBConv (Mobile Inverted Bottleneck Convolution)**
It is used in deeper layers of the network, consisting of a depthwise separable convolution followed by a pointwise convolution. This layer includes Squeeze-and-Excitation (SE) blocks to enhance essential features.

(iii) **Convolutional Layers**
A 3×3 convolutional layer at the beginning for initial feature extraction and a 1×1 convolution at the end before fully connected layers.

(iv) **Squeeze-and-Excitation (SE) Blocks**
It is used inside MBConv blocks to enhance essential features by recalibrating channel-wise activations and helps distinguish subtle differences in mammogram images.

(v) **Progressive Learning Strategy**
The model starts training with lower-resolution images and gradually increases image size while adjusting regularisation. This part prevents overfitting and speeds up convergence.

Table 3. Comparison Among EfficientNetV2 Variants

Variant	No. Parameters	FLOPs*	Application
EfficientNetV2S	22 M	8.8 B	Edge devices, mobile apps
EfficientNetV2M	54 M	24 B	General-purpose classification
EfficientNetV2L	120 M	53 B	Large-scale datasets, medical imaging
EfficientNetV2XL	208 M	94 B	Research, high-performance tasks

*(Floating Point Operations per Second) is a metric that measures the computational complexity of a neural network model.

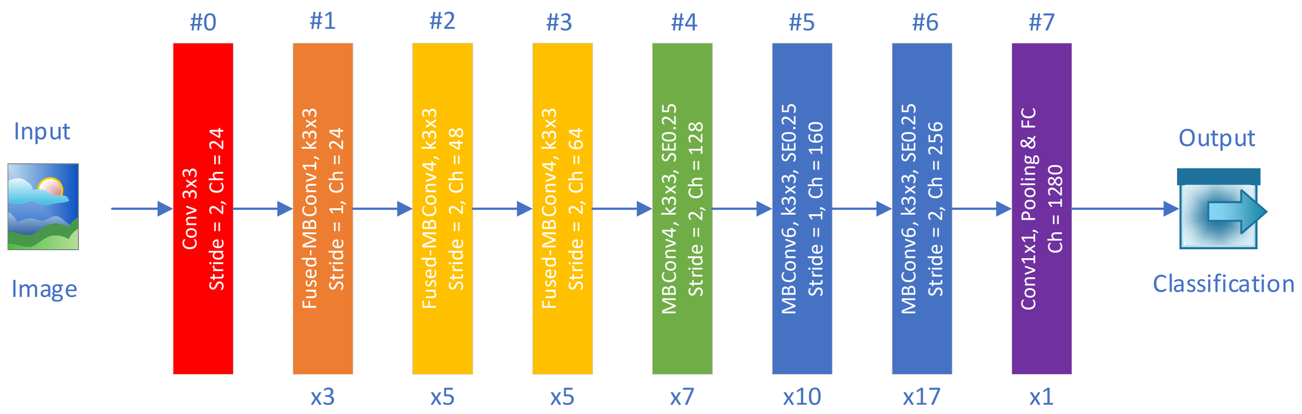


Figure 3. EfficientNetV2L Architecture

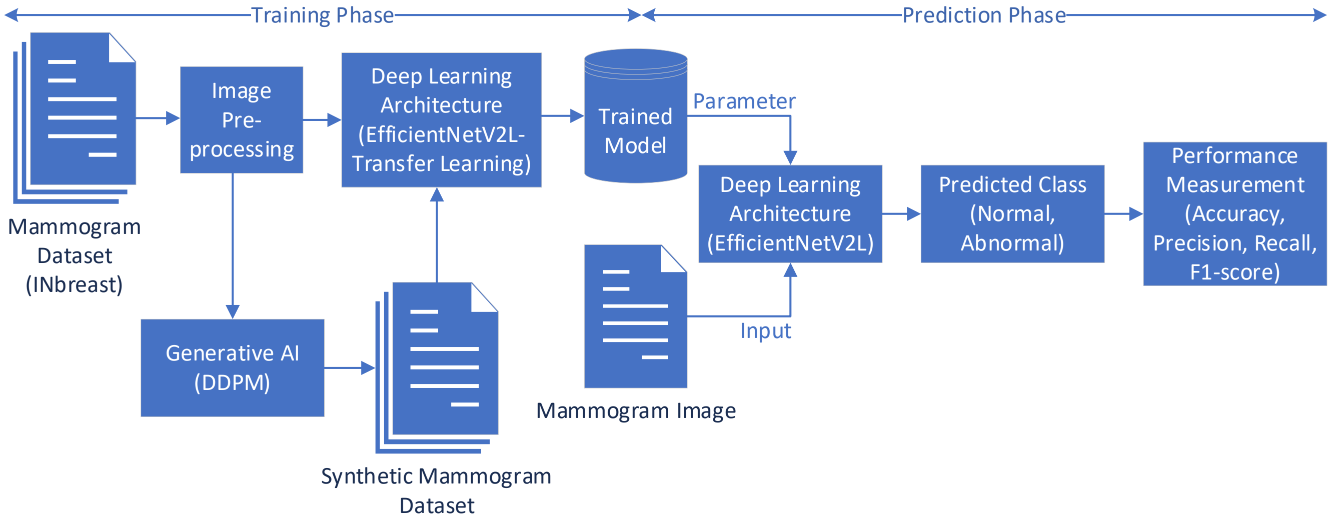


Figure 4. Block Diagram of CAD System Utilising Synthetic Mammogram Dataset

In the context of breast cancer detection, EfficientNetV2L is the most suitable model for building an accurate and computationally efficient deep-learning system. Its optimised architecture enables the extraction of fine-grained and complex structural details in female breast imaging, facilitating the differentiation between normal and abnormal classification with high sensitivity and specificity. The model's enhanced computational efficiency makes integration into CAD systems feasible, where rapid and reliable classification of mammogram images is essential. Figure 4 presents the block diagram of the CAD system for mammography, which is designed based on the EfficientNetV2L architecture, utilising the combination of real and synthetic mammogram datasets.

The INbreast dataset is the primary training dataset for the deep-learning system. Before training, the INbreast dataset undergoes image pre-processing, which includes resizing the images to 256×256 pixels and aligning breast object positions to the left side of the images to standardise their orientation. This step aims to reduce image variation, thereby optimising the system's accuracy. The combined real and synthetic datasets are then used to train the EfficientNetV2L model through transfer learning, utilising pre-trained ImageNet weights. The system is evaluated using multiple dataset configurations, incorporating varying proportions of real and synthetic data, and assessed based on performance metrics such as accuracy, precision, recall, and F1-score.

3. Results and Discussion

This study aims to evaluate the effectiveness of the EfficientNetV2L deep learning architecture for breast cancer detection, specifically in classifying mammogram images into two categories: normal and abnormal. The research is structured into four experimental scenarios to assess the impact of synthetic image augmentation on model performance. In the first scenario, the model is trained using 410 real mammogram images from the INbreast dataset, which serves as the baseline. The second scenario supplements the real dataset with 100 synthetic images, evenly distributed between the two classes (50 synthetic images per class). The third scenario expands the dataset further by adding 250 synthetic images, comprising 125 per class. In the fourth and final scenario, 500 synthetic images (250 per class) are added to the original dataset.

For each scenario, the combined dataset is partitioned into training and testing subsets with a ratio of 80:20. This consistent split ensures a fair comparison across all scenarios. It allows the evaluation of the model's generalisation ability. For the training and testing stage, the parameters of EfficientNetV2L are configured as follows:

- Adam Optimizer: Learning Rate = 0.0001, Beta 1 = 0.9, Beta 2 = 0.999, AMSGrad = True.
- Class Weights: Balanced.
- Model: loss = binary crossentropy, metrics = accuracy, batch size = 16, epoch = 50.

A quantitative evaluation of the EfficientNetV2L model's classification performance was conducted under each experimental configuration. The assessment utilized four key performance indicators: accuracy, precision, recall, and the F1-score. The selection of these metrics ensures a comprehensive analysis of the model's effectiveness—a critical requirement in medical imaging where diagnostic sensitivity and specificity are paramount. Specifically, accuracy gauges overall correctness, precision measures the positive predictive value, recall (or sensitivity) assesses the true positive rate, and the F1-score provides a balanced evaluation vital in contexts of class imbalance or significant misclassification consequences.

The central objective of this research is to quantify the impact of synthetically generated images on the efficacy of deep learning frameworks for breast cancer classification. Through a methodical analysis of the four aforementioned scenarios, this investigation seeks to yield valuable insights into the utility of synthetic data augmentation. The anticipated outcomes include a clearer understanding of how such techniques can enhance diagnostic precision and provide robust support for clinical decision-making processes.

The results in Table 4 show a clear trend: including synthetic images substantially improves the EfficientNetV2L model's performance in classifying normal and abnormal mammograms. In Scenario 1, where only 410 real images from the INbreast dataset were used, the model achieved an overall accuracy of 0.84. However, the two classes have a noticeable disparity in performance. While the abnormal class exhibited strong results—with precision, recall, and F1-score values of 0.88, 0.94, and 0.91, respectively—the normal class showed considerably lower performance, with a precision of 0.50, a recall of 0.31, and an F1-score of only 0.38. This suggests that when trained solely on a limited number of real samples, the model tends to be biased toward the abnormal class, likely due to the subtle features of normal cases being more challenging to learn with limited data.

In Scenario 2, where 100 synthetic images (50 per class) were added to the original dataset, overall performance improved notably. The accuracy increased to 0.90, and the model showed balanced improvement across both classes. The F1-score rose significantly to 0.78 for the normal class, while the abnormal class maintained strong performance with an F1-score of 0.94. These results indicate that even a moderate amount of synthetic data helps mitigate class imbalance and enhances the model's generalisation ability.

Scenario 3, with 250 synthetic images added, maintained the accuracy level of 0.93 but demonstrated further refinement in class-specific metrics. The precision, recall, and F1-scores remained consistently high and balanced across both classes, suggesting that additional synthetic data reinforces the model's learning without introducing overfitting or noise.

In Scenario 4, augmenting the dataset with 500 synthetic images (250 per class) resulted in the highest performance across all metrics. The model achieved an overall accuracy of 0.94. Both classes reported identical and robust evaluation scores, with precision, recall, and F1-scores of 0.90 or higher. This outcome illustrates the strong positive effect of large-scale synthetic data in improving accuracy and the stability and consistency of predictions across both classes.

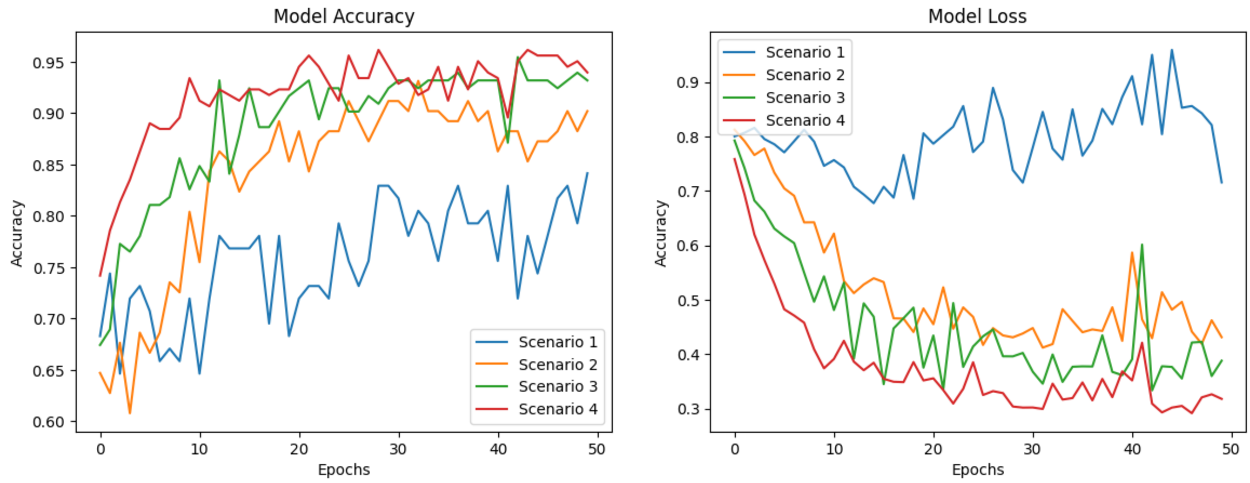
Figure 5 illustrates the EfficientNetV2L model's training dynamics in accuracy and loss over 50 epochs for the four experimental scenarios above. Figure 5(a) shows the model's accuracy throughout the training. It is evident that Scenario 1, which utilises only 410 real images, demonstrates the lowest and most unstable accuracy curve. The accuracy in this scenario fluctuates significantly, ranging between 0.65 and 0.84, indicating challenges in convergence and generalisation due to the limited training data size. In contrast, Scenario 2, which includes 100 synthetic images, substantially improves accuracy and stability, maintaining values consistently above 0.80. Scenario 3 and Scenario 4 further enhance performance, with Scenario 4 (including 500 synthetic images) achieving the highest and most stable accuracy, consistently remaining above 0.90. These results indicate that increasing the quantity of synthetic data leads to better generalisation and training stability, likely due to the improved diversity and balance in the training set.

Figure 5(b) depicts the model loss over the same training period. A similar trend is observed, wherein Scenario 1 exhibits the highest and most fluctuating loss values, often exceeding 0.8. This variability further supports the observation that the model struggles to converge when trained on a small dataset. As synthetic data is introduced in Scenarios 2 to 4, the loss decreases progressively, and the training process becomes more stable. Scenario 4, which includes the most significant volume of synthetic data, consistently achieves the lowest loss values (around 0.30), indicating more effective learning and reduced prediction errors.

Taken together, these results confirm that synthetic data augmentation has a significant positive impact on training stability and model performance. Including a larger and class-balanced synthetic dataset enhances classification accuracy and facilitates smoother and faster convergence by reducing training loss. These findings highlight the value of synthetic data in medical imaging tasks, particularly when real data availability is limited.

Table 4. The Research Result for Four Different Scenarios

#	INbreast Dataset	Synthetic Dataset	Class	Evaluation Metrics			
				Accuracy	Precision	Recall	F1-score
1	410	0	Normal	0.84	0.50	0.31	0.38
			Abnormal		0.88	0.94	0.91
2	410	100	Normal	0.90	0.78	0.78	0.78
			Abnormal		0.94	0.94	0.94
3	410	250	Normal	0.93	0.84	0.95	0.89
			Abnormal		0.98	0.93	0.95
4	410	500	Normal	0.94	0.92	0.90	0.91
			Abnormal		0.95	0.96	0.95



(a) (b)

Figure 5. Model Accuracy and Loss for Four Different Scenarios

Despite the benefits, the use of synthetic data is not without risks. The generative model could introduce subtle, systematic artefacts into the images that the classification model might learn, potentially leading to overfitting on synthetic features rather than true pathological indicators. Furthermore, if the initial real dataset is not diverse enough, the DDPM may only learn to reproduce a limited set of variations, which could create a false sense of security regarding model generalisation.

4. Conclusion

This research demonstrates the effectiveness of synthetic mammogram augmentation in improving deep-learning-based breast cancer detection using the EfficientNetV2L architecture. By incorporating synthetic images generated via Denoising Diffusion Probabilistic Models (DDPM), the model's performance was evaluated across four training scenarios with increasing volumes of synthetic images. The results consistently indicate that augmenting real mammograms with synthetic data substantially improves classification accuracy, stability, and generalisation.

When trained solely on real data, the EfficientNetV2L model exhibited significant class imbalance in performance, particularly underperforming in detecting normal cases. However, as synthetic data was progressively introduced and balanced across classes, the model achieved notable gains across all evaluation metrics—including accuracy, precision, recall, and F1-score. In the final scenario, which incorporated 500 synthetic images, the model reached peak performance, with all metrics exceeding 0.90 for normal and abnormal classifications. In addition to improved evaluation metrics, synthetic augmentation made the training process more stable, as evidenced by reduced loss and smoother convergence.

The findings underscore the critical role of high-quality, class-balanced synthetic data in addressing data scarcity, class imbalance, and generalisation challenges in medical imaging.. This research reinforces the potential of generative AI techniques such as DDPM to create scalable, privacy-preserving datasets that can significantly enhance deep-learning models in healthcare.

Future work may explore integrating multimodal synthetic data and extending this framework to multiclass or multi-view mammography classification tasks, further advancing the development of reliable, AI-assisted diagnostic tools in clinical environments. Also, future work should focus on cross-dataset validation to ensure the model performs robustly on mammograms from different sources and patient populations.

References

[1] Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025 Feb 24;

[2] Guo F, Kuo Y, Shih YCT, Giordano SH, Berenson AB. Trends in breast cancer mortality by stage at diagnosis among young women in the United States. Cancer. 2018 Sep 6;124(17):3500–9.

[3] Nicosia L, Gnocchi G, Gorini I, Venturini M, Fontana F, Pesapane F, et al. History of Mammography: Analysis of Breast Imaging Diagnostic Achievements over the Last Century. Healthcare. 2023 May 30;11(11):1596.

[4] Wang J. Breast cancer detection via wavelet energy and feed-forward neural network trained by genetic algorithm. EAI Endorsed Transactions on AI and Robotics. 2023 Sep 5;2.

[5] Ginsburg O, Yip C, Brooks A, Cabanes A, Caleffi M, Dunstan Yataco JA, et al. Breast cancer early detection: A phased approach to implementation. Cancer. 2020 May 15;126(S10):2379–93.

[6] Iranmakani S, Mortezazadeh T, Sajadian F, Ghaziani MF, Ghafari A, Khezerloo D, et al. A review of various modalities in breast imaging: technical aspects and clinical outcomes.

Egyptian Journal of Radiology and Nuclear Medicine. 2020 Dec 16;51(1):57.
[7] Kavitha T, Mathai PP, Karthikeyan C, Ashok M, Kohar R, Avanija J, et al. Deep Learning Based Capsule Neural Network Model for Breast Cancer Diagnosis Using Mammogram Images. Interdiscip Sci. 2022 Mar 2;14(1):113–29.
[8] Jose Escorcia-Gutierrez, F. Mansour R, Beleno K, Jimenez-Cabas J, Perez M, Madera N, et al. Automated Deep Learning Empowered Breast Cancer Diagnosis Using Biomedical Mammogram Images. Computers, Materials & Continua [Internet]. 2022;71(3):4221–35. Available from: https://www.techscience.com/cmc/v71n3/46473
[9] El Houby EMF, Yassin NIR. Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. Biomed Signal Process Control. 2021 Sep;70:102954.
[10] Qasrawi R, Daraghmeh O, Qdaih I, Thwib S, Vicuna Polo S, Owienah H, et al. Hybrid ensemble deep learning model for advancing breast cancer detection and classification in clinical applications. Heliyon. 2024 Oct;10(19):e38374.
[11] Montesinos López OA, Montesinos López A, Crossa J. Overfitting, Model Tuning, and Evaluation of Prediction Performance. In: Multivariate Statistical Machine Learning Methods for Genomic Prediction. Cham: Springer International Publishing; 2022. p. 109–39.
[12] Zhang N, Zhang E, Li F. A Soybean Classification Method Based on Data Balance and Deep Learning. Applied Sciences. 2023 May 24;13(11):6425.
[13] Fan FJ, Shi Y. Effects of data quality and quantity on deep learning for protein-ligand binding affinity prediction. Bioorg Med Chem. 2022 Oct;72:117003.
[14] Sendari S, Buana FK, Wirawan IM, Wibowo FS, Tibyani. Smart Parking: Symbol Detection Using Haar Cascade Classifier Algorithm To Monitor Empty Parking Slots. In: 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE). IEEE; 2021. p. 431–5.
[15] Utama ABP, Wibawa AP, Handayani AN, Chuttur MY. Exploring the Role of Deep Learning in Forecasting for Sustainable Development Goals: A Systematic Literature Review. International Journal of Robotics and Control Systems. 2024 Mar 31;4(1):365–400.
[16] World Medical Association. World Medical Association Declaration of Helsinki. JAMA. 2013 Nov 27;310(20):2191.
[17] Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: Toward a Full-field Digital Mammographic Database. Acad Radiol. 2012 Feb;19(2):236–48.
[18] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. Mammographic Image Analysis Society (MIAS) database v1.21. [Dataset]. Apollo - University of Cambridge Repository. University of Cambridge; 2015.
[19] 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 Dec 19;4(1):170177.
[20] Iswahyudi W, Farhan Ali Irfani M, Dwi Mahandi Y, Ari Elbaith Zaeni I, Sendari S, Widiyaningtyas T. Analyzing the Effectiveness of MobileNetV2, Xception, and DenseNet for Classifying Chest Diseases: Pneumonia, Pneumothorax,and Cardiomegaly. In: 2024 International Conference on Electrical Engineering and Computer Science (ICECOS). IEEE; 2024. p. 251–5.
[21] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020 Oct 22;63(11):139–44.
[22] Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models. In: Advances in Neural Information Processing Systems 33. 2020.
[23] Kingma DP, Welling M. An Introduction to Variational Autoencoders. Foundations and Trends® in Machine Learning. 2019;12(4):307–92.
[24] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal. 2019 Dec;58:101552.
[25] Cheng ST, Yeh CW. Two-Stage Metal Surface Defect Detection and Classification System Based on VAEGAN and Class-Embedding. EAI Endorsed Transactions on AI and Robotics. 2023 Nov 29;2.
[26] Yilmaz B, Korn R. A Comprehensive guide to Generative Adversarial Networks (GANs) and application to individual electricity demand. Expert Syst Appl. 2024 Sep;250:123851.
[27] Rais K, Amroune M, Benmachiche A, Haouam MY. Exploring Variational Autoencoders for Medical Image Generation: A Comprehensive Study. In: Proceedings of ACN International Conference. Istanbul; 2024. p. 1–5.
[28] Müller-Franzes G, Niehues JM, Khader F, Arasteh ST, Haarburger C, Kuhl C, et al. A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. Sci Rep. 2023 Dec 1;13(1).
[29] Xiao Z, Kreis K, Vahdat A. Tackling the Generative Learning Trilemma with Denoising Diffusion GANs. In: International Conference on Learning Representations (ICLR). 2022.
[30] Sutjiadi R, Sendari S, Herwanto HW, Kristian Y. Generating High-quality Synthetic Mammogram Images Using Denoising Diffusion Probabilistic Models: a Novel Approach for Augmenting Deep Learning Datasets. In: 2024 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE; 2024. p. 386–92.
[31] Sutjiadi R, Sendari S, Herwanto HW, Kristian Y. Deep Learning for Segmentation and Classification in Mammograms for Breast Cancer Detection: A Systematic Literature Review. Advanced Ultrasound in Diagnosis and Therapy. 2024;8(3):94–105.
[32] Tan M, Le Q V. EfficientNetV2: Smaller Models and Faster Training. Proceedings of the 38th International Conference on Machine Learning [Internet]. 2021 Apr 1;10096–106. Available from: http://arxiv.org/abs/2104.00298